

Stationary subspace analysis of nonstationary covariance processes: Eigenstructure description and testing

RAANJU R. SUNDARARAJAN¹, VLADAS PIPIRAS² and MOHSEN POURAHMADI³

¹*Department of Statistical Science, Southern Methodist University. E-mail: rsundararajan@smu.edu*

²*Department of Statistics and Operations Research, University of North Carolina. E-mail: pipiras@email.unc.edu*

³*Department of Statistics, Texas A&M University. E-mail: pourahm@stat.tamu.edu*

Stationary subspace analysis (SSA) searches for linear combinations of the components of nonstationary vector time series that are stationary. These linear combinations and their number define an associated stationary subspace and its dimension. SSA is studied here for zero mean nonstationary covariance processes. We characterize stationary subspaces and their dimensions in terms of eigenvalues and eigenvectors of certain symmetric matrices. This characterization is then used to derive formal statistical tests for estimating dimensions of stationary subspaces. Eigenstructure-based techniques are also proposed to estimate stationary subspaces, without relying on previously used computationally intensive optimization-based methods. Finally, the introduced methodologies are examined on simulated and real data.

Keywords: dimension test; eigen-decomposition; local and global dimensions; multivariate nonstationarity

1. Introduction

The goal of this work is to provide new and basic insights into the so-called stationary subspace analysis (SSA), a technique for finding linear combinations of components of a multivariate time series that are stationary. More precisely, consider an ideal setup where the observed p -vector nonstationary time series X_t is a linear transformation of a d -vector stationary series U_t^s and a $(p-d)$ -vector nonstationary series U_t^n through

$$X_t = MU_t = \begin{bmatrix} M_s & M_n \end{bmatrix} \begin{bmatrix} U_t^s \\ U_t^n \end{bmatrix}, \quad (1.1)$$

and M is an unknown $p \times p$ (invertible) mixing matrix, M_s and M_n are $p \times d$ and $p \times (p-d)$ matrices, respectively. It is further assumed that no linear transformation of U_t^n is stationary. Given the data X_1, X_2, \dots, X_T , SSA seeks to find a demixing matrix $B = (M^{-1})'$ so that $B'X_t = U_t$ is naturally partitioned into its stationary and nonstationary sources. The space spanned by the first d columns of B is referred to as a stationary subspace and d as its dimension.

SSA was introduced and studied by von Bünau et al. [24], with applications to analyzing EEG data from neuroscience experiments. In that work, the observed vector time series is assumed to be independent across time and the notion of stationarity is with respect to the first two moments, that is, the mean and lag-0 covariance are required to be time invariant. The demixing matrix in SSA is found in the spirit of ANOVA by dividing the observed time series data into N segments and minimizing a Kullback–Leibler (KL) divergence between Gaussian distributions measuring differences in the means and covariances across these segments. A sequential likelihood ratio test is used in von Bünau et al.

[24] and Blythe et al. [1] to determine the dimension of the stationary subspace d under the additional assumption of normality of the data. The frequency domain or dependent SSA (DSSA) in Sundararajan and Pourahmadi [22] avoids independence and dividing the data into segments. However, conceptually DSSA reduces this more general formulation to that of the SSA by relying on the approximate uncorrelatedness of the discrete Fourier transform (DFT) of a second-order stationary time series at Fourier frequencies. There, the sum of the Frobenius norms of the estimated covariances of the DFTs at the first few lags is used as a discrepancy measure and the demixing matrix is obtained by optimizing this measure. Finally, a sequential test of second-order stationarity is used to determine d and the consistency of the estimated d is studied using the asymptotic distribution of the test statistic under the alternative hypothesis of local stationarity of the time series (Dahlhaus [4,5]). Another technique closely related to SSA is the notion of “costationarity” introduced and studied by Cardinali and Nason [3] wherein stationary linear transformations of bivariate locally stationary time series is sought. The linear transformation is obtained by optimizing over a measure of nonstationarity that involves a time-varying spectral density.

Overall, the research thus far suggests the need for a better mathematical formulation and understanding of the problem, ideally providing a transparent and interpretable solution of the SSA problem. In this work, we shall tackle these issues for a special but general case of (1.1), namely, that of zero mean vectors U_t , assuming $U_t = C(t)Y_t$ with i.i.d. zero mean vectors Y_t . We shall further write this model formulation as

$$X_t = A\left(\frac{t}{T}\right)Y_t, \quad t = 1, 2, \dots, T, \quad (1.2)$$

where T is the sample size and the time dependence is brought into $A\left(\frac{t}{T}\right) = M \cdot C(t)$. Additional assumptions on the matrix-valued function $A : (0, 1) \rightarrow \mathbb{R}$ and the i.i.d. vectors Y_t can be found below in Section 2. The modification (1.2) of (1.1) takes the heterogeneity out of U_t and places it into the deterministic matrix-valued function $A(\cdot)$. The nonstationary covariance process (1.2) will be said to follow a *varying covariance* (VC) model. The assumption of zero mean in (1.2) is made for several reasons. For one, all previous works that identify EEG data analysis as the motivating application involve this assumption. We are currently working in parallel on analogous approaches to the SSA problem for time-varying means (Düker et al. [9]), and will possibly look at the combined model in the future. In the latter regard, we should also note that dealing with SSA for varying covariances is seemingly much more involved than for varying means. Indeed, as seen in this work, the SSA for the VC model has a surprisingly rich structure.

In contrast to optimization-based approaches to SSA reviewed earlier, a key feature of our approach for VC models is its direct reliance on eigenanalysis of certain symmetric matrices. This is inline with the initial heuristic formulation of popular multivariate statistical techniques such as principal components analysis (PCA), factor analysis and cointegration analysis, and their eventual formulations/solutions involving eigenanalysis of certain covariance matrices. More specifically, our contributions to the SSA for the VC model (1.2) are as follows. First, by using basic ideas from linear algebra, we provide an interpretation of a stationary subspace and its dimension d in terms of eigenvalues and eigenvectors of certain symmetric matrices. This interpretation, in fact, is given at two levels: “local” or for fixed $u \in (0, 1)$, and “global” or for $(0, 1) = \{u : u \in (0, 1)\}$, where u is thought here as a variable of $A(u)$ replaced by t/T in (1.2). Second, in the context of the obtained interpretation, we develop a formal statistical test for dimensions of stationary subspaces. Together with the algebraic interpretation, this test is the key theoretical contributions of this work. The test involves “local” quantities through an integral and in that sense is “global”. Third, by leveraging the new interpretation of stationary subspaces, we provide more direct and algebraic ways to construct them. These are shown to outperform the computationally more expensive optimization-based solutions of the previous SSA approaches in

a number of simulation settings. We should also note that the proposed dimension tests in Section 4 assume the existence of a common stationary subspace across the “local” levels (see Section 3.2 for more details); testing for the latter remains an open problem. Furthermore, this work concerns the asymptotics under $T \rightarrow \infty$ with a fixed p . Fourth, we revisit an SSA application to EEG data from a Brain-Computer Interface (BCI) experiment and provide additional insights by using the proposed methods.

The outline of the paper is as follows. Section 2 reintroduces more formally the VC model and its stationary subspace and dimension. Section 3 gives an eigenstructure-based characterization of a stationary subspace and its dimension. Section 4 introduces the “local” quantities of interest and considers a statistical test for the dimension of a stationary subspace at the “global” level. Section 5 discusses estimation methods for stationary subspaces that are based on algebraic constructs and do not involve iterative and computationally heavy optimization methods. Sections 6 and 7 illustrate the proposed methodologies using simulated and real data. Section 8 concludes.

2. Model of interest and its stationary subspace

We focus throughout this work on the varying covariance (VC) model (1.2), where $A : (0, 1) \rightarrow \mathbb{R}^{p \times p}$ is a smoothly varying matrix-valued function and Y_t are i.i.d. random vectors with i.i.d. entries, $\mathbb{E}(Y_t) = 0$ and $\mathbb{E}(Y_t Y_t') = I_p$. Further technical assumptions can be found below.

Definition 2.1. If d is the largest integer in $\{0, 1, 2, \dots, p\}$ for which there is a $p \times d$ matrix B_1 such that

$$B_1' A^2(u) B_1 = \Sigma, \quad \forall u \in (0, 1), \tag{2.1}$$

where $A^2(u) = A(u)A(u)'$, Σ does not depend on u and $B_1' B_1 = I_d$, then the space \mathcal{B}_1 spanned by the columns of B_1 will be called a (second-order) *stationary subspace* of dimension d of the model (1.2).

Note that (2.1) states effectively that the covariance matrix of $B_1' X_t$ does not depend on t . It can be reformulated as follows: Let $\bar{A}^2 = \int_0^1 A^2(u) du$ and define a $p \times p$ symmetric matrix $M(u)$ as

$$M(u) = A^2(u) - \bar{A}^2. \tag{2.2}$$

Then the condition (2.1) is equivalent to

$$B_1' M(u) B_1 = 0, \quad \forall u \in (0, 1). \tag{2.3}$$

Indeed, (2.1) implies (2.3) after integrating (2.1) over $u \in (0, 1)$, noting that $\int_0^1 \Sigma du = \Sigma$ and subtracting the two sides of the resulting relation $B_1' \bar{A}^2 B_1 = \Sigma$ from those of (2.1). Similarly (2.3) implies (2.1) with $\Sigma = B_1' \bar{A}^2 B_1$. The matrix $M(u)$ will play a central role henceforth.

Our approach to finding a matrix B_1 and the corresponding stationary subspace \mathcal{B}_1 and dimension d is based on the relation (2.3) for a fixed u , that is, $B_1' M(u) B_1 = 0$ for a fixed u and $B_1 = B_1(u)$ of dimension $d = d(u)$. As shown in the next section, for a fixed u , the matrix $B_1(u)$ and its dimension $d(u)$ can be characterized using the eigenstructure of the matrix $M(u)$. When it comes to $M(u)$, we shall be using the terminology of the following definition.

Definition 2.2. Let $B_1 = B_1(u)$ be a matrix with $d = d(u)$ columns that satisfies (2.3) for a fixed $u \in (0, 1)$. The space $\mathcal{B}_1(u)$ spanned by the columns of $B_1(u)$ will be called a *local stationary subspace*

of local dimension $d(u)$. The respective quantities in Definition 2.1 will be referred to as a *global stationary subspace* and a *global dimension*.

Relationships between local and global stationary subspaces and their dimensions are discussed in Section 3.2. The estimation of the local and global dimensions $d(u)$ and d is discussed in Section 4.

3. Matrix pseudo nullity and pseudo null space

In this section, we characterize a stationary subspace and its dimension from a matrix eigenstructure perspective. In view of the relation (2.3) and Definition 2.2, we start with the following definition.

Definition 3.1. Let M be a $p \times p$ symmetric matrix. A *pseudo nullity* of M , denoted by $d(M)$, is defined as the largest non-negative integer d_1 such that

$$C_1' M C_1 = 0, \quad (3.1)$$

for a $p \times d_1$ matrix C_1 with $C_1' C_1 = I_{d_1}$. A *pseudo null space* of M , denoted as $\mathcal{P}(M)$, is defined as the linear span of the d_1 columns of the matrix C_1 in (3.1). A column of C_1 , that is, a column vector s such that $s' M s = 0$ will be called a *pseudo eigenvector*.

If M is positive semi-definite, note that its pseudo nullity is its nullity (i.e., the number of zero eigenvalues of M) and its pseudo null space is its null space; thus, the pseudo- terminology is used to draw attention to the contrast between these two cases. Otherwise, we should caution the reader against drawing other parallels between the two contexts. For example, if s_1 and s_2 are two pseudo eigenvectors (which can be either orthogonal or non-orthogonal), note that there is a priori no reason to have $s_1' M s_2 = 0$ and hence $C_1' M C_1 = 0$ with $C_1 = (s_1 \ s_2)$. In particular, if for example, $d(M) = 2$, and s_1 and s_2 are orthogonal, the linear space spanned by s_1 and s_2 is not necessarily a pseudo null space.

Another word of caution is that $\mathcal{P}(M)$ is not unique in general. This, perhaps surprising, fact will be explained below. By writing $\mathcal{P}(M)$, we mean one of the pseudo null spaces.

3.1. Characterization of pseudo nullity and pseudo null space

We characterize first the pseudo nullity $d(M)$ of a symmetric matrix M in terms of its inertia. Let

$$d_0 = d_0(M), \quad d_+ = d_+(M), \quad d_- = d_-(M) \quad (3.2)$$

be the number of zero, positive and negative eigenvalues of M , respectively.

Proposition 3.1. Let M be a symmetric matrix. Then, $d(M) = d_0 + \min(d_+, d_-)$.

The next result characterizes the pseudo null space $\mathcal{P}(M)$ and its pseudo eigenvectors. Let s_i ($s_{0,i}$, $s_{+,i}$ and $s_{-,i}$, resp.) be the orthonormal eigenvectors (associated with the zero, positive and negative eigenvalues, resp.) of M . The corresponding eigenvalues are denoted λ_i ($\lambda_{0,i} = 0$, $\lambda_{+,i}$, $\lambda_{-,i}$ resp.). We also let $\mathcal{N}_0(M)$ denote the null space of the matrix M , that is, the linear space spanned by the eigenvectors $s_{0,i}$, $i = 1, \dots, d_0$.

Proposition 3.2. *Let M be a symmetric matrix let $\mathcal{P}(M)$ be a pseudo null space of M . Then,*

$$\mathcal{P}(M) = \mathcal{N}_0(M) \oplus \mathcal{N}_\pm(M), \tag{3.3}$$

where $\mathcal{N}_0(M)$ is the null space of M and $\mathcal{N}_\pm(M)$ is a linear space spanned by orthogonal eigenvectors $w_{k,\pm}$, $k = 1, 2, \dots, \min(d_-, d_+)$, expressed as

$$w_{k,\pm} = \sum_{i=1}^{d_+} \alpha_{k,+,i} s_{+,i} + \sum_{i=1}^{d_-} \alpha_{k,-,i} s_{-,i}, \tag{3.4}$$

where $\alpha_{k,+,i}, \alpha_{k,-,i} \in \mathbb{R}$ are such that for all (possibly the same) k_1, k_2 ,

$$\sum_{i=1}^{d_+} \alpha_{k_1,+,i} \alpha_{k_2,+,i} \lambda_{+,i} + \sum_{i=1}^{d_-} \alpha_{k_1,-,i} \alpha_{k_2,-,i} \lambda_{-,i} = 0. \tag{3.5}$$

Propositions 3.1 and 3.2 are proved in Appendix A. But we would like to provide here the basic idea behind the results, as the underlying perspectives are important to keep in mind through the rest of the paper. By using the above notation, note first that

$$s'_{0,i_1} M s_{0,i_2} = 0, \quad i_1, i_2 = 1, 2, \dots, d_0.$$

This immediately implies that $d(M) \geq d_0$ and that all eigenvectors associated with the zero eigenvalue are also pseudo eigenvectors. But other, quite different pseudo eigenvectors can also be constructed by exploiting the following observation. Consider two eigenvectors $s_{+,i}$ and $s_{-,i}$ associated with a positive eigenvalue $\lambda_{+,i}$ and negative eigenvalue $\lambda_{-,i}$, respectively. Since $s'_{+,i} M s_{+,i} = \lambda_{+,i}$, similarly for the negative eigenvector/value and $s'_{-,i} M s_{-,i} = 0$, note that a linear combination $\alpha_+ s_{+,i} + \alpha_- s_{-,i}$ with weights $\alpha_+, \alpha_- \in \mathbb{R}$ satisfies

$$(\alpha_+ s_{+,i} + \alpha_- s_{-,i})' M (\alpha_+ s_{+,i} + \alpha_- s_{-,i}) = \alpha_+^2 \lambda_{+,i} + \alpha_-^2 \lambda_{-,i}.$$

Note further that since $\lambda_{+,i} > 0$ and $\lambda_{-,i} < 0$, a suitable choice of weights α_+, α_- can ensure that $(\alpha_+ s_{+,i} + \alpha_- s_{-,i})' M (\alpha_+ s_{+,i} + \alpha_- s_{-,i}) = 0$ and hence that $\alpha_+ s_{+,i} + \alpha_- s_{-,i}$ is a pseudo eigenvector. Since $(\alpha_+ s_{+,i} + \alpha_- s_{-,i})' M s_{0,j} = 0$ by orthogonality of eigenvectors, observe also that both this pseudo eigenvector and all eigenvectors associated with the zero eigenvalue can be part of a pseudo null space. Finally, note that there is considerable flexibility in the described construction, in particular, related to which positive and negative eigenvectors/values are paired. Consequences of this flexibility are the stated results of Propositions 3.1 and 3.2.

We next illustrate the above results through an example. The example also shows that a pseudo null space is not unique in general.

Example 3.1. Let M be a symmetric matrix defined as

$$M = \text{diag}(1, -1, 1). \tag{3.6}$$

The eigenvalues of M are 1 (of multiplicity 2) and -1 . Then, $\lambda_{+,1} = \lambda_{+,2} = 1$, $\lambda_{-,1} = -1$ and $d_+ = 2$, $d_- = 1$, $d_0 = 0$. By Proposition 3.1, $d(M) = 0 + \min(1, 2) = 1$. The corresponding eigenvectors are $s_{+,1} = (1 \ 0 \ 0)'$, $s_{+,2} = (0 \ 0 \ 1)'$ and $s_{-,1} = (0 \ 1 \ 0)'$. By Proposition 3.2, a pseudo null space of M can be expressed as

$$\mathcal{P}(M) = \text{lin}\{\alpha_{-,1} s_{-,1} + \alpha_{+,1} s_{+,1} + \alpha_{+,2} s_{+,2}\}$$

such that

$$-\alpha_{-,1}^2 + \alpha_{+,1}^2 + \alpha_{+,2}^2 = 0$$

and $\alpha_{-,1}^2 + \alpha_{+,1}^2 + \alpha_{+,2}^2 = 1$ for the norm to be 1, where “lin” indicates a linear span. The latter two expressions yield $\alpha_{-,1} = (\alpha_{+,1}^2 + \alpha_{+,2}^2)^{1/2}$ (after choosing a positive sign for the square root) and $\alpha_{+,1}^2 + \alpha_{+,2}^2 = 1/2$. This further yields $\alpha := \alpha_{+,1} \in [1/\sqrt{2}, -1/\sqrt{2}]$, $\alpha_{+,2} = \pm(1/2 - \alpha^2)^{1/2}$ and $\alpha_{-,1} = 1/\sqrt{2}$. Thus, a pseudo null space can also be expressed as

$$\begin{aligned} \mathcal{P}(M) &= \text{lin}\{(1/\sqrt{2})s_{-,1} + \alpha s_{+,1} \pm (1/2 - \alpha^2)^{1/2} s_{+,2}\} \\ &= \text{lin}\{(\alpha, 1/\sqrt{2}, \pm(1/2 - \alpha^2)^{1/2})'\}, \end{aligned} \tag{3.7}$$

where $\alpha \in [1/\sqrt{2}, -1/\sqrt{2}]$. Note that these spaces (vectors) are generally different for different α 's. For example, for $\alpha = 0$,

$$\mathcal{P}(M) = \text{lin}\{(0, 1/\sqrt{2}, 1/\sqrt{2})'\} = \text{lin}\{(0 \quad 1 \quad 1)'\} \tag{3.8}$$

and for $\alpha = 1/\sqrt{2}$,

$$\mathcal{P}(M) = \text{lin}\{(1/\sqrt{2}, 1/\sqrt{2}, 0)'\} = \text{lin}\{(1 \quad 1 \quad 0)'\}. \tag{3.9}$$

3.2. Implications for stationary subspace and its dimension

In view of Definitions 2.1 and 2.2 and their notation, the global stationary subspace \mathcal{B}_1 and its dimension d are given by:

$$d = d(u) = d(M(u)), \quad \mathcal{B}_1 = \mathcal{B}_1(u) = \mathcal{P}(M(u)), \quad \forall u \in (0, 1). \tag{3.10}$$

In view of Proposition 3.1, the first relation in (3.10) can now be reformulated as

$$d = d(u) = d(M(u)) = d_0(u) + \min(d_+(u), d_-(u)), \quad \forall u \in (0, 1), \tag{3.11}$$

where

$$d_0(u) = d_0(M(u)), \quad d_+(u) = d_+(M(u)), \quad d_-(u) = d_-(M(u)) \tag{3.12}$$

are the numbers of zero, positive and negative eigenvalues of $M(u)$, respectively.

On the other hand, if

$$d(M(u)) = d_0(u) + \min(d_+(u), d_-(u)) \equiv d^*, \quad u \in (0, 1), \tag{3.13}$$

for some d^* , this does not necessarily mean that the dimension d of the stationary subspace is d^* . The latter is because (3.13) does not guarantee that

$$\mathcal{B}_1(u) = \mathcal{P}(M(u)) \equiv \mathcal{B}_1^*, \quad u \in (0, 1), \tag{3.14}$$

for some \mathcal{B}_1^* , playing the role of a stationary subspace. But if (3.14) holds, then (3.13) does imply that d^* is the dimension d of the stationary subspace. The statistical tests developed in the subsequent sections will, in fact, be for testing (3.13) and hence will lead to the dimension d assuming (3.14).

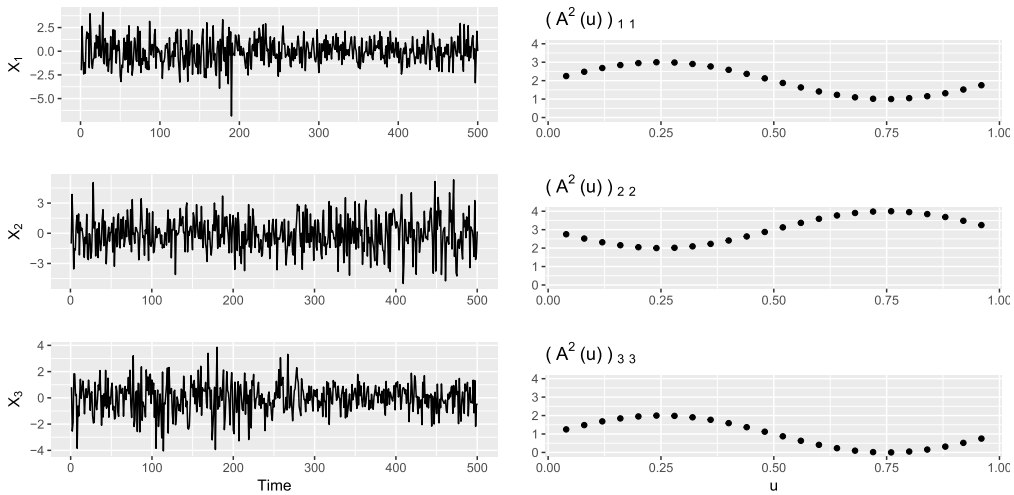


Figure 1. Left: Plot of a sample realization of length $T = 500$ based on the VC model using $A^2(u)$ from Example 3.2. Right: Visualization of the diagonal entries of $A^2(u)$ from Example 3.2.

How testing can be done for (3.14) remains an open question, though we shall also suggest new ways to estimate \mathcal{B}_1 , based on developments in Section 3.1.

Finally, we illustrate the observations made above through a simple but instructive example. The example will use some findings of Example 3.1. See also a subsequent remark.

Example 3.2. Consider a VC model with

$$A^2(u) = \text{diag}(2 + \sin(2\pi u), 3 - \sin(2\pi u), 1 + \sin(2\pi u)). \tag{3.15}$$

Then, $\overline{A}^2 = \text{diag}(2, 3, 1)$ and

$$M(u) = A^2(u) - \overline{A}^2 = \sin(2\pi u) \cdot \text{diag}(1, -1, 1). \tag{3.16}$$

Figure 1 (left) plots a sample realization of the VC model using $A(u)$ in (3.15) with Y_t being i.i.d. $\mathcal{N}(0, I_3)$. Figure 1 (right) plots the diagonal entries of $A^2(u)$.

For fixed $u \neq 1/2$, the eigenvalues of $M(u)$ are $\sin(2\pi u) \cdot 1$ (of multiplicity 2) and $\sin(2\pi u) \cdot (-1)$. For further illustration, suppose $u \in (0, 1/2)$, so that $\sin(2\pi u) > 0$. Then, $\lambda_{+,1} = \lambda_{+,2} = \sin(2\pi u) \cdot 1$, $\lambda_{-,1} = \sin(2\pi u) \cdot (-1)$ and $d_+ = 2, d_- = 1, d_0 = 0$, by using the notation in Section 3.1 with $M = M(u)$. By Proposition 3.1, $d(u) = d(M(u)) = 0 + \min(1, 2) = 1$.

The corresponding eigenvectors are $s_{+,1} = (1 \ 0 \ 0)'$, $s_{+,2} = (0 \ 0 \ 1)'$ and $s_{-,1} = (0 \ 1 \ 0)'$. As in Example 3.1, a local stationary subspace or a pseudo null space of $M(u)$ can be expressed as

$$\mathcal{B}_1(u) = \mathcal{P}(M(u)) = \text{lin}\{(\alpha, 1/\sqrt{2}, \pm(1/2 - \alpha^2)^{1/2})'\}, \tag{3.17}$$

where $\alpha \in [1/\sqrt{2}, -1/\sqrt{2}]$.

Remark 3.1. The fact that a pseudo null space in Example 3.2 is not unique should not be surprising from the following perspective. Let $X_t = (X_{1,t}, X_{2,t}, X_{3,t})'$ be a 3-vector process following the VC

model with (3.15). The pseudo eigenvectors $w_1 = (0 \ 1 \ 1)'$ in (3.8) and $w_2 = (1 \ 1 \ 0)'$ in (3.9) can be checked easily to be such that $w_1'X_t$ and $w_2'X_t$ are stationary.

It is also interesting and important to note here that the stationary subspace dimension for this model is *not* 2. For example, note that while $w_1'X_t$ and $w_2'X_t$ are indeed stationary, the 2-vector process $(w_1'X_t, w_2'X_t)'$ is *not* stationary. Indeed, this is the case since e.g. $\mathbb{E}(w_1'X_t)(w_2'X_t) = \mathbb{E}(X_{2,t} + X_{3,t})(X_{1,t} + X_{2,t}) = \mathbb{E}X_{2,t}^2 = 3 - \sin(2\pi t/T)$ depends on t .

4. Inference of stationary subspace dimension

According to (3.11)–(3.12), the dimension of a stationary subspace of the VC model is the local dimension $d(u)$ or pseudo nullity $d(M(u))$ of the matrix $M(u)$ in (2.2), assuming it is the same across u , which can further be expressed in terms of $d_0(u)$, $d_+(u)$ and $d_-(u)$. We are interested here in the statistical testing for $d_0(u)$, $d_+(u)$, $d_-(u)$ and hence also for $d(u) = d(M(u))$. In view of (3.11), we shall focus on “global” tests, that is, over an interval of u . Our global test, however, will be the aggregate of local quantities, defined for fixed u . These quantities of interest are introduced and discussed first in Section 4.1. (A byproduct of this discussion are local dimension tests but these will not be our focus – see Remarks 4.1 and 4.2). The global dimension testing is treated in Sections 4.2 and 4.3.

4.1. Local quantities of interest

For statistical inference, we obviously need an estimator of the matrix $M(u)$. It is set naturally as

$$\widehat{M}(u) = \widehat{A}^2(u) - \widehat{A}^2 := \frac{1}{T} \sum_{t=1}^T X_t X_t' K_h\left(u - \frac{t}{T}\right) - \frac{1}{T} \sum_{t=1}^T X_t X_t', \tag{4.1}$$

where $K_h(u) = h^{-1}K(h^{-1}u)$, $K(\cdot)$ is a kernel function and h denotes the bandwidth. A kernel is a symmetric function which integrates to 1, with further regularity assumptions possibly made as well. In simulations and data applications, we work with the triangle kernel $K(u) = 1 - |u|$ if $|u| < 1$ and 0, otherwise.

We shall use the estimator $\widehat{M}(u)$ to make inference about the quantities $d_0(u)$, $d_+(u)$ and $d_-(u)$ characterizing $M(u)$. What makes this more challenging is that we would like this inference to be across u 's. To achieve this, we shall introduce here relevant quantities for this problem with u fixed, and then integrate them over u in Section 4.2 to devise a global test. For fixed u , inference about $d_0(u)$, $d_+(u)$ and $d_-(u)$ could be made by adapting the matrix rank tests found in Donald et al. [7].

The underlying assumption in Donald et al. [7] is an asymptotic normality of the matrix estimator, which is $\widehat{M}(u)$ as considered here. We show in the supplementary technical appendix (Sundararajan et al. [21]) that under suitable assumptions,

$$a_T F(u) (\widehat{M}(u) - M(u)) F(u)' \xrightarrow{d} \mathcal{Z}_p, \tag{4.2}$$

where

$$a_T = \frac{\sqrt{Th}}{\|K\|_2 \mu_4^{1/2}}, \quad F(u) = A(u)^{-1}, \tag{4.3}$$

$\|K\|_2^2 = \int_{\mathbb{R}} K(v)^2 dv$, $\mu_4 = \mathbb{E}(Y_{i,t}^2 - 1)^2 = \mathbb{E}Y_{i,t}^4 - 1$ and \mathcal{Z}_p is a symmetric $p \times p$ matrix having independent normal entries with variance 1 on the diagonal and variance $1/\mu_4$ off the diagonal. In

addition,

$$\widehat{F}(u) \xrightarrow{P} F(u) \tag{4.4}$$

with $\widehat{F}(u) = \widehat{A}(u)^{-1}$. We assume here for simplicity that μ_4 is known. Construction of consistent estimators $\widehat{\mu}_4$ is discussed in the supplementary technical appendix (Sundararajan et al. [21]), and would be sufficient for the results below to hold assuming that μ_4 is estimated through $\widehat{\mu}_4$. Furthermore, if one is willing to assume Gaussianity of Y_t , note that $\mu_4 = 2$ can also be used.

The matrix $F(u) = A(u)^{-1}$ plays the role of standardization in obtaining the standardized limit in (4.2). In this sense, it is natural to focus on $F(u)M(u)F(u)'$ rather than $M(u)$. Note that $d_0(u) = d_0(M(u)) = d_0(F(u)M(u)F(u)')$ and also $d_{\pm}(M(u)) = d_{\pm}(F(u)M(u)F(u)')$. We let

$$\gamma_1(u) \leq \dots \leq \gamma_p(u) \quad \text{and} \quad \widehat{\gamma}_1(u) \leq \dots \leq \widehat{\gamma}_p(u) \tag{4.5}$$

be the ordered eigenvalues of $F(u)M(u)F(u)'$ and $\widehat{F}(u)\widehat{M}(u)\widehat{F}(u)'$, respectively. We have

$$\begin{aligned} \gamma_1(u) \leq \dots \leq \gamma_{d_-(u)}(u) < 0 = \gamma_{d_-(u)+1}(u) = \dots = \gamma_{d_-(u)+d_0(u)}(u) \\ < \gamma_{d_-(u)+d_0(u)+1}(u) \leq \dots \leq \gamma_p(u). \end{aligned} \tag{4.6}$$

Let also

$$0 \leq \gamma_{2,1}(u) \leq \dots \leq \gamma_{2,p}(u) \quad \text{and} \quad \widehat{\gamma}_{2,1}(u) \leq \dots \leq \widehat{\gamma}_{2,p}(u) \tag{4.7}$$

be the ordered eigenvalues of $(F(u)M(u)F(u)')^2$ and $(\widehat{F}(u)\widehat{M}(u)\widehat{F}(u)')^2$, respectively. We have $(\gamma_i(u))^2 = \gamma_{2,j(i)}(u)$ for some $j(i)$, and a similar expression with the hats and also

$$0 = \gamma_{2,1}(u) = \dots = \gamma_{2,d_0(u)}(u) < \gamma_{2,d_0(u)+1}(u) \leq \dots \leq \gamma_{2,p}(u). \tag{4.8}$$

In particular, $d_0(u) = d_0(F(u)M(u)F(u)') = d_0((F(u)M(u)F(u)')^2)$. By combining these observations, we have

$$d_{\pm}(u) = \#\{i : \gamma_i(u) \geq 0, (\gamma_i(u))^2 = \gamma_{2,k}(u) \text{ for some } k = d_0(u) + 1, \dots, p\}. \tag{4.9}$$

In the proofs for Section 4.2, we shall also use eigenspaces associated with the eigenvalues above but these will not be discussed here.

To make inference about $d_0(u)$ and in view of (4.8), one would expect $\widehat{\gamma}_{2,1}(u), \dots, \widehat{\gamma}_{2,d_0(u)}$ to be small, while the rest of the eigenvalues is larger. Following Donald et al. [7], a natural test statistic to consider in this regard is: for $r = 0, 1, \dots, p$,

$$\widehat{\xi}_r(u) = a_T^2 \sum_{i=1}^r \widehat{\gamma}_{2,i}(u) = \frac{Th}{\|K\|_2^2 \mu_4} \sum_{i=1}^r \widehat{\gamma}_{2,i}(u). \tag{4.10}$$

This statistic will be integrated over u in Section 4.2 to devise a global test.

Remark 4.1. Under the assumptions leading to (4.2), it follows from Theorem 4.7 in Donald et al. [7] that, under $H_0 : d_0(u) = r$,

$$\widehat{\xi}_r(u) \xrightarrow{d} \chi^2(r(r+1)/2) \tag{4.11}$$

and under $H_1 : d_0(u) < r$, $\widehat{\xi}_r(u) \rightarrow_p +\infty$. This result can be used to test for $d_0(u)$ in a standard way sequentially, namely, testing for $H_0 : d_0(u) = r$ starting with $r = p$ and subsequently decreasing r by

1 till the null hypothesis is not rejected. Letting $\widehat{d}_0(u)$ be the resulting estimator, it can be shown to be consistent for $d_0(u)$ under a suitable choice of critical values in the sequential testing (see, e.g., Theorem 4.3 in Fortuna [10]).

Remark 4.2. If $\widehat{d}(u)$ is a consistent estimator of $d_0(u)$ in the previous remark and in view of (4.9), it is natural to estimate $d_{\pm}(u)$ as

$$\widehat{d}_{\pm}(u) = \#\{i : \widehat{\gamma}_i(u) \geq 0, (\widehat{\gamma}_i(u))^2 = \widehat{\gamma}_{2,k}(u), \text{ for some } k = \widehat{d}_0(u) + 1, \dots, p\} \tag{4.12}$$

and $d(u)$ as

$$\widehat{d}(u) = \widehat{d}(M(u)) = \widehat{d}_0(u) + \min(\widehat{d}_+(u), \widehat{d}_-(u)). \tag{4.13}$$

The consistency of $\widehat{d}_{\pm}(u)$ and $\widehat{d}(u)$ could be proved by using that of $\widehat{d}_0(u)$ where $\widehat{d}_0(u)$ is defined in Remark 4.1.

4.2. Global dimension test

In this section, we consider testing for $d_0(u)$, $d_+(u)$, $d_-(u)$ and $d(M(u))$ ‘‘globally’’, that is, for an interval of u . Since our approach to $d(M(u))$ goes through $d_0(u)$, $d_+(u)$ and $d_-(u)$, we shall make inference about these quantities first. To deal with the possibility that these quantities might differ across $u \in (0, 1)$, we shall work under the assumption that

$$d_0(u) \equiv d_0, \quad d_+(u) \equiv d_+, \quad d_-(u) \equiv d_-, \quad \forall u \in \mathcal{H} \subset (0, 1), \tag{4.14}$$

and develop a global dimension test about d_0 , d_+ and d_- in (4.14) for fixed \mathcal{H} . In practice, we shall apply the developed test over refined dyadic partitions, first for $(0, 1)$, then for $(0, 1/2)$ and $(1/2, 1)$, then for $(0, 1/4)$, $(1/4, 1/2)$, $(1/2, 3/4)$ and $(3/4, 1)$, etc. What to expect under this splitting of the interval $(0, 1)$ is discussed in Section 4.3, and will be illustrated in Sections 6 and 7.

We first focus on inference about d_0 in (4.14). Our global test will be based on the quantity $\int_{\mathcal{H}} \widehat{\xi}_r(u) du$, where $\widehat{\xi}_r(u)$ is the test statistic (4.10). Its asymptotics are described in the next result, which also defines the global statistic $\widehat{\xi}_r$. Recall the notation $\mu_4 = \mathbb{E}(Y_{i,t} - 1)^4$ and the discussion following (4.4). We also let $|\mathcal{H}|$ be the length of the interval \mathcal{H} and $\overline{K}(u) = \int_{\mathbb{R}} K(v)K(u - v) dv$ be the so-called convolution kernel.

Proposition 4.1. *Suppose that the assumptions stated at the beginning of Appendix A.3 hold. Then, under $H_0 : d_0(u) \equiv r$ for all $u \in \mathcal{H}$,*

$$\widehat{\xi}_r := \widehat{\xi}_{r,\mathcal{H}} := \frac{\int_{\mathcal{H}} \widehat{\xi}_r(u) du - |\mathcal{H}|^{\frac{r(\mu_4+r-1)}{\mu_4}}}{\sqrt{h \frac{\|\overline{K}\|_2^2}{\|\overline{K}\|_4^4} \frac{2(r\mu_4^2+2r(r-1))}{\mu_4^2} |\mathcal{H}|}} \xrightarrow{d} \mathcal{N}(0, 1) \tag{4.15}$$

and under $H_1 : d_0(u) < r$ for some $u \in \mathcal{H}$, $\widehat{\xi}_r \rightarrow_p \infty$.

The proof of the proposition can be found in Appendix A.3, and follows the approach taken in Donald et al. [8]. The required assumptions on Y_t to establish the above result are that Y_t is i.i.d. with $E(Y_t) = 0$, $E(Y_t Y_t') = I_p$ and has finite absolute moments of order $4 + \epsilon$ for some $\epsilon > 0$. The entries of $A^2(u)$ are assumed to be real analytic (possessing all order derivatives) for $u \in \mathcal{H}$, a closed

subinterval of $(0, 1)$. Additionally, the assumption on h is that $h \rightarrow 0$, $Th^{3/2} \rightarrow \infty$, $T^{\epsilon/(4+\epsilon)}h^{1/2} \rightarrow \infty$; see Appendix A.3 for details.

As in Remark 4.1, Proposition 4.1 can be used to test for d_0 sequentially, namely, testing for $H_0 : d_0 = r$ starting with $r = p$ and subsequently decreasing r by 1 till the null hypothesis is not rejected. Let \widehat{d}_0 be the resulting estimator obtained using this sequential procedure with significance level α_T .

Corollary 4.1. *Suppose that the assumptions of Proposition 4.1 hold. If $\alpha_T \rightarrow 0$ and $(-\log \alpha_T)/Th \rightarrow 0$, then $\widehat{d}_0 \rightarrow_p d_0$.*

Proof. The result can be proved as in for example, Theorem 4.3 in Fortuna [10], by noting that the two conditions provided in the statement of the theorem are in fact equivalent. \square

We now turn to inference about d_+ and d_- in (4.14). Recall the definition of the eigenvalues $\widehat{\gamma}_i(u)$ in (4.5), whose squares are the eigenvalues $\widehat{\gamma}_{2,j}(u)$ in (4.7) entering the test statistics $\widehat{\xi}_r(u)$ and $\widehat{\zeta}_r$. From the discussions surrounding (4.5)–(4.8), the \widehat{d}_0 consecutive eigenvalues $\widehat{\gamma}_i(u)$ can be thought to be associated with the d_0 zero eigenvalues of $F(u)M(u)F(u)'$. If we can estimate the starting index for these consecutive eigenvalues, we could then deduce the numbers d_{\pm} of positive and negative eigenvalues of $F(u)M(u)F(u)'$. In the case $\widehat{d}_0 \geq 1$, the above suggests to consider

$$\widehat{\zeta}_r = \left| \int_{\mathcal{H}} (\widehat{\gamma}_r(u) + \dots + \widehat{\gamma}_{r+\widehat{d}_0-1}(u)) du \right|, \quad r = 1, \dots, p - \widehat{d}_0 + 1, \tag{4.16}$$

that is, the quantities involving the sums of the \widehat{d}_0 consecutive eigenvalues $\widehat{\gamma}_i(u)$, and to estimate this starting index through

$$\widehat{r} = \underset{r=1, \dots, p-\widehat{d}_0+1}{\operatorname{argmin}} \widehat{\zeta}_r. \tag{4.17}$$

(Whenever r does not match the starting index, a larger value of $\widehat{\zeta}_r$ is expected, since it will be driven by $\widehat{\gamma}_i(u)$ associated with the positive or negative eigenvalues of $F(u)M(u)F(u)'$.) With the estimated index \widehat{r} , it is then natural to set further

$$\widehat{d}_- = \widehat{r} - 1, \quad \widehat{d}_+ = p - \widehat{d}_0 - \widehat{d}_-. \tag{4.18}$$

If $\widehat{d}_0 = 0$, the preceding argument does not apply and, in fact, the quantity (4.16) is not even defined. In this case, we suggest to consider

$$\widehat{\eta}_r = \left| \int_{\mathcal{H}} (\widehat{\gamma}_1(u) + \dots + \widehat{\gamma}_r(u)) du \right| + \left| \int_{\mathcal{H}} (\widehat{\gamma}_{r+1}(u) + \dots + \widehat{\gamma}_p(u)) du \right|, \quad r = 0, \dots, p, \tag{4.19}$$

and to set

$$\widehat{d}_- = \underset{r=0, \dots, p}{\operatorname{argmax}} \widehat{\eta}_r, \quad \widehat{d}_+ = p - \widehat{d}_-. \tag{4.20}$$

The idea behind this definition and further motivation for using (4.16)–(4.18) can be found in the proof of the following corollary.

Corollary 4.2. *Under the assumptions of Corollary 4.1, we have $\widehat{d}_+ \rightarrow_p d_+$ and $\widehat{d}_- \rightarrow_p d_-$.*

Proof. The result follows from the following two observations. First, by Corollary 4.1, we may assume that $\widehat{d}_0 = d_0$ on an event indexed by T , whose probability converges to 1 as $T \rightarrow \infty$. Second, the

Algorithm 1: Estimating stationary subspace dimension

Output: Estimate \widehat{d} of stationary subspace dimension d in Definition 2.1.

Input: p -variate time series data $X_t, t = 1, 2, \dots, T$, the sub-interval $\mathcal{H} \subset (0, 1)$ given by a discretization $U = \{u_1 \leq u_2 \leq \dots \leq u_{n_u}\}$ and bandwidth h .

- 1: Calculate the estimates $\widehat{M}(u)$ and $\widehat{F}(u) = \widehat{A}^{-1}(u)$ in (4.1), the eigenvalues $\widehat{\gamma}_1(u) \leq \dots \leq \widehat{\gamma}_p(u)$ of $\widehat{M}(u)$ and $\widehat{\gamma}_{2,1}(u) \leq \dots \leq \widehat{\gamma}_{2,p}(u)$ of $(\widehat{F}(u)\widehat{M}(u)\widehat{F}(u)')^2$ for every $u \in U$.
 - 2: Set $r = p$ and compute the test statistic $\widehat{\xi}_r$ in (4.15) with integral replaced by Riemannian sum over U and $|U| = u_{n_u} - u_1$. Sequentially decrease r by 1 until $H_0 : d_0(u) = r$ for all $u \in U$ is not rejected by using normal critical values.
 - 3: With r from Step 2, set $\widehat{d}_0 = r$. If $\widehat{d}_0 \geq 1$, do Step 4(a) else if $\widehat{d}_0 = 0$, do Step 4(b).
 - 4(a): Compute $\widehat{\zeta}_r$ for $r = 1, 2, \dots, p - \widehat{d}_0 + 1$ in (4.16) and the estimates $\widehat{d}_-, \widehat{d}_+$ in (4.17) and (4.18).
 - 4(b): Compute $\widehat{\eta}_r$ for $r = 0, 1, \dots, p$ in (4.19) and the estimates $\widehat{d}_-, \widehat{d}_+$ in (4.20).
 - 5: Output the estimate $\widehat{d} = \widehat{d}_0 + \min(\widehat{d}_+, \widehat{d}_-)$.
-

eigenvalues entering (4.16) and (4.19) converge (in probability) to their population counterparts, so that the relations (4.16) and (4.19) become in the limit, respectively, the relations $\zeta_r = |\int_{\mathcal{H}}(\gamma_r(u) + \dots + \gamma_{r+d_0-1}(u)) du|$ and $\eta_r = |\int_{\mathcal{H}}(\gamma_1(u) + \dots + \gamma_r(u)) du| + |\int_{\mathcal{H}}(\gamma_{r+1}(u) + \dots + \gamma_p(u)) du|$. The conclusion follows by observing that the population quantities satisfy (4.17), (4.18) and (4.20) with all the hats removed. \square

Finally, a natural estimator for $d := d_0 + \min(d_+, d_-)$ is

$$\widehat{d} = \widehat{d}_0 + \min(\widehat{d}_+, \widehat{d}_-). \tag{4.21}$$

Corollaries 4.1–4.2 imply the consistency of this estimator, which is the main result of this section.

Theorem 4.1. *Under the assumptions of Corollary 4.1, we have $\widehat{d} \rightarrow_p d$.*

We conclude this section with Algorithm 1 that summarizes the steps leading to \widehat{d} .

4.3. Global dimension tests under interval splitting

The global pseudo dimension test was developed in Section 4.2 under the assumption (4.14) for a subinterval $\mathcal{H} \subset (0, 1)$. When global testing is to be performed on $(0, 1)$, we suggest to carry out the introduced global test over subintervals of refined partitions. In this section, we describe how the procedure is carried out and the results can be interpreted.

We shall need the following basic property of the global test statistic $\widehat{\xi}_{r,\mathcal{H}}$ in (4.15). Suppose \mathcal{H}_1 and \mathcal{H}_2 are two disjoint intervals such that

$$\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2. \tag{4.22}$$

Then, since $\int_{\mathcal{H}} = \int_{\mathcal{H}_1} + \int_{\mathcal{H}_2}$ and $|\mathcal{H}| = |\mathcal{H}_1| + |\mathcal{H}_2|$, it follows from the definition (4.15) of $\widehat{\xi}_{r,\mathcal{H}}$ that

$$\widehat{\xi}_{r,\mathcal{H}} = \widehat{\xi}_{r,\mathcal{H}_1} \left(\frac{|\mathcal{H}_1|}{|\mathcal{H}|} \right)^{1/2} + \widehat{\xi}_{r,\mathcal{H}_2} \left(\frac{|\mathcal{H}_2|}{|\mathcal{H}|} \right)^{1/2}. \tag{4.23}$$

In particular, in view of Proposition 4.1, under $d_0(u) \equiv r$, $u \in \mathcal{H}_i$,

$$\widehat{\xi}_{r, \mathcal{H}_i} \xrightarrow{d} \mathcal{N}(0, 1) =: Z_i, \quad i = 1, 2, \tag{4.24}$$

where Z_i 's can be considered independent because the VC model involves independent variables across \mathcal{H}_1 and \mathcal{H}_2 . The relations (4.23) and (4.24) then imply

$$\widehat{\xi}_{r, \mathcal{H}} \xrightarrow{d} Z_1 \left(\frac{|\mathcal{H}_1|}{|\mathcal{H}|} \right)^{1/2} + Z_2 \left(\frac{|\mathcal{H}_2|}{|\mathcal{H}|} \right)^{1/2} \stackrel{d}{=} \mathcal{N}(0, 1), \tag{4.25}$$

which is consistent with what is expected under $d_0(u) \equiv r$, $u \in \mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2$ by Proposition 4.1. Such considerations will allow having some consistency over refined partitions in the sense described below. We first consider the case of $d_0(u)$, and then discuss those of $d_{\pm}(u)$ and $d(u) = d(M(u))$.

Thus, let

$$\mathcal{H}_i^{(k)} = \left(\frac{i-1}{2^k}, \frac{i}{2^k} \right], \quad i = 1, \dots, 2^k, \tag{4.26}$$

form refined partitions of $(0, 1]$ for $k \geq 0$. Let $\widehat{d}_{0,i}^{(k)}$ be the global estimator of d_0 over $\mathcal{H}_i^{(k)}$ by using the test statistic $\widehat{\xi}_{r, \mathcal{H}_i^{(k)}}$. In view of (4.23)–(4.25) and for the sake of consistency, when estimating d_0 over finer partitions, we suggest to adjust the normal critical value for comparing $\widehat{\xi}_{r, \mathcal{H}_i^{(k)}}$. More precisely, if $c_{\alpha}^{(0)} = c_{\alpha}$ is a normal critical value used at the level $k = 0$, we use the critical value $c_{\alpha}^{(k)} = 2^{-k/2} c_{\alpha}$ at level k .

As a consequence of the choice of the critical values, there are only the following three possibilities for estimators $\widehat{d}_{0,i}^{(k)}$, $\widehat{d}_{0,2i-1}^{(k+1)}$ and $\widehat{d}_{0,2i}^{(k+1)}$ when refining a partition from k to $k + 1$:

- (P1) $\widehat{d}_{0,i}^{(k)} = r$, $\widehat{d}_{0,2i-1}^{(k+1)} = r$, $\widehat{d}_{0,2i}^{(k+1)} = r$;
- (P2) $\widehat{d}_{0,i}^{(k)} = r$ and either $\widehat{d}_{0,2i-1}^{(k+1)} = r$, $\widehat{d}_{0,2i}^{(k+1)} < r$ or $\widehat{d}_{0,2i-1}^{(k+1)} < r$, $\widehat{d}_{0,2i}^{(k+1)} = r$;
- (P3) $\widehat{d}_{0,i}^{(k)} = r$ and either $\widehat{d}_{0,2i-1}^{(k+1)} > r$, $\widehat{d}_{0,2i}^{(k+1)} > \widehat{d}_{0,2i}^{(k+1)}$ or $\widehat{d}_{0,2i}^{(k+1)} > r$, $\widehat{d}_{0,2i}^{(k+1)} > \widehat{d}_{0,2i-1}^{(k+1)}$.

Indeed, let us explain the first two of these possibilities, and also indicate a case which is not one of the possibilities.

The possibility (P1) arises in the following scenario: one has $\widehat{d}_{0,i}^{(k)} = r$ when $\widehat{\xi}_{r, \mathcal{H}_i^{(k)}} > 2^{-k/2} c_{\alpha}$ for $j = p, p - 1, \dots, r + 1$ and $\widehat{\xi}_{r, \mathcal{H}_i^{(k)}} \leq 2^{-k/2} c_{\alpha}$. Similarly, $\widehat{d}_{0,2i-1}^{(k+1)} = r$ and $\widehat{d}_{0,2i}^{(k+1)} = r$ when $\widehat{\xi}_{j, \mathcal{H}_{2i-1}^{(k+1)}} \leq 2^{-(k+1)/2} c_{\alpha}$ for $j = p, p - 1, \dots, r + 1$ and $\widehat{\xi}_{r, \mathcal{H}_{2i-1}^{(k+1)}}, \widehat{\xi}_{r, \mathcal{H}_{2i}^{(k+1)}} \leq 2^{-(k+1)/2} c_{\alpha}$. This is consistent with the special case of (4.23),

$$\widehat{\xi}_{j, \mathcal{H}_i^{(k)}} = \widehat{\xi}_{j, \mathcal{H}_{2i-1}^{(k+1)}} \cdot 2^{1/2} + \widehat{\xi}_{j, \mathcal{H}_{2i}^{(k+1)}} \cdot 2^{1/2}, \tag{4.27}$$

in the sense that the relationships of the two summands of (4.27) to the respective critical values is the same as that for their sum.

The possibility (P2), on the other hand, corresponds to the scenario when $\widehat{\xi}_{j, \mathcal{H}_{2i-1}^{(k+1)}}, \widehat{\xi}_{j, \mathcal{H}_{2i}^{(k+1)}} > 2^{-(k+1)/2} c_{\alpha}$ for $j = p, p - 1, \dots, r + 1$, but then either $\widehat{\xi}_{r, \mathcal{H}_{2i-1}^{(k+1)}} \leq 2^{-(k+1)/2} c_{\alpha}$ and $\widehat{\xi}_{r, \mathcal{H}_{2i}^{(k+1)}} > 2^{-(k+1)/2} c_{\alpha}$ or $\widehat{\xi}_{r, \mathcal{H}_{2i-1}^{(k+1)}} > 2^{-(k+1)/2} c_{\alpha}$ and $\widehat{\xi}_{r, \mathcal{H}_{2i}^{(k+1)}} \leq 2^{-(k+1)/2} c_{\alpha}$. The case that is excluded from the possibilities listed above, is $\widehat{d}_{0,2i-1}^{(k+1)} < r$ and $\widehat{d}_{0,2i}^{(k+1)} < r$ (and $\widehat{d}_{0,i}^{(k)} = r$) which would happen if

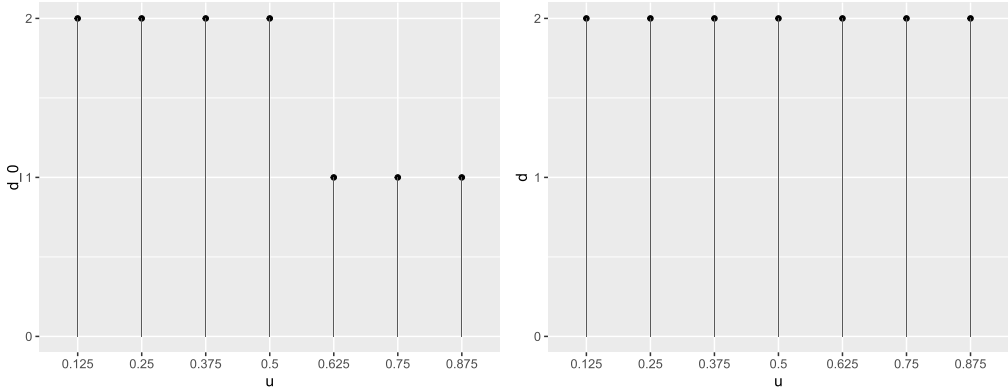


Figure 2. Example 4.1, Left: Visualization of the estimates of $\widehat{d}_{0,i}^{(k)}$ for $k = 0, 1, 2$ from one realization of a VC model. Right: Similar visualization of $\widehat{d}_i^{(k)}$.

$\widehat{\xi}_{r, \mathcal{H}_{2i-1}^{(k+1)}}, \widehat{\xi}_{r, \mathcal{H}_{2i}^{(k+1)}} > 2^{-(k+1)/2} c_\alpha$ but is impossible in view of (4.27). In our experiments with simulated and real data, the possibilities (P1) and (P2) seem to be the most common.

Following (4.16)–(4.20), the estimators $\widehat{d}_{0,i}^{(k)}$ lead to the corresponding estimates $\widehat{d}_{\pm,i}^{(k)}$ and $\widehat{d}_i^{(k)}$.

Example 4.1. We consider a VC model with $p = 3$, $d_0(u) = 2$ and $d(u) = 2$ if $u < 0.5$ and $d_0(u) = 1$ and $d(u) = 2$ if $u > 0.5$ and

$$A^2(u) = 1_{(0,0.5)}(u) \text{diag}(2 + \sin(2\pi u), 2.0901, 3) + 1_{(0.5,1)}(u) \text{diag}(2 + \sin(2\pi u), 4 + 3 \sin(2\pi u), 3).$$

Figure 2, left plot, presents global testing results under splitting for one realization of the above model. In this plot, the estimate $\widehat{d}_{0,1}^{(0)}$ for $(0, 1]$ is presented at the point $1/2$ of the x -axis which is the midpoint of $(0, 1]$. The estimator $\widehat{d}_{0,1}^{(1)}$ and $\widehat{d}_{0,2}^{(1)}$ are presented at points $1/4$ and $3/4$ respectively, which are the midpoints of the intervals $(0, 1/2]$ and $(1/2, 1]$. The presentation is continued in the same way till the level $k = 2$ is reached and the estimates $\widehat{d}_{0,1}^{(2)}, \widehat{d}_{0,2}^{(2)}, \widehat{d}_{0,3}^{(2)}, \widehat{d}_{0,4}^{(2)}$ are presented at the finest considered level. The corresponding estimates $\widehat{d}_i^{(k)}$ could be presented similarly as in the described left plot of Figure 2. This is illustrated in the right plot of Figure 2.

We note from Figure 2, left plot, that the splitting results for $k = 0$ and $k + 1 = 1$ correspond to the possibility (P2) and those for $k = 1$ and $k + 1 = 2$ correspond to the possibility (P1). The results reflect correctly the underlying assumed varying values of $d_0(u)$.

5. Estimation of the stationary subspace

We turn here to the estimation of a stationary subspace of the VC model, which is related to pseudo null spaces, given in Definition 3.1, of matrices $M(u)$ through (3.10). We shall not provide here any formal statistical tests related to a stationary subspace but rather make a number of related comments, inspired by the developments in Section 3.1. In Section 5.1, we introduce a particular class of local stationary subspaces. In Section 5.2, a graph-based method is provided that aims at forming clusters of the many local stationary subspaces introduced in Section 5.1. Finally, a technique to select one subspace out

of the clusters is provided and this serves as our estimate of a stationary subspace. Numerical and performance advantages of this estimator are discussed in Section 6.

5.1. (1, 1)-local stationary subspaces

The results of Section 3.1 show that pseudo null spaces have a rich structure and are typically not unique, for a given matrix. Motivated by the discussion following Proposition 3.2, we shall restrict our attention to their special cases, given in the following definition. The definition and subsequent developments use the notation of Section 3, namely that of $M, d_0, d_{\pm}, \lambda_i, \lambda_{0,i}, \lambda_{\pm,i}, s_i, s_{0,i}, s_{\pm,i}$.

Definition 5.1. Let M be a symmetric matrix and suppose $\min(d_+, d_-) \geq 1$. A (1, 1)-pseudo null space of M is defined as

$$\mathcal{P}_{(1,1)}(M) = \text{lin}\{s_{0,1}, \dots, s_{0,d_0}, \alpha_i s_{+,p(i)} + \beta_i s_{-,n(i)}, i = 1, \dots, \min(d_+, d_-)\}, \tag{5.1}$$

for fixed α_i, β_i , where $p(i) \in \{1, \dots, d_+\}$ are different across $i, n(i) \in \{1, \dots, d_-\}$ are different across i and

$$\alpha_i^2 \lambda_{+,p(i)} + \beta_i^2 \lambda_{-,n(i)} = 0, \quad i = 1, \dots, \min(d_+, d_-). \tag{5.2}$$

When $M = M(u)$ with $M(u)$ as in (2.2), a (1,1)-pseudo null space $\mathcal{P}(M(u))$ will be called a (1,1)-local stationary subspace and denoted as $\mathcal{B}_{(1,1)}(u)$.

The fact that $\mathcal{P}_{(1,1)}(M)$ defines a pseudo null space for M follows from Proposition 3.2.

Remark 5.1. The total number of (1, 1)-pseudo null spaces of M is

$$n(M) = \binom{d_+}{\min(d_+, d_-)} \cdot \binom{d_-}{\min(d_+, d_-)} \cdot (\min(d_+, d_-)!), \tag{5.3}$$

where the first two terms account for the selection of eigenvectors associated with the positive and negative eigenvalues, and the last term for pairing them off. Depending on the values of d_{\pm} , the total number (5.3) can be quite large: see, for example, with $d_+ = 3$ and $d_- = 5$, the number is 60.

Remark 5.2. The prefix “(1, 1)-” in Definition 5.1 refers to the fact that a pseudo eigenvector of a pseudo null space is constructed by taking one (1) eigenvector $s_{+,p(i)}$ associated with the positive eigenvalues and one (1) eigenvector $s_{-,n(i)}$ associated with the negative eigenvalues. More elaborate constructions are possible as well, for example, by taking two (2) eigenvectors associated with the positive eigenvalues and one (1) eigenvector with the negative ones, as in Example 3.2, which could be called a (2, 1)-pseudo null space. In this work though, we shall consider only (1, 1)-pseudo null spaces.

Example 5.1. Consider again the VC model from Example 3.2 given by

$$\begin{aligned} A^2(u) &= \text{diag}(2 + \sin(2\pi u), 3 - \sin(2\pi u), 1 + \sin(2\pi u)), \\ M(u) &= A^2(u) - \overline{A}^2 = \sin(2\pi u) \cdot \text{diag}(1, -1, 1). \end{aligned}$$

Following Definition 5.1 with M replaced by $M(u)$, a (1, 1)-pseudo null space is given by

$$\mathcal{P}_{(1,1)}(M(u)) = \text{lin}\{\alpha_1 (1 \ 0 \ 0)' + \beta_1 (0 \ 1 \ 0)', \alpha_1 (0 \ 0 \ 1)' + \beta_1 (0 \ 1 \ 0)'\}, \tag{5.4}$$

where $\alpha_1^2 - \beta_1^2 = 0$ and $0 < u < 1/2$.

If \widehat{M} estimates M with the analogous estimators $\widehat{d}_0, \widehat{d}_\pm, \widehat{\lambda}_i, \widehat{\lambda}_{0,i}, \widehat{\lambda}_{\pm,i}, \widehat{s}_i, \widehat{s}_{0,i}, \widehat{s}_{\pm,i}$ of the respective quantities, we would similarly like to have the sample counterparts of (1, 1)-pseudo null spaces in (5.1). Definition 5.1 may not, however, extend directly to the sample quantities since \widehat{d}_\pm may not necessarily represent the actual number of positive/negative eigenvalues $\widehat{\lambda}_{\pm,i}$ and hence the condition (5.2) may not be satisfied. To deal with this possibility, we define the sample counterparts $\mathcal{P}_{(1,1)}(\widehat{M})$ in the same way as in (5.1) by using the quantities with the hats, except that \widehat{d}_+ and \widehat{d}_- are replaced by \widetilde{d}_+ and \widetilde{d}_- , which are defined as

$$\widetilde{d}_+ = \max\{r : r \leq \widehat{d}_+, 0 < \widehat{\lambda}_{p-r+1} \leq \dots \leq \widehat{\lambda}_p\} \tag{5.5}$$

and

$$\widetilde{d}_- = \max\{r : r \leq \widehat{d}_-, \widehat{\lambda}_1 \leq \dots \leq \widehat{\lambda}_r < 0\}, \tag{5.6}$$

where we assumed that the eigenvalues $\widehat{\lambda}_i$ appear in the non-decreasing order. That is, we define a sample (1, 1)-pseudo null space as

$$\mathcal{P}_{(1,1)}(\widehat{M}) = \text{lin}\{\widehat{s}_{0,1}, \dots, \widehat{s}_{0,\widehat{d}_0}, \alpha_i \widehat{s}_{+,p(i)} + \beta_i \widehat{s}_{-,n(i)}, i = 1, \dots, \min(\widetilde{d}_+, \widetilde{d}_-)\}, \tag{5.7}$$

where $p(i) \in \{1, \dots, \widetilde{d}_+\}$ are different across i , $n(i) \in \{1, \dots, \widetilde{d}_-\}$ are different across i and

$$\alpha_i^2 \widehat{\lambda}_{+,p(i)} + \beta_i^2 \widehat{\lambda}_{-,n(i)} = 0, \quad i = 1, \dots, \min(\widetilde{d}_+, \widetilde{d}_-). \tag{5.8}$$

Replacing \widehat{M} above by $\widehat{M}(u)$ from (4.1) for $u \in (0, 1)$, estimators $\widehat{\mathcal{B}}_{(1,1)}(u) = \mathcal{P}_{(1,1)}(\widehat{M}(u))$ can be defined for (1, 1)-local stationary subspaces. Techniques to cluster these (1, 1)-local stationary subspaces are discussed next.

5.2. Clustering (1, 1)-local stationary subspaces

According to (3.10), there is a stationary subspace \mathcal{B}_1 for a VC model if there is at least one identical stationary subspace for all $u \in (0, 1)$. In Section 5.1, we defined (1, 1)-local stationary subspaces $\widehat{\mathcal{B}}_{(1,1)}(u)$ whose number can already be quite large for a fixed u ; see Remark 5.1. A natural possibility in defining a candidate for a stationary subspace \mathcal{B}_1 is to consider all (1, 1)-local stationary subspaces across u 's and select a subspace representing their ‘‘majority’’ in some suitable sense. In light of this observation, using clustering seems natural and this approach is pursued here on the estimated (1, 1)-local stationary subspaces.

More specifically, we discuss a graph-based approach to clustering the many (1, 1)-local stationary subspaces, or equivalently, the many (1,1)-pseudo null spaces using distances that are computed based on canonical angles between spaces. In addition, a method to select one (1, 1)-local stationary subspace out of the many is discussed and this is considered as our estimate of a stationary subspace.

As in Section 5.1, we let $\widehat{\mathcal{B}}_{(1,1)}(u) = \mathcal{P}_{(1,1)}(\widehat{M}(u))$ denote a sample (1, 1)-local stationary subspace of the matrix $\widehat{M}(u)$ from (4.1). Let $\widehat{\mathcal{B}}_{(1,1)} = \bigcup_u \widehat{\mathcal{B}}_{(1,1)}(u)$ be the union of all (1, 1)-local stationary subspaces over u . In practice, the union \bigcup_u is replaced by a set of discrete points $\{u_1, u_2, \dots, u_{n_u}\}$ in $(0, 1)$.

Consider a graph $G = (V, E)$ where every vertex $v \in V$ corresponds to a (1, 1)-local stationary subspace in $\widehat{\mathcal{B}}_{(1,1)}$. The adjacency matrix E will be defined in terms of a distance between (1, 1)-local stationary subspaces using canonical angles. Let \mathcal{B}_1 and \mathcal{B}_2 be two (1, 1)-local stationary subspaces in $\widehat{\mathcal{B}}_{(1,1)}$ of dimensions d_1 and d_2 , respectively. Letting $d = \min(d_1, d_2)$, the canonical angles computed

between these two spaces are given by $\theta_1 \leq \theta_2 \leq \dots \leq \theta_d$, where

$$\theta_1 = \min_{x_1 \in \mathcal{B}_1, y_1 \in \mathcal{B}_2} \arccos\left(\frac{x_1 y_1}{\|x_1\| \cdot \|y_1\|}\right), \tag{5.9}$$

$$\theta_j = \min_{\substack{x_j \in \mathcal{B}_1, y_j \in \mathcal{B}_2; \\ x_j \perp x_1, x_2, \dots, x_{j-1}, y_j \perp y_1, y_2, \dots, y_{j-1}}} \arccos\left(\frac{x_j y_j}{\|x_j\| \cdot \|y_j\|}\right), \quad j = 2, 3, \dots, d. \tag{5.10}$$

The vectors x_1, x_2, \dots, x_d and y_1, y_2, \dots, y_d are called canonical vectors. We measure the distance between spaces \mathcal{B}_1 and \mathcal{B}_2 as $\max_{1 \leq i \leq d} \theta_i$, and define the adjacency matrix $E = (e_{i,j})$ for $i, j = 1, 2, \dots, |V|$, as

$$e_{i,j} = 1, \quad \text{if } \theta(v_i, v_j) < \theta^*, \tag{5.11}$$

where $\theta(v_i, v_j)$ is the maximum canonical angle between the $(1, 1)$ -local stationary subspaces corresponding to vertices v_i and v_j , and θ^* is a threshold. The choice of θ^* dictates the number of clusters estimated with more being formed for lower values of θ^* . In our numerical work, we set $\theta^* = 20^\circ$.

Finally, in order to obtain the clusters of vertices in the graph G , we utilize the Walktrap algorithm of Pons and Latapy [18]. Here, a transition probability matrix $P = (p_{ij})$ is constructed with $p_{ij} = \frac{A_{ij}}{d(i)}$ where $A = (A_{ij})$ denotes the adjacency matrix of G and $d(i)$ denotes the degree of vertex v_i . A random walk process defined on this graph G is based on the powers of the matrix P , that is, the probability of moving from vertex v_i to v_j through a random walk of length t is given by P_{ij}^t . The closeness of vertices in the graph is measured by these probabilities from the observation that if two vertices v_i and v_j are in the same cluster, P_{ij}^t must be high.

Let C_1, C_2, \dots, C_L be the L clusters of vertices produced by the Walktrap algorithm that have a size of at least 3 vertices. We first obtain the centers $\{c_1, c_2, \dots, c_L\}$ of these L clusters by computing the sine of the maximum canonical angle,

$$c_l = \operatorname{argmin}_{v \in C_l} \sum_{v' \neq v} \sin(\theta(v', v)), \quad l = 1, 2, \dots, L. \tag{5.12}$$

The $(1, 1)$ -local stationary subspaces corresponding to the L cluster centers are considered as the candidate stationary subspaces returned by our method. Additionally, in order to select a single $(1, 1)$ -local stationary subspace (our stationary subspace estimator) out of these L representative subspaces, we assess the ‘‘denseness’’ of each cluster through

$$T(C_l) = \frac{\frac{1}{|C_l|} \sum_{v \in C_l} \log(d(v))}{\frac{1}{|C_l|} \sum_{v \in C_l} \left(\frac{1}{|C_l| - 1} \sum_{v' \in C_l, v' \neq v} \sin(\theta(v', v)) \right)}, \quad l = 1, 2, \dots, L, \tag{5.13}$$

where $d(v)$ denotes the degree of vertex v within cluster C_l . We then identify the cluster with maximum $T(\cdot)$ among the L clusters and select the final $(1, 1)$ -local stationary subspace corresponding to the center of that cluster. That is, we select our stationary subspace estimate as the $(1, 1)$ -local stationary subspace corresponding to the cluster center c_s in cluster C_s , where

$$s = \operatorname{argmax}_l T(C_l). \tag{5.14}$$

The procedure to select our stationary subspace is summarized in Algorithm 2.

Algorithm 2: Finding stationary subspace

Output: Estimate $\widehat{\mathcal{B}}_1$ of the stationary subspace \mathcal{B}_1 from Definition 2.1.

Input: p -variate time series data $X_t, t = 1, 2, \dots, T$, the sub-interval $\mathcal{H} \subset (0, 1)$ given by a discretization $U = \{u_1 \leq u_2 \leq \dots \leq u_{n_u}\}$ and bandwidth h .

- 1: Using \mathcal{H}, U, h , find estimates \widehat{d} and $\widehat{d}_-, \widehat{d}_+$ from Algorithm 1. For every $u \in U$, compute the eigenvalues $(\widehat{\lambda}_{0,i}(u), \widehat{\lambda}_{+,i}(u), \widehat{\lambda}_{-,i}(u))$, the eigenvectors $(\widehat{s}_{0,i}(u), \widehat{s}_{+,i}(u), \widehat{s}_{-,i}(u))$ of $\widehat{M}(u)$, the estimates $\widehat{d}_+(u), \widehat{d}_-(u)$ and $\mathcal{P}_{(1,1)}(\widehat{M}(u))$ in (5.5)–(5.8).
- 2: Define a graph $G = (V, E)$, where every vertex $v \in V$ corresponds to a $(1, 1)$ -local stationary subspace in $\widehat{\mathcal{P}}_{(1,1)} = \bigcup_u \mathcal{P}_{(1,1)}(\widehat{M}(u))$ and the adjacency matrix $E = (e_{i,j})$ is given by (5.11).
- 3: Find the optimal number of clusters L and the resulting clusters C_1, C_2, \dots, C_L using the Walktrap algorithm in Section 4 of Pons and Latapy [18].
- 4: Find the L cluster centers $\{c_1, c_2, \dots, c_L\}$ using (5.12) and also compute the “denseness” of each cluster $T(C_l)$ using (5.13).
- 5: Output stationary subspace estimate $\widehat{\mathcal{B}}_1$ as the $(1, 1)$ -local stationary subspace corresponding to the cluster center c_s in cluster C_s , where $s = \operatorname{argmax}_l T(C_l)$.

Example 5.2. To illustrate the above technique, we consider the VC model from Example 5.1. The stationary subspace dimension for this model is $d = 1$ and $(1, 1)$ -local stationary subspaces are given by $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0), (0, \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}), (\frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}}, 0)$ and $(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0)$ for all u 's. We generated one realization of the series X_t based on this model and obtained all the estimated $(1, 1)$ -local stationary subspaces across a set of points $\{0.04, 0.08, 0.12, \dots, 0.96\}$. Figure 3 depicts a 3D plot that includes the population $(1, 1)$ -local stationary subspaces (solid circles), the estimated $(1, 1)$ -local stationary subspaces (crosses), the cluster centers c_l for $l = 1, 2, 3, 4$ of the 4 clusters (open circles) and the selected $(1, 1)$ -local stationary subspace based on (5.14) (solid square, marked as VC Final). Finally, the estimated stationary subspace from DSSA (Sundararajan and Pourahmadi [22]) is plotted along with the other spaces (diamond).

In Table 1, we list the cluster centers of the 4 main clusters in Figure 3, the sizes of those 4 clusters and the proportions of $u \in \{0.04, 0.08, 0.12, \dots, 0.96\}$ with points in the respective clusters. Observe that the $(1, 1)$ -local stationary subspace selected as the stationary subspace based on (5.14) lies in the most “dense” cluster and the DSSA stationary subspace also lies in the same cluster. Here, the 4 biggest clusters formed by the method comprise roughly 84% of the total number of $(1, 1)$ -local stationary subspaces. The $(1, 1)$ -local stationary subspace selected as the stationary subspace and identified as the solid square point in Figure 3 lies in the biggest cluster that contains roughly 24% of the $(1, 1)$ -local stationary subspaces. This subspace also lies in the cluster with the highest proportion of u 's with points in that cluster.

6. Simulation study

In Section 6.1, we evaluate the empirical performance of the proposed method in estimating the dimension of a stationary subspace for several VC models. In Section 6.2, we assess the ability of the proposed method in estimating a stationary subspace using a few discrepancy measures.

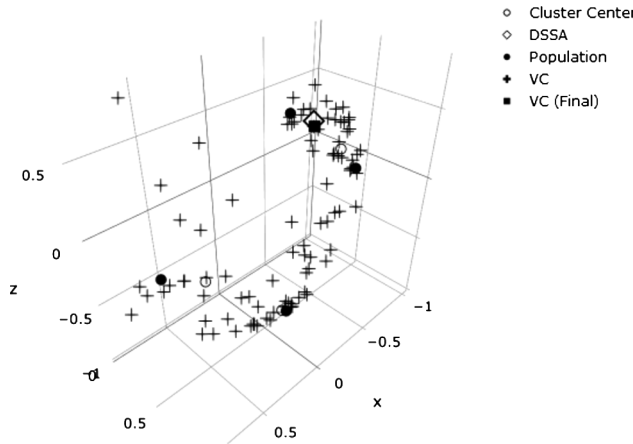


Figure 3. Local and global stationary subspaces in Example 5.2: 3D plot of the population (1, 1)-local stationary subspaces (solid circles), the estimated (1, 1)-local stationary subspaces (crosses), the cluster centers of the 4 clusters (open circles), the stationary subspace based on (5.14) (solid square, marked as VC Final), the estimated stationary subspace from DSSA (diamond).

6.1. Dimension estimation comparison

We first consider a few VC models (1.2), characterized through $A(u)$, for which the dimensions $d_0(u) = d_0$, $\min(d_+(u), d_-(u)) = \min(d_+, d_-)$ and $d(u) \equiv d$ do not depend on u (and neither do the local stationary subspaces). The model matrices $A(u)$, the respective matrices in (2.2) and the dimensions $p, d_0, \min(d_+, d_-)$ are:

Model 1: $p = 3, d_0 = 0, \min(d_+, d_-) = 1, d = 1,$

$$A^2(u) = \text{diag}(2 + 0.5 \sin(2\pi u), 3 - \sin(2\pi u), 1.5 + \sin(2\pi u)),$$

$$M(u) = \text{diag}(0.5 \sin(2\pi u), -\sin(2\pi u), \sin(2\pi u)).$$

Model 2: $p = 4, d_0 = 1, \min(d_+, d_-) = 1, d = 2,$

$$A^2(u) = \begin{pmatrix} e^1 & 1 & 0 & 0 \\ 1 & 2 + \sin(2\pi u) & 0 & 0 \\ 0 & 0 & 3 - 2u & 0.5 \\ 0 & 0 & 0.5 & 3 - \sin(2\pi u) \end{pmatrix},$$

Table 1. The 4 cluster centers, cluster sizes (in percentage), and the proportions of u 's for the 4 biggest clusters in Figure 3. The selected (1, 1)-local stationary subspace identified as the solid square marked as VC Final in Figure 3 corresponds to $(-0.114, 0.778, 0.618)$

Cluster center	Cluster size (in %)	Proportion of u 's
(0.661, 0.750, -0.003)	19.79	79.16
(0.029, -0.641, 0.767)	23.95	91.67
(-0.114, 0.778, 0.618)	23.95	95.83
(0.938, -0.346, -0.018)	17.70	70.85

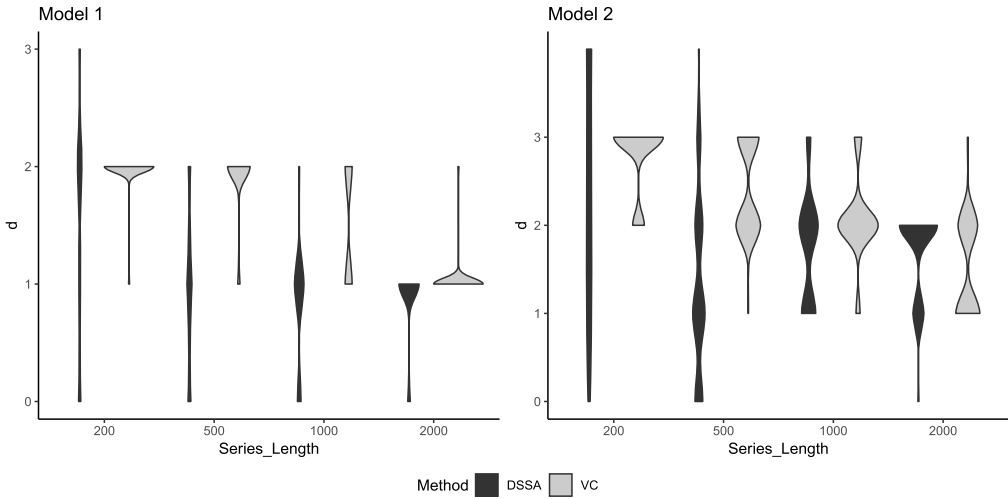


Figure 4. Models 1–2: Violin plots of the estimates of d for the indicated sample sizes for the two competing methods: DSSA and VC (proposed method).

$$M(u) = \text{diag}(0, \sin(2\pi u), 1 - 2u, -\sin(2\pi u)).$$

We suppose that the models above are Gaussian with i.i.d. $\mathcal{N}(0, I_p)$ vectors Y_t . In the simulations, we take $T \in \{200, 500, 1000, 2000\}$ as the sample sizes. In estimating $A^2(u)$ in (4.1), bandwidth choices h ranging from $T^{-0.1}$ to $T^{-0.5}$ were attempted and the best results were obtained for $h \in (T^{-0.3}, T^{-0.4})$. Here, we also attempted a leave-one-out cross-validation estimation of h and we found that the optimal bandwidth was close to $h = T^{-0.35}$. Hence, we take this value for h and present the results for this choice. We focus on testing for $\mathcal{H} = (0, 1)$ only and use 100 Monte Carlo replications.

We compare the performance of the proposed dimension estimator \hat{d} given in (4.21) with the sequential estimation procedure provided in Section 2.2.5 of Sundararajan and Pourahmadi [22]. The method found in the latter work will be referred to as DSSA and the proposed method will be denoted as VC. The estimation results for the two methods and the two considered models are presented through violin plots in Figure 4. Violin plots are intended as a somewhat qualitative visualization of the results – perhaps slightly more informative histograms for the results can be found in Sundararajan et al. [21]. The plots show that estimation improves with increasing sample size T since the widths of the violin plots tend to increase at true values. For Model 1, the VC method performs better than the competing DSSA method in detecting the true dimension d whereas the same is also true for Model 2 except for the largest sample size $T = 2000$.

We now turn to VC models whose dimensions depend on u 's. We consider the following models:

Model 3: $p = 3$, $d_0(u) = 2$ and $d(u) = 2$ if $u < 0.5$ and $d_0(u) = 1$ and $d(u) = 1$ if $u > 0.5$,

$$A^2(u) = 1_{(0,0.5)}(u) \text{diag}(2 + \sin(2\pi u), 2.0901, 3) + 1_{(0.5,1)}(u) \text{diag}(2 + \sin(2\pi u), 4 + 3 \sin(2\pi u), 3).$$

Model 4: $p = 3$, $d_0(u) = 3$ and $d(u) = 3$ if $u < 0.5$ and $d_0(u) = 2$ and $d(u) = 2$ if $u > 0.5$,

$$A^2(u) = 1_{(0,0.5)}(u) \begin{pmatrix} 4 & 0.5 & 0 \\ 0.5 & 3.125 & 0 \\ 0 & 0 & 1 \end{pmatrix} + 1_{(0.5,1)}(u) \begin{pmatrix} 4 & 0.5 & 0 \\ 0.5 & 3u^2 + 2u & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

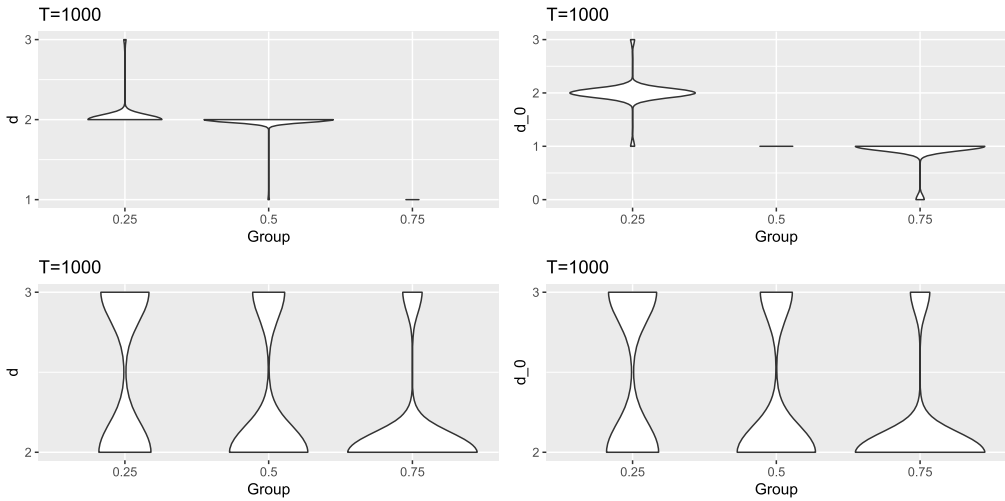


Figure 5. Top panel (Model 3): Violin plots of the estimates of $d(u)$ (left) and $d_0(u)$ (right) obtained over $(0, 0.5)$, $(0, 1)$, $(0.5, 1)$ depicted at $u = 0.25$, $u = 0.5$ and $u = 0.75$, respectively. Bottom panel (Model 4): Similar violin plots for Model 4.

The matrix entries 2.0901 in Model 3 and 3.125 in Model 4 ensure smoothness of $A^2(u)$ at $u = 0.5$.

We report the estimation results for the two models in Figure 5. The structure of the plots is similar to that of Figure 2. That is, the estimates over $\mathcal{H} = (0, 1)$ are depicted at $u = 0.5$, over $\mathcal{H} = (0, 0.5)$ at $u = 0.25$ and those over $\mathcal{H} = (0.5, 1)$ at $u = 0.75$. The only difference here is that the results are reported over 100 realizations in the form of violin plots.

Several observations are in place regarding Figures 5. For Model 3, the estimates of $d_0(u)$ and $d(u)$ in the violin plots closely align with the true dimension values. In comparison, for Model 4, the estimates are less closely aligned to the true dimension values but are sensitive to the choice of the subinterval, $(0, 0.5)$ or $(0.5, 1)$, in the direction of the true dimension values.

Finally, we comment on the running time of the two competing methods, VC and DSSA, in estimating the dimension d of the stationary subspace. The running time is recorded for 100 replications of Models 1 and 2, and sample size $T = 500$. For Model 1, the running time for the VC method (Algorithm 1) in estimating the subspace dimension d had a mean of 0.545 seconds and a standard deviation of 0.021. Similarly for DSSA and Model 1, the mean was 13.922 seconds with a standard deviation of 0.77. For Model 2, the VC method had a mean running time of 0.412 seconds and a standard deviation of 0.01. Similarly for DSSA and Model 2, the mean was 66.181 seconds with a standard deviation of 5.29 seconds. It was observed that VC method is faster than DSSA for both models by at least an order of magnitude on average.

6.2. Subspace estimation comparison

We compare here the performance of the proposed method in estimating a subspace (from Section 5.2) to DSSA of Sundararajan and Pourahmadi [22], in terms of three discrepancy measures. The first measure concerns departure from stationarity based on the sizes of the DFT covariances as given in Eq. (9) of Sundararajan and Pourahmadi [22]. More precisely, for an estimated stationary subspace

process, we set

$$D_1(\widehat{B}_1) = \sum_{r=1}^m \|\Re(\widehat{\Gamma}_r)\|_F^2 + \|\Im(\widehat{\Gamma}_r)\|_F^2, \tag{6.1}$$

where $\|A\|_F$ denotes the Frobenius norm of a matrix $A \in \mathbb{R}^{d \times d}$, $\Re(\cdot)$ and $\Im(\cdot)$ denote the entrywise real and imaginary parts respectively, and $\widehat{\Gamma}_r$ is the $d \times d$ lag- r DFT sample autocovariance matrix given by

$$\widehat{\Gamma}_r = \frac{1}{T} \sum_{k=1}^T J(\omega_k) J(\omega_{k+r})^*, \tag{6.2}$$

with $\omega_k = 2\pi k/T$ referring to a Fourier frequency, $J(\cdot)$ being the discrete Fourier transform (DFT) of the d -variate stationary subspace process and $J(\cdot)^*$ denoting the complex conjugate transpose. The number of DFT covariance lags m in (6.1) is fixed to 3.

The second measure is based on the relation (2.3). For any candidate subspace \widehat{B}_1 , we set

$$D_2(\widehat{B}_1) = \sum_k \|\widehat{B}_1'(A^2(u_k) - \overline{A^2})\widehat{B}_1\|_F = \sum_k \|\widehat{B}_1' M(u_k) \widehat{B}_1\|_F.$$

The last measure is based on canonical angles computed between the population and estimated subspaces B_1 and \widehat{B}_1 that are spanned by the columns of the $p \times d$ matrices B_1 and \widehat{B}_1 , respectively. As in (5.9), let $\theta_1 \leq \theta_2 \leq \dots \leq \theta_d$ be the d canonical angles between the spaces B_1 and \widehat{B}_1 . Then, set

$$D_3(\widehat{B}_1) = \left(\sum_{j=1}^d \sin^2(\theta_j) \right)^{1/2}. \tag{6.3}$$

In the simulations here, we consider the same two models, Models 1–2, as in Section 6.1. We first present estimation results for the measure D_3 in Table 2. For our method, labeled as VC in the table, the stationary subspace estimate is taken as discussed in Section 5.2. At the population level, the corresponding stationary subspace is also selected using the same technique. In the case when such multiple subspaces are available at the population level, we compute the distance D_3 to all of them and then take the minimum. For the DSSA method, we take the subspace estimate as defined in Sundararajan and Pourahmadi [22].

As seen from the table, the VC method generally performs better than DSSA in yielding smaller average distances over 100 replications, though possibly not statistically significant if standard deviations (in parentheses) are taken into account. As noted below, the VC method is computationally much less intensive than DSSA.

We now turn to the measures D_1 and D_2 . Here, we shall examine the VC and DSSA methods through these measures from a different perspective. The relevant plots can be found in Figure 6 for Model 1, associated with the indicated sample sizes. In the left two plots of Figure 6, the horizontal solid circles in the plots represent the measure $D_1(\widehat{B}_1)$ for the DSSA estimate \widehat{B}_1 , averaged over multiple realizations. The other two curves correspond to the measure $D_1(\widehat{B}_1(u))$ computed for the estimates of $\widehat{B}_1(u)$ of (1, 1)-local stationary subspaces from Section 5, either averaged over multiple realizations (triangles) or with the minimum value taken (squares). The interpretation of the right two plots of Figure 6 is analogous but for the measure D_2 . The plots show that VC method performs better than DSSA even “locally” for most values of u under measure D_2 whereas DSSA, not surprisingly, performs better under measure D_1 .

Table 2. Models 1–2: The canonical angle based measure $D_3(\hat{B}_1)$, between the population and estimated subspace. The empirical standard errors of the respective measures are provided in parentheses

T	Model 1		Model 2	
	DSSA	VC	DSSA	VC
200	0.3277 (0.1429)	0.3489 (0.1479)	0.3633 (0.1693)	0.3489 (0.1401)
500	0.2976 (0.1399)	0.2878 (0.1288)	0.3575 (0.1657)	0.3301 (0.1329)
1000	0.2757 (0.1306)	0.2968 (0.1301)	0.3241 (0.1453)	0.2649 (0.1228)
2000	0.2678 (0.1323)	0.2374 (0.1311)	0.2461 (0.1347)	0.1931 (0.1202)

Finally, as in Section 6.1, we comment on the running time of the two competing methods, VC and DSSA, in obtaining the stationary subspace. The running time is recorded for 100 replications of Models 1 and 2 are reported for sample size $T = 500$. For Model 1, the running time for the VC method (Algorithm 2) in estimating the stationary subspace had a mean of 0.718 seconds and a standard deviation of 0.036. Similarly for DSSA and Model 1, the mean was 5.115 seconds with a standard deviation of 0.588. For Model 2, the VC method had a mean running time of 0.807 seconds and standard deviation of 0.033. Similarly for DSSA and Model 2, the mean was 6.285 seconds with a standard deviation of 0.724. Here again, the VC method is computationally much more efficient than DSSA for both models.

7. Application to BCI and EEG data

Brain-Computer Interface (BCI) aims at connecting the human brain and the computer in a non-invasive manner. During the BCI study used here, individuals are asked to imagine movements with their left and right hands and these are referred to as trials. The trials are interspersed with break periods. The multivariate EEG brain signal is recorded during the entire course of the experiment and the objective is to associate the movements imagined by the individuals with the corresponding EEG signal. Note that the EEG signal is recorded through different channels (locations on the scalp) and each channel constitutes a component of the multivariate signal.

EEG signals measuring brain activity have been treated as a multivariate nonstationary time series. In Ombao et al. [17], EEG recorded during an epileptic seizure is considered and the nonstationary

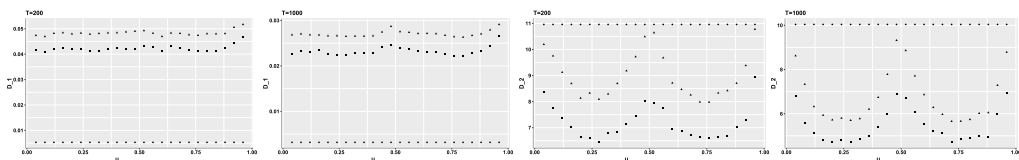


Figure 6. Model 1 – Left 2 plots: Plot of $D_1(\hat{B}_1(u))$ against u for the competing methods DSSA and VC and indicated sample sizes. VC (avg.) in triangles, VC (min.) in squares and DSSA in solid circles. Right 2 plots: Analogous plot but with measure $D_2(\hat{B}_1(u))$.

features of this signal are used to identify changes in electrical activity due to seizure. In von Bünau et al. [24,25] and Sundararajan et al. [20], the nonstationarity is regarded as the noise or background activity in the brain signal (Kaplan et al. [12]) and removing this nonstationarity was seen to improve classification accuracy in brain related experiments. The key observation made in the latter references was that nonstationary sources in the brain signal were associated with variations in the mental state that are unrelated to the motor imagery task. For instance, alpha oscillations in the brain lead to nonstationarity in the EEG signal. These oscillations are due to physical movement, blinking or fatigue in individuals and deemed unrelated to the experimental task. Hence, the aim of SSA is to eliminate the nonstationary sources in the signal and work with only the stationary sources.

We study the classification performance of the proposed VC method using the BCI Competition IV¹ dataset II in Naeem et al. [16]. It consists of EEG signals from 9 subjects performing 4 different motor imagery tasks: 1-left hand, 2-right hand, 3-feet and 4-tongue. We analyze the EEG signals only from classes 1 and 2 and treat the problem as a two-group classification. The continuous signal was sampled at discrete time steps at the sampling rate of 250 Hz where 1 second corresponds to 250 observations on the digital signal scale. The signal was recorded through 22 electrodes over the course of the experiment and the signal was band-pass filtered as in Lotte and Guan [14]. The experiment involved 144 trials for each subject wherein 72 trials belonged to Class 1 (left-hand side) and the other 72 to Class 2 (right-hand side). Every trial is followed by an adequate resting period for the subject before the start of the next one. In each trial, we use the observations between 0.5 seconds to 2.5 seconds after the cue instructing the subject to perform the motor imagery task. More precisely, for trial j where $j = 1, 2, \dots, 144$, this interval comprises of 500 observations on the digital signal scale, denoted by $X_t^{(j)}$, $t = 1, 2, \dots, 500$.

We first restrict attention to 5 EEG electrodes and treat the input signal to have dimension $p = 5$. These are 5 locations that can be viewed as representatives of the different regions on the brain, namely, Frontal (Fz), Pre-Frontal (Pz) and Cortical (C3, C4, Cz).² We use the VC method to obtain a d dimensional stationary process where $d < p$. For every trial $j = 1, 2, \dots, 144$, we have $\{X_t^{(j)}\} \in \mathbb{R}^p$ as the observed multivariate signal and $\{Y_t^{(j)}\} \in \mathbb{R}^d$ as the stationary transformation.

We first report on the estimated pseudo dimension d for the observed signals for the 9 subjects in this study labeled S1, S2, ..., S9. For subjects S3, S5 and S8, the percentage of times the candidate dimensions ($d = 0, 1, 2, 3, 4$) were estimated by the 2 competing methods, DSSA or VC, out of the 144 trials is provided in the left panel of Figure 7. This plot also includes the estimates of d over the first and second halves of the data denoted as VC (First) and VC (Second). It is noted that DSSA always provides a lower estimate of d than the VC method. Similar plots for all 9 subjects can be found Sundararajan et al. [21].

Given the d -variate stationary processes $Y_t^{(j)}$, we aim to find differences between the two classes (1 and 2) based on the covariance structure. For a given subject, this is achieved by computing the average spectral density matrices for the two classes over the Fourier frequencies:

$$\overline{g^i}(\omega_k) = \frac{1}{n_i} \sum_{j \in \text{Class } i} g_j(\omega_k), \quad i = 1, 2, \quad (7.1)$$

where $g_j(\omega_k)$ is the estimated $d \times d$ spectral matrix for trial j using observations $\{Y_t^{(j)}\}$, $n_i = 72$, for $i = 1, 2$ and $\omega_k = \frac{2\pi k}{500}$, $k = 1, 2, \dots, 500$, are the Fourier frequencies.

¹ See <http://www.bbci.de/competition/iv/>.

² See http://www.bbci.de/competition/iv/desc_2a.pdf for additional details on the dataset.

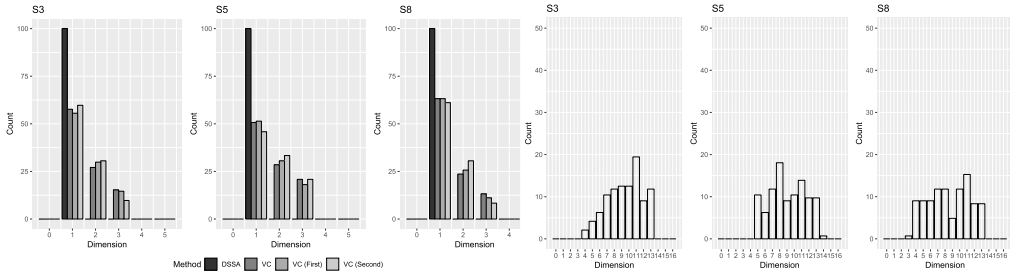


Figure 7. Left: ($p = 5$) Histogram of the dimension estimates d by the two competing methods based on the 144 trials. Right: ($p = 22$) Histogram of the stationary subspace dimension estimates for the VC method based on the 144 trials.

Table 3. Left: Out-of-sample classification accuracy (in %) for the 3 subjects S3, S5 and S8 for the two indicated methods with $p = 5$. Right: Out-of-sample classification accuracy (in %) for 3 subjects S3, S5 and S8 corresponding to $d = 7, 9, 11,$ and 13 for the VC method with $p = 22$

d		S3	S5	S8	d	S3	S5	S8	
1	DSSA	56.25	49.03	51.11	7	DSSA	70	60.48	68.30
	VC	50.69	52.08	46.15		VC	70.83	73.61	68.53
2	DSSA	54.86	54.16	56.45	9	DSSA	77.90	65.50	70.38
	VC	48.61	55.56	45.05		VC	72.22	81.25	71.32
3	DSSA	59.02	59.33	64.39	11	DSSA	72.92	69.58	71.33
	VC	47.22	57.63	54.54		VC	74.30	88.19	81.25
4	DSSA	56.25	62.50	66.28	13	DSSA	85.10	70.83	78.38
	VC	65.97	64.58	59.44		VC	85.41	89.58	84.72

In order to train a classifier, for every trial $j = 1, 2, \dots, 144$, a distance vector $p_{j,AB} = (p_{0,j,AB}, p_{1,j,AB})$ is computed, where

$$p_{i,j,AB} = \frac{1}{250} \sum_{k=1}^{250} \|g_j(\omega_k) - \bar{g}^i(\omega_k)\|_F^2, \quad i = 1, 2, \tag{7.2}$$

and $\|\cdot\|_F$ is the Frobenius norm of a matrix. It measures the distance to the center of each of the two classes. This distance measure serves as our two-dimensional feature vector to be used in constructing a logistic regression classifier and assessing its out-of-sample classification accuracy.

Table 3 (left) shows the out-of-sample classification accuracies for three subjects corresponding to $d = 1, 2, 3, 4$. A similar table containing results for all 9 subjects can be found in Sundararajan et al. [21]. The accuracy rates reflect a comparable average performance in the two competing methods. Finally, the accuracy rate increases as the pseudo dimension d increases from 1 to 4, a phenomenon witnessed and discussed in von Büнау et al. [24] and Sundararajan and Pourahmadi [22]. In dealing with brain signals from healthy individuals in this experiment, the nonstationarity is believed to be caused by artifacts such as as fatigue, physical movement, blinking. Hence, more stationary sources means greater elimination of nonstationary sources in the signal that are unrelated to the experimental task at hand.

The results in Figure 7 (right) and Table 3 (right) are analogous to those on their respective left panels but taking the input signal to have dimension $p = 22$, that is, without restricting attention to 5 EEG electrodes. For the out-of-sample classification accuracies, the values $d = 7, 9, 11, 13$ are considered. The results in Table 3 (right) indicate a better performance of VC in comparison to results on the same dataset (with $p = 22$) in Sundararajan and Pourahmadi [22].

It must be noted that the results in Table 3 are based on an estimated stationary subspace for a prefixed dimension d . The estimates of the dimensions of the stationary subspaces from (4.21), namely $\widehat{d}_0, \widehat{d}_+$ and \widehat{d}_- , could potentially be different across different trials. For each trial, we obtained these estimates and compared them with the fixed d and investigated results for several different combinations of $\widehat{d}_0, \widehat{d}_+$ and \widehat{d}_- that results in $d = \widehat{d}_0 + \min(\widehat{d}_+, \widehat{d}_-)$. Having witnessed very similar results for the various combinations, we only present, in Table 3, the case wherein $d = \widehat{d}_0$.

8. Concluding remarks

Our goal in this work is to (i) study existence of linear combinations of components of a multivariate nonstationary process which are stationary, and (ii) find the number of such stationary linear combinations. The true nature of the problem and richness of its solution present themselves naturally when the general dependence setup in Sundararajan and Pourahmadi [22] is abandoned in favor of heterogenous independent observations. In this simplified setup (von Büнау et al. [24]), solution of the problem reduces to the study of inertia or signs of the eigenvalues and the corresponding eigenvectors of certain symmetric time-varying matrices constructed from varying covariance or heterogeneity of the vector observations. This enables us to provide a direct linear-algebraic method to construct stationary subspaces which outperforms the earlier computationally more expensive optimization-based SSA solutions.

Several directions related to this work could be explored in the future. The developed framework involving pseudo spaces and dimensions is general enough to apply to zero mean locally stationary processes, when working with their time-frequency spectra. More specifically, let $X_{t,T}$ be a p -vector locally stationary time series (Dahlhaus [4]) represented in a standard integral form as

$$X_{t,T} = \int_{-\pi}^{\pi} A_{t,T}(\omega)e^{it\omega} dZ(\omega), \quad t = 1, \dots, T,$$

where $A_{t,T}(\omega)$ is a $p \times p$ complex-valued matrix and $dZ(\omega)$ is a suitable complex-valued random measure. With the common assumption that $A_{t,T}(\omega)$ equals approximately $A(t/T, \omega)$ for a suitable matrix-valued function $A(u, \omega)$, the time-varying spectral matrix of the series can be written as $f(u, \omega) = A(u, \omega)A(u, \omega)^*$. The dimension d of a stationary subspace can then be defined as the largest integer in $\{0, 1, \dots, p\}$ for which there is a $d \times p$ matrix B_1 such that

$$B_1 f(u, \omega)B_1' = G(\omega), \quad \forall u \in (0, 1), \omega \in (-\pi, \pi), \tag{8.1}$$

where $G(\omega)$ does not depend on u . The column space of B_1 being the stationary subspace. The formulation in (8.1) is the same as in (2.1) but with the additional requirement for it hold for $\omega \in (-\pi, \pi)$. Similarly, the condition (8.1) is equivalent to

$$B_1 F(u, \omega)B_1' = 0, \quad \forall u \in (0, 1), \omega \in (-\pi, \pi), \tag{8.2}$$

where $F(u, \omega) = f(u, \omega) - \bar{f}(\omega)$ and $\bar{f}(\omega) = \int_0^1 f(u, \omega) du$. An estimator $\widehat{F}(u, \omega)$ of $F(u, \omega)$ could then be defined naturally from the time-varying periodogram. To proceed as in this work, one requires an asymptotic result of the type (4.2) but for $\widehat{F}(u, \omega)$, which would suggest a local test statistic

$\widehat{\xi}_r(u, \omega)$ to use, and then also an asymptotic result for $\int_{\mathcal{H}_1 \times \mathcal{H}_2} \widehat{\xi}_r(u, \omega) du d\omega$, analogous to that in Proposition 4.1.

Another future possibility is to try to combine the models of this work and Düker et al. [9], so that both mean and the covariance are allowed to vary in time. Yet another direction is to explore connections to multivariate stochastic volatility models that are concerned with modeling the changing covariance across time.

Appendix A: Proofs

The appendix concerns the technical aspects of this work, including various assumptions and proofs of some of the stated results.

A.1. Proof of Proposition 3.1

Proof. By the Poincare Separation theorem (see, e.g., Magnus and Neudecker [15]),

$$\lambda_i \leq \mu_i \leq \lambda_{n-d_1+i}, \quad i = 1, \dots, d_1,$$

where $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_p$ are the ordered eigenvalues of M and $\mu_1 \leq \dots \leq \mu_{d_1}$ are the ordered eigenvalues of $C_1' M C_1$ for any $p \times d_1$ matrix C_1 such that $C_1' C_1 = I_{d_1}$. Taking C_1 as in (3.1) with $d_1 = d(M)$, we have $C_1' M C_1 = 0$, and hence

$$\lambda_i \leq 0 \leq \lambda_{n-d_1+i}, \quad i = 1, \dots, d_1.$$

This shows that $d(M) \leq d_0 + \min(d_+, d_-)$.

To prove the reverse inequality $d_0 + \min(d_+, d_-) \leq d(M)$, let $d_* = d_0 + \min(d_+, d_-)$. The rest of the proof constructs a $p \times d_*$ matrix C_1 such that $C_1' M C_1 = 0$ and $C_1' C_1 = I_{d_*}$, which yields the desired inequality. By the Schur decomposition (see, e.g., Magnus and Neudecker [15]),

$$S' M S = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_p), \tag{A.1}$$

where the columns s_1, \dots, s_p of S are orthonormal and represent the eigenvectors associated with the negative, positive and zero eigenvalues of M . We now need to separate the eigenvectors into those associated with the eigenvalues $\lambda_1, \dots, \lambda_p$ of M . Let $s_{0,i}$ be the eigenvectors associated with the zero eigenvalues $\lambda_{0,i}$, $i = 1, \dots, d_0$, $s_{+,i}$ be the eigenvectors associated with the positive eigenvalues $\lambda_{+,i}$, $i = 1, \dots, d_+$, and $s_{-,i}$ be the eigenvectors associated with the negative eigenvalues $\lambda_{-,i}$, $i = 1, \dots, d_-$. Note by (A.1) that

$$s_i' s_j = \delta_{ij}, \quad s_i' M s_j = \delta_{ij} \lambda_i, \tag{A.2}$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.

For the zero eigenvalues and the corresponding eigenvectors, we have $M s_{0,i} = 0 \cdot s_{0,i}$ and hence

$$s_{0,i}' M s_{0,i} = 0, \quad i = 1, \dots, d_0. \tag{A.3}$$

Similarly, for the positive and negative eigenvalues, we have

$$\begin{aligned} s_{+,i}' M s_{+,i} &= \lambda_{+,i} > 0, \quad i = 1, \dots, d_+, \\ s_{-,i}' M s_{-,i} &= \lambda_{-,i} < 0, \quad i = 1, \dots, d_-. \end{aligned}$$

Let $d_{\pm} = \min(d_+, d_-)$. Then, for some $\alpha_i \in (0, 1)$ and $\beta_i = 1 - \alpha_i$, we have

$$(\alpha_i^{1/2} s_{+,i} + \beta_i^{1/2} s_{-,i})' M (\alpha_i^{1/2} s_{+,i} + \beta_i^{1/2} s_{-,i}) = \alpha_i (s'_{+,i} M s_{+,i}) + \beta_i (s'_{-,i} M s_{-,i}) = 0, \tag{A.4}$$

for $i = 1, \dots, d_{\pm}$. Set

$$s_{\pm,i} = \frac{\alpha_i^{1/2} s_{+,i} + \beta_i^{1/2} s_{-,i}}{\|\alpha_i^{1/2} s_{+,i} + \beta_i^{1/2} s_{-,i}\|_2}, \quad i = 1, \dots, d_{\pm}, \tag{A.5}$$

and define a $(d_0 + d_{\pm}) \times p = d_* \times p$ matrix C_1 as

$$C_1 = (s_{0,1} \ \cdots \ s_{0,d_0} \ s_{\pm,1} \ \cdots \ s_{\pm,d_{\pm}}).$$

By using (A.2), (A.3) and (A.4), we have $C_1' M C_1 = 0$. Since s_i are orthonormal and in view of (A.5), we also have $C_1' C_1 = I_{d_*}$. This shows that $d_0 + d_{\pm} \leq d(M)$ and concludes the proof. \square

A.2. Proof of Proposition 3.2

Proof. Any vector w can be expressed as a linear combination of the basis vectors $s_{0,i}, s_{+,i}, s_{-,i}$ as

$$w = w_0 + w_{k,\pm}$$

where $w_0 \in \mathcal{N}_0(M)$, $w_{k,\pm} \in \mathcal{N}_{\pm}(M)$, and $k = 1, 2, \dots, \min(d_-, d_+)$. The relation (3.5) follows since for any $i_1, i_2 = 1, 2, \dots, d_0$,

$$\begin{aligned} 0 &= w' M w = \sum_{i=1}^{d_0} \alpha_{0,i_1} \alpha_{0,i_2} s'_{0,i_1} M s_{0,i_2} + \sum_{i=1}^{d_+} \alpha_{k_1,+,i} \alpha_{k_2,+,i} s'_{+,i} M s_{+,i} + \sum_{i=1}^{d_-} \alpha_{k_1,-,i} \alpha_{k_2,-,i} s'_{-,i} M s_{-,i} \\ &= \sum_{i=1}^{d_+} \alpha_{k_1,+,i} \alpha_{k_2,+,i} \lambda_{+,i} + \sum_{i=1}^{d_-} \alpha_{k_1,-,i} \alpha_{k_2,-,i} \lambda_{-,i} \end{aligned}$$

for all k_1, k_2 . The converse statement follows similarly. \square

A.3. Assumptions and proof of Proposition 4.1

We use the following assumptions for Proposition 4.1, labeled according to the quantities they concern.

- (Y1) $Y_t, t = 1, \dots, T$, are i.i.d. random vectors with i.i.d. entries, $\mathbb{E}(Y_t) = 0$ and $\mathbb{E}(Y_t Y_t') = I_p$.
- (Y2) The entries of Y_t have finite absolute moments of order $4 + \epsilon$ for some $\epsilon > 0$.
- (K) The kernel function K is even (i.e. $K(u) = K(-u), u \in \mathbb{R}$), has bounded support $(-S, S)$, where it is positive, and is continuously differentiable on $(0, S)$. Furthermore, $\int_{\mathbb{R}} K(u) du = 1$.
- (A1) The matrix $A^2(u)$ is positive definite for $u \in (0, 1)$.
- (A2g) The entries of the matrix $A^2(u)$ are real analytic for $u \in \mathcal{H}$.

The assumptions on $h = h(T)$ as $T \rightarrow \infty$ can be found in Lemma A.2.

We assume implicitly as stated in Section 4.2 that \mathcal{H} is a closed subinterval of $(0, 1)$. Concerning Assumption (A2g), according to one possible definition, a function f is real analytic if its Taylor series converges to the function f in a neighborhood of each point. As noted for similar assumptions G4, G5 in Donald et al. [8], analyticity is assumed to have smoothness of the eigenvectors of analytic matrices involving $A^2(u)$. It is well known that smoothness of a matrix is not sufficient to have smooth eigenvectors (see, e.g., Kato [13], Bunse-Gerstner et al. [2]). Alternatively, the smoothness of the eigenvectors of interest can be assumed. We note that (A1) and (A2g) imply the continuous differentiability of $A(u)$, $A(u)^{-1}$ and $A(u)^{-2}$ for $u \in (0, 1)$.

The proof of Proposition 4.1 follows the path taken by Donald et al. [8]; see, in particular, its technical appendix online. We shall try to minimize repetitions by indicating only the key needed assertions. Some of the developments will be somewhat simpler since many of the considered matrices are symmetric. But we shall also need several new auxiliary results to account for the key difference from Donald et al. [8] in that smoothing through a kernel is carried out here in time t while this was over the values of a random variable in Donald et al. [8]. Some of the auxiliary results for the proof of Proposition 4.1 can be found in Appendix C of Sundararajan et al. [21] and the reader interested in proofs might want to look at them first, before going through the arguments in this section.

A number of comments concerning the notation are also in place. To simplify the notation, we shall drop the dependence on u , and write \widehat{A} , A , \widehat{M} , M , $\widehat{\xi}_r$, $\widehat{\gamma}_{2,i}$, etc. instead of $\widehat{A}(u)$, $A(u)$, $\widehat{M}(u)$, $M(u)$, $\widehat{\xi}_r(u)$, $\widehat{\gamma}_{2,i}(u)$, etc. Similarly, \sup will denote $\sup_{u \in \mathcal{H}}$ and $\xi = O_{p,\sup}(b_T)$ will stand for $\sup|\xi| = O_p(b_T)$. As throughout this work, $G_h(u)$ will stand for a scaled kernel function $h^{-1}G(h^{-1}u)$. The kernel function G will be normalized to integrated to 1 when it is important. We shall use both K in Assumption (K) and other functions related to K , which will be denoted as

$$\overline{K}(u) = \int_{\mathbb{R}} K(u - v)K(v) dv, \quad K_p^{(q)}(u) = C|u|^q K^p(u),$$

where C is such that $\|K_p^{(q)}\|_1 = 1$. Note that $K_1^{(0)}(u) = K(u)$. When $q = 0$, we shall simply write $K_p(u)$.

Proof of Proposition 4.1. We shall first prove (4.15). The first step consists of showing that the ordered eigenvalues $0 \leq \widehat{\gamma}_{2,1} \leq \dots \leq \widehat{\gamma}_{2,r}$ entering $\widehat{\xi}_r$ can be replaced in the asymptotic limit by the ordered eigenvalues $0 \leq \widehat{\lambda}_{2,1} \leq \dots \leq \widehat{\lambda}_{2,r}$ of the $r \times r$ matrix

$$D'(\widehat{M} - M)D^2(\widehat{M} - M)D = D'\widehat{M}D^2\widehat{M}D, \tag{A.6}$$

where D is a $p \times r$ matrix described below. The key here is that the matrix (A.6) is $r \times r$ so that the sum of its r eigenvalues is just its trace, which is amenable to easier manipulations. The matrix D will also play an important role of standardization.

The $p \times r$ matrix D and another $p \times (p - r)$ matrix \widetilde{D} enter into a $p \times p$ matrix $D_0 = (\widetilde{D}, D)$ characterized as follows: D_0 consists of the ‘‘eigenvectors’’ associated with the eigenvalues $\gamma_{2,i}$ through the characteristic equation

$$|(FMF)^2 - \gamma_{2,i}I_p| = |FMF^2MF - \gamma_{2,i}I_p| = |MF^2M - \gamma_{2,i}F^{-2}| = 0 \tag{A.7}$$

that satisfy

$$D'_0F^{-2}D_0 = F^{-1}D_0^2F^{-1} = I_p. \tag{A.8}$$

Said differently, $F^{-1}D_0$ consists of the eigenvectors of $(FMF)^2$, that is,

$$FMF^2MD_0 = F^{-1}D_0 \text{diag}(\gamma_{2,p}, \dots, \gamma_{2,1}). \quad (\text{A.9})$$

Since we deal with squared symmetric matrices, $F^{-1}D_0$ also consists of the eigenvectors of FMF , whose squared eigenvalues are $\gamma_{2,i}$, so that

$$FMD_0 = F^{-1}D_0 \text{diag}(\gamma_{p'}, \dots, \gamma_{1'}), \quad (\text{A.10})$$

where $(\gamma_{i'})^2 = \gamma_{2,i}$ and the prime indicates that the order is not necessarily that of the increasing order in $\gamma_1 \leq \dots \leq \gamma_p$. This fact will be used below. Note also that $\gamma_{2,r} = \dots = \gamma_{2,1} = 0$ by assumption and hence that $D'M = 0$.

The next result will justify the replacement of the eigenvalues $\widehat{\gamma}_{2,i}$.

Lemma A.1. *With the above notation and under Assumptions (A.1), (A.2g), we have for $i = 1, \dots, r$,*

$$\sup |a_T^2 \widehat{\gamma}_{2,i} - a_T^2 \widehat{\lambda}_i| = O_p((Th/\ln T)^{1/2}). \quad (\text{A.11})$$

Proof. Let $|B|$ denote the determinant of a matrix B . As on p. 173 of Robin and Smith [19], we have

$$\begin{aligned} 0 &= |\widehat{M}\widehat{F}^2\widehat{M} - \widehat{\gamma}_{2,i}\widehat{F}^{-2}| = |(\widetilde{D}, a_T D)'(\widehat{M}\widehat{F}^2\widehat{M} - \widehat{\gamma}_{2,i}\widehat{F}^{-2})(\widetilde{D}, a_T D)| \\ &= \left| \begin{pmatrix} \widetilde{D}'(\widehat{M}\widehat{F}^2\widehat{M} - \widehat{\gamma}_{2,i}\widehat{F}^{-2})\widetilde{D} & a_T \widetilde{D}'(\widehat{M}\widehat{F}^2\widehat{M} - \widehat{\gamma}_{2,i}\widehat{F}^{-2})D \\ a_T D'(\widehat{M}\widehat{F}^2\widehat{M} - \widehat{\gamma}_{2,i}\widehat{F}^{-2})\widetilde{D} & a_T^2 D'(\widehat{M}\widehat{F}^2\widehat{M} - \widehat{\gamma}_{2,i}\widehat{F}^{-2})D \end{pmatrix} \right|. \end{aligned}$$

By using the relation $|(B_{11} \ B_{12}; B_{21} \ B_{22})| = |B_{11}| \cdot |B_{22} - B_{21}B_{11}^{-1}B_{12}|$, we have further that

$$0 = |\widehat{S}| \cdot |\widehat{W} - a_T^2 \widehat{\gamma}_{2,i} \widehat{V}^{-1}|, \quad (\text{A.12})$$

where

$$\begin{aligned} \widehat{S} &= \widetilde{D}'(\widehat{M}\widehat{F}^2\widehat{M} - \widehat{\gamma}_{2,i}\widehat{F}^{-2})\widetilde{D}, \\ \widehat{W} &= a_T^2 \widetilde{D}'\widehat{M}\widehat{F}^2\widehat{M}D - a_T^2 D'\widehat{M}\widehat{F}^2\widehat{M}\widetilde{D}\widehat{S}^{-1}\widetilde{D}'\widehat{M}\widehat{F}^2\widehat{M}D, \\ \widehat{V}^{-1} &= D'\widehat{F}^{-2}D + \widehat{\gamma}_{2,i}D'\widehat{F}^{-2}\widetilde{D}\widehat{S}^{-1}\widetilde{D}'\widehat{F}^{-2}D \\ &\quad - D'\widehat{M}\widehat{F}^2\widehat{M}\widetilde{D}\widehat{S}^{-1}\widetilde{D}'\widehat{F}^{-2}D - D'\widehat{F}^{-2}\widetilde{D}\widehat{S}^{-1}\widetilde{D}'\widehat{M}\widehat{F}^2\widehat{M}D. \end{aligned}$$

By Proposition A.1 and the smoothness of \widetilde{D} by Proposition C.2 in Sundararajan et al. [21], note that

$$\widehat{S} = \widetilde{D}'MF^2M\widetilde{D} + O_{p,\sup}((Th/\ln T)^{-1/2}) = \text{diag}(\gamma_{2,p}, \dots, \gamma_{2,r+1}) + O_{p,\sup}((Th/\ln T)^{-1/2}),$$

where the second equality follows from (A.9) and (A.8). This shows that, asymptotically, $|\widehat{S}| > 0$. Hence, in view of (A.12), we may suppose without loss of generality that $a_T^2 \widehat{\gamma}_{2,i}$ are the eigenvalues of the matrix $\widehat{W}\widehat{V}$. The matrix \widehat{V} is symmetric and its eigenvalues are positive asymptotically since $\widehat{V} \rightarrow_p D'\widehat{F}^{-2}D$. Then, \widehat{V} may be assumed to be positive definite, and $a_T^2 \widehat{\gamma}_{2,i}$ be taken as eigenvalues of $\widehat{V}^{1/2}\widehat{W}\widehat{V}^{1/2}$. Since the matrix is symmetric, by applying the Wielandt–Hoffman theorem (see, e.g., Golub and Van Loan [11]), we get that

$$\sup |a_T^2 \widehat{\gamma}_{2,i} - a_T^2 \widehat{\lambda}_i| \leq \sup |\widehat{V}^{1/2}\widehat{W}\widehat{V}^{1/2} - a_T^2 D'(\widehat{M} - M)D^2(\widehat{M} - M)D|.$$

By using Proposition C.1 in Sundararajan et al. [21] and the fact that the square root is a continuous operation on positive definite matrices, $\widehat{V}^{1/2} = (D'F^{-2}D)^{1/2} + O_{p,\text{sup}}((Th/\ln T)^{-1/2}) = I_r + O_{p,\text{sup}}((Th/\ln T)^{-1/2})$. Similarly, for \widehat{W} , we have $\widehat{W} = a_T^2 D' \widehat{M} D^2 \widehat{M} D + O_{p,\text{sup}}((Th/\ln T)^{-1/2}) = a_T^2 D'(\widehat{M} - M)D^2(\widehat{M} - M)D + O_{p,\text{sup}}((Th/\ln T)^{-1/2})$, where the first equality follows from the relation

$$F^2 - F^2 M \widetilde{D} \text{diag}(\gamma_{2,p}, \dots, \gamma_{2,r+1})^{-1} \widetilde{D}' M F^2 = D^2.$$

The latter relation holds by the following argument. Note that it is equivalent to

$$I = F M \widetilde{D} \text{diag}(\gamma_{2,p}, \dots, \gamma_{2,r+1})^{-1} \widetilde{D}' M F + F^{-1} D^2 F^{-1}$$

and in view of (A.9), follows from $F M \widetilde{D} \text{diag}(\gamma_{2,p}, \dots, \gamma_{2,r+1})^{-1} \widetilde{D}' M F = F^{-1} \widetilde{D}^2 F^{-1}$ or

$$F^2 M \widetilde{D} \text{diag}(\gamma_{2,p}, \dots, \gamma_{2,r+1})^{-1} \widetilde{D}' M F^2 = \widetilde{D}^2,$$

which is a consequence of (A.10). \square

By Lemma A.1, instead of working with $\int \widehat{\xi}_r du$, we can focus instead on

$$\widehat{L}_r = a_T^2 \sum_{i=1}^r \widehat{\lambda}_i = a_T^2 \text{tr}\{D' \widehat{M} D^2 \widehat{M} D\} = a_T^2 \text{tr}\{D'(\widehat{M} - M)D^2(\widehat{M} - M)D\} \quad (\text{A.13})$$

and $\int \widehat{L}_r du$. Write $\widehat{M} - M = S_1 + S_2 - S_3 - S_4$, where

$$\begin{aligned} S_1 &= \frac{1}{T} \sum_{t=1}^T \left(X_t X_t' - A^2\left(\frac{t}{T}\right) \right) K_h\left(u - \frac{t}{T}\right), \\ S_2 &= \frac{1}{T} \sum_{t=1}^T A^2\left(\frac{t}{T}\right) K_h\left(u - \frac{t}{T}\right) - A^2(u), \\ S_3 &= \frac{1}{T} \sum_{t=1}^T \left(X_t X_t' - A^2\left(\frac{t}{T}\right) \right), \\ S_4 &= \frac{1}{T} \sum_{t=1}^T A^2\left(\frac{t}{T}\right) - \int_0^1 A^2(u) du. \end{aligned} \quad (\text{A.14})$$

Then,

$$\int \widehat{L}_r du = \sum_{j,k=1}^4 (\pm 1) \int \widehat{L}_{r,jk} du =: \sum_{j,k=1}^4 \widehat{G}_{r,jk}, \quad (\text{A.15})$$

where

$$\widehat{L}_{r,jk} = a_T^2 \text{tr}\{D' S_j D^2 S_k D\} \quad (\text{A.16})$$

and (± 1) in (A.15) accounts for the signs of S_j, S_k . The next two lemmas concern the asymptotics of $\widehat{G}_{r,jk}$.

Lemma A.2. Under Assumptions (Y1), (Y2), (K), (A1), (A2g), and

$$T \rightarrow \infty, \quad h \rightarrow 0, \quad Th^{3/2} \rightarrow \infty, \quad T^{\epsilon/(4+\epsilon)}h^{1/2} \rightarrow \infty,$$

we have

$$\frac{\widehat{G}_{r,11} - |\mathcal{H}|^{\frac{r(\mu_4+r-1)}{\mu_4}}}{\sqrt{h \frac{\|\bar{K}\|_2^2}{\|K\|_2^2} \frac{2(r\mu_4^2+2r(r-1))}{\mu_4^2} |\mathcal{H}|}} \xrightarrow{d} \mathcal{N}(0, 1). \quad (\text{A.17})$$

As noted following Assumption (A2g), this assumption is needed to apply Proposition C.2 in Sundararajan et al. [21] to have the smoothness of $D = D(u)$.

Proof. In view of the definition of S_1 in (A.14), we can write

$$\begin{aligned} \widehat{G}_{r,11} &= \int_{\mathcal{H}} a_T^2 \operatorname{tr}\{D' S_1 D^2 S_1 D\} du = I_1 + I_2 \\ &:= \frac{a_T^2}{T^2} \sum_{t=1}^T \int_{\mathcal{H}} \operatorname{tr}\left\{D'\left(X_t X'_t - A^2\left(\frac{t}{T}\right)\right) D^2\left(X_t X'_t - A^2\left(\frac{t}{T}\right)\right) D\right\} K_h^2\left(u - \frac{t}{T}\right) du \\ &\quad + \frac{2a_T^2}{T^2} \sum_{t_1 < t_2} \int_{\mathcal{H}} \operatorname{tr}\left\{D'\left(X_{t_1} X'_{t_1} - A^2\left(\frac{t_1}{T}\right)\right) D^2\left(X_{t_2} X'_{t_2} - A^2\left(\frac{t_2}{T}\right)\right) D\right\} \\ &\quad \times K_h\left(u - \frac{t_1}{T}\right) K_h\left(u - \frac{t_2}{T}\right) du. \end{aligned}$$

We shall argue first that $D = D(u)$ in I_1 and I_2 above can be replaced by $D(t/T)$, and shall denote the respective terms by \tilde{I}_1 and \tilde{I}_2 . The relation (A.8) will then be used to simplify \tilde{I}_1 and \tilde{I}_2 . Finally, we will show that \tilde{I}_1 produces the centering in (A.17) and \tilde{I}_2 yields the asymptotic normality in (A.17).

To see why I_1 can be replaced by \tilde{I}_1 , that is, $D = D(u)$ be replaced by $D(t/T)$, consider one of the terms in the difference between I_1 and \tilde{I}_1 , namely,

$$R_1 = \frac{a_T^2}{T^2} \sum_{t=1}^T \int_{\mathcal{H}} \operatorname{tr}\left\{\left(D - D\left(\frac{t}{T}\right)\right)' \left(X_t X'_t - A^2\left(\frac{t}{T}\right)\right) D^2\left(X_t X'_t - A^2\left(\frac{t}{T}\right)\right) D\right\} K_h^2\left(u - \frac{t}{T}\right) du.$$

(The other terms in the difference can be dealt with similarly.) Then, by using the smoothness of $D(u)$ by Proposition C.2 in Sundararajan et al. [21], the expression (4.3) for a_T and Lemma C.1 in Sundararajan et al. [21], (ii), we have

$$\begin{aligned} |R_1| &\leq \frac{C}{T} \sum_{t=1}^T \left\| X_t X'_t - A^2\left(\frac{t}{T}\right) \right\|_2 \int_{\mathcal{H}} \left\| D(u) - D\left(\frac{t}{T}\right) \right\| K_{2,h}\left(u - \frac{t}{T}\right) du \\ &\leq \frac{C'h}{T} \sum_{t=1}^T \left\| X_t X'_t - A^2\left(\frac{t}{T}\right) \right\|_2 \int_{\mathcal{H}} K_{2,h}^{(1)}\left(u - \frac{t}{T}\right) du \\ &\leq \frac{C''h}{T} \sum_{t=1}^T \left\| X_t X'_t - A^2\left(\frac{t}{T}\right) \right\|_2^2 \leq C(w)h, \end{aligned}$$

for a random constant $C(w)$. The term R_1 would then not affect the asymptotics (A.17) since $h/\sqrt{h} \rightarrow 0$.

For I_2 and \tilde{I}_2 , consider similarly a term from their difference given by

$$\begin{aligned} R_2 &= \sum_{t_1 < t_2} b_{T,t_1,t_2}(X_{t_1} X'_{t_1}, X_{t_2} X'_{t_2}) \\ &:= \sum_{t_1 < t_2} \frac{2a_T^2}{T^2} \int_{\mathcal{H}} \text{tr} \left\{ \left(D - D\left(\frac{t}{T}\right) \right)' \left(X_{t_1} X'_{t_1} - A^2\left(\frac{t_1}{T}\right) \right) D^2 \left(X_{t_2} X'_{t_2} - A^2\left(\frac{t_2}{T}\right) \right) D \right\} \\ &\quad \times K_h\left(u - \frac{t_1}{T}\right) K_h\left(u - \frac{t_2}{T}\right) du. \end{aligned}$$

We shall use Proposition C.3 in Sundararajan et al. [21] to obtain a convergence rate for R_2 . To apply the proposition, a number of (conditional) expectations involving $b_{T,t_1,t_2}(X_{t_1} X'_{t_1}, X_{t_2} X'_{t_2})$ need to be evaluated. Note that $\mathbb{E}b_{T,t_1,t_2}(X_{t_1} X'_{t_1}, X_{t_2} X'_{t_2}) = 0$ and $\mathbb{E}(b_{T,t_1,t_2}(X_{t_1} X'_{t_1}, X_{t_2} X'_{t_2}) | X_{t_1} X'_{t_1}) = 0$ and similarly when conditioning on $X_{t_2} X'_{t_2}$. We thus only need to consider $\mathbb{E}(b_{T,t_1,t_2}(X_{t_1} X'_{t_1}, X_{t_2} X'_{t_2}))^2$. By using the generalized Minkowski's inequality and the smoothness of $D(u)$ by Proposition C.2 in Sundararajan et al. [21], note that

$$\begin{aligned} &(\mathbb{E}(b_{T,t_1,t_2}(X_{t_1} X'_{t_1}, X_{t_2} X'_{t_2}))^2)^{1/2} \\ &\leq \int_{\mathcal{H}} \left(\mathbb{E} \left(\frac{2a_T^2}{T^2} \text{tr} \left\{ \left(D - D\left(\frac{t}{T}\right) \right)' \left(X_{t_1} X'_{t_1} - A^2\left(\frac{t_1}{T}\right) \right) D^2 \left(X_{t_2} X'_{t_2} - A^2\left(\frac{t_2}{T}\right) \right) D \right\} \right. \right. \\ &\quad \left. \left. \times K_h\left(u - \frac{t_1}{T}\right) K_h\left(u - \frac{t_2}{T}\right) \right)^2 \right)^{1/2} du \\ &\leq \frac{Ca_T^2}{T^2} \int_{\mathcal{H}} \left\| D(u) - D\left(\frac{t_1}{T}\right) \right\|_2 K_h\left(u - \frac{t_1}{T}\right) K_h\left(u - \frac{t_2}{T}\right) du \\ &\leq \frac{C'h^2}{T} \int_{\mathcal{H}} K_h^{(1)}\left(u - \frac{t_1}{T}\right) K_h\left(u - \frac{t_2}{T}\right) du, \end{aligned}$$

where we used the definition of a_T . It follows from Lemma C.2 in Sundararajan et al. [21] that

$$\sum_{t_1 < t_2} \mathbb{E}(b_{T,t_1,t_2}(X_{t_1} X'_{t_1}, X_{t_2} X'_{t_2}))^2 \leq \frac{Ch^4}{T^2} \sum_{t_1 < t_2} \left(\int_{\mathcal{H}} K_h^{(1)}\left(u - \frac{t_1}{T}\right) K_h\left(u - \frac{t_2}{T}\right) du \right)^2 \leq C'h^3.$$

Hence, by Proposition C.3 in Sundararajan et al. [21], R_2 is of the order $O_p(h^{3/2})$ and hence does not affect the asymptotics (A.17) since $h^{3/2}/\sqrt{h} \rightarrow 0$.

We can thus replace I_1 and I_2 by \tilde{I}_1 and \tilde{I}_2 , respectively, which in view of (A.8) can be expressed as

$$\begin{aligned} \tilde{I}_1 &= \frac{a_T^2}{T^2} \sum_{t=1}^T \text{tr} \{ (\tilde{Y}_t \tilde{Y}'_t - I_r) (\tilde{Y}_t \tilde{Y}'_t - I_r) \} \int_{\mathcal{H}} K_h^2\left(u - \frac{t}{T}\right) du, \\ \tilde{I}_2 &= \frac{2a_T^2}{T^2} \sum_{t_1 < t_2} \text{tr} \{ (\tilde{Y}_{t_1} \tilde{Y}'_{t_1} - I_r) (\tilde{Y}_{t_2} \tilde{Y}'_{t_2} - I_r) \} \int_{\mathcal{H}} K_h\left(u - \frac{t_1}{T}\right) K_h\left(u - \frac{t_2}{T}\right) du, \end{aligned}$$

where $\tilde{Y}_t = D'(t/T)X_t$.

For \tilde{I}_1 , write it as

$$\begin{aligned} \tilde{I}_1 &= \frac{1}{\mu_4 T} \sum_{t=1}^T \sum_{i,j=1}^r (\tilde{Y}_t \tilde{Y}'_t - I_r)_{ij}^2 \int_{\mathcal{H}} K_{2,h} \left(u - \frac{t}{T} \right) du \\ &= \frac{1}{\mu_4 T} \sum_{t=1}^T \sum_{i,j=1}^r \left((\tilde{Y}_t \tilde{Y}'_t - I_r)_{ij}^2 - \mathbb{E}(\tilde{Y}_t \tilde{Y}'_t - I_r)_{ij}^2 \right) \int_{\mathcal{H}} K_{2,h} \left(u - \frac{t}{T} \right) du \\ &\quad + \sum_{i,j=1}^r \mathbb{E}(\tilde{Y}_1 \tilde{Y}'_1 - I_r)_{ij}^2 \frac{1}{\mu_4 T} \sum_{t=1}^T \int_{\mathcal{H}} K_{2,h} \left(u - \frac{t}{T} \right) du = \tilde{I}_{1,1} + \tilde{I}_{1,2}. \end{aligned}$$

It can be checked that

$$\sum_{i,j=1}^r \mathbb{E}(\tilde{Y}_1 \tilde{Y}'_1 - I_r)_{ij}^2 = r\mu_4 + r(r-1).$$

Hence, by Lemma C.1 in Sundararajan et al. [21], (ii),

$$\tilde{I}_{1,2} = |\mathcal{H}| \frac{r\mu_4 + r(r-1)}{\mu_4} + O\left(\frac{1}{Th}\right).$$

For $\tilde{I}_{1,1}$, setting $\xi_t = \sum_{i,j=1}^r ((\tilde{Y}_t \tilde{Y}'_t - I_r)_{ij}^2 - \mathbb{E}(\tilde{Y}_t \tilde{Y}'_t - I_r)_{ij}^2)$, $p = 1 + \epsilon/4$ and using the von Bahr-Esseen inequality (von Bahr and Esseen [23]), we obtain that

$$\mathbb{E}|\tilde{I}_{1,1}|^p = C \mathbb{E} \left| \frac{1}{T} \sum_{t=1}^T \xi_t \int_{\mathcal{H}} K_{2,h} \left(u - \frac{t}{T} \right) du \right|^p \leq \frac{C' \mathbb{E}|\xi_1|^p}{T^p} \sum_{t=1}^T \left| \int_{\mathcal{H}} K_{2,h} \left(u - \frac{t}{T} \right) du \right|^p \leq \frac{C''}{T^{p-1}}$$

and hence $\tilde{I}_{1,1} = O_p(T^{1/p-1})$. By assumption, $Th^{3/2} \rightarrow \infty$ and $T^{1-1/p}h^{1/2} = T^{\epsilon/(4+\epsilon)}h^{1/2} \rightarrow \infty$. This shows that the error term in $\tilde{I}_{1,1}$ and $\tilde{I}_{1,2}$ do not affect the asymptotics (A.17), and that $\tilde{I}_{1,2}$ produces the desired asymptotic mean.

For \tilde{I}_2 , consider

$$\frac{\tilde{I}_2}{\sqrt{h}} = \sum_{t_1 < t_2} b_{T,t_1,t_2},$$

where

$$b_{T,t_1,t_2} = \frac{2h^{1/2}}{\|K\|_2^2 \mu_4 T} \text{tr} \{ (\tilde{Y}_{t_1} \tilde{Y}'_{t_1} - I_r) (\tilde{Y}_{t_2} \tilde{Y}'_{t_2} - I_r) \} \int_{\mathcal{H}} K_h \left(u - \frac{t_1}{T} \right) K_h \left(u - \frac{t_2}{T} \right) du.$$

We will argue that \tilde{I}_2/\sqrt{h} is asymptotically normal with the desired limiting variance. By Proposition 3.2 in de Jong [6], it is enough to show that

1. $\text{Var}(\frac{\tilde{I}_2}{\sqrt{h}}) \rightarrow \frac{\|\bar{K}\|_2^2}{\|K\|_2^4} \frac{2(r\mu_4^2 + 2r(r-1))}{\mu_4^2} |\mathcal{H}|$;
2. $G_{T,i} = o(1), i = 1, 2, 4$, where

$$G_{T,1} = \sum_{t_1 < t_2} \mathbb{E}b_{T,t_1,t_2}^4,$$

$$\begin{aligned}
 G_{T,2} &= \sum_{t_1 < t_2 < t_3} \mathbb{E} b_{T,t_1,t_2}^2 b_{T,t_1,t_3}^2 + \mathbb{E} b_{T,t_1,t_2}^2 b_{T,t_2,t_3}^2 + \mathbb{E} b_{T,t_1,t_3}^2 b_{T,t_2,t_3}^2, \\
 G_{T,4} &= \sum_{t_1 < t_2 < t_3 < t_4} \mathbb{E} b_{T,t_1,t_2} b_{T,t_1,t_3} b_{T,t_2,t_4} b_{T,t_3,t_4} + \mathbb{E} b_{T,t_1,t_2} b_{T,t_1,t_4} b_{T,t_2,t_3} b_{T,t_3,t_4} \\
 &\quad + \mathbb{E} b_{T,t_1,t_3} b_{T,t_1,t_4} b_{T,t_2,t_4} b_{T,t_2,t_4}.
 \end{aligned}$$

The first point above follows from

$$\text{Var}\left(\frac{\tilde{I}_2}{\sqrt{h}}\right) = \frac{4h}{\|K\|_2^4 \mu_4^2 T^2} \sum_{t_1 < t_2} \mathbb{E} \text{tr}^2\{(\tilde{Y}_{t_1} \tilde{Y}'_{t_1} - I_r)(\tilde{Y}_{t_2} \tilde{Y}'_{t_2} - I_r)\}^2 \left(\int_{\mathcal{H}} K_h\left(u - \frac{t_1}{T}\right) K_h\left(u - \frac{t_2}{T}\right) du\right),$$

the observation that

$$\mathbb{E} \text{tr}^2\{(\tilde{Y}_{t_1} \tilde{Y}'_{t_1} - I_r)(\tilde{Y}_{t_2} \tilde{Y}'_{t_2} - I_r)\} = r\mu_4^2 + 2r(r-1)$$

and Lemma C.2 in Sundararajan et al. [21]. For the second point above, $G_{T,1}$ can be bounded up to a constant by

$$\frac{h^2}{T^4} \sum_{t_1 < t_2} \left(\int_{\mathcal{H}} K_h\left(u - \frac{t_1}{T}\right) K_h\left(u - \frac{t_2}{T}\right) du\right)^4.$$

For example, the first term in the sum of $G_{T,2}$ can be bounded up to a constant by

$$\frac{h^2}{T^4} \sum_{t_1 < t_2 < t_3} \left(\int_{\mathcal{H}} K_h\left(u - \frac{t_1}{T}\right) K_h\left(u - \frac{t_2}{T}\right) du\right)^2 \left(\int_{\mathcal{H}} K_h\left(u - \frac{t_1}{T}\right) K_h\left(u - \frac{t_3}{T}\right) du\right)^2.$$

For example, the first term in the sum of $G_{T,4}$ can be bounded up to a constant by

$$\begin{aligned}
 &\frac{h^2}{T^4} \sum_{t_1 < t_2 < t_3 < t_4} \left(\int_{\mathcal{H}} K_h\left(u - \frac{t_1}{T}\right) K_h\left(u - \frac{t_2}{T}\right) du\right) \left(\int_{\mathcal{H}} K_h\left(u - \frac{t_1}{T}\right) K_h\left(u - \frac{t_3}{T}\right) du\right) \\
 &\quad \times \left(\int_{\mathcal{H}} K_h\left(u - \frac{t_2}{T}\right) K_h\left(u - \frac{t_4}{T}\right) du\right) \left(\int_{\mathcal{H}} K_h\left(u - \frac{t_3}{T}\right) K_h\left(u - \frac{t_4}{T}\right) du\right).
 \end{aligned}$$

The rate $o(1)$ for each of these bounds follows from Lemma C.2 in Sundararajan et al. [21]. This completes the proof of the lemma. \square

Lemma A.3. *Under Assumptions (Y1), (Y2), (K), (A1), (A2g), and*

$$T \rightarrow \infty, \quad h \rightarrow 0, \quad Th^{5/2} \rightarrow \infty, \quad Th^3 \rightarrow 0,$$

we have, for $(j, k) \neq (1, 1)$,

$$\widehat{G}_{r,jk} = \int \widehat{L}_{r,jk} du = o_p(h^{1/2}). \tag{A.18}$$

In fact, the proof of lemma establishes sharper rates of convergence of $\widehat{G}_{r,jk}$ to 0. The rate given in the lemma is what is needed to conclude the convergence (4.15). Indeed, the latter now follows immediately from the arguments above and, in particular, Lemmas A.2 and A.3.

Proof. We consider only the cases $(j, k) = (2, 2), (3, 3), (4, 4), (1, 2), (1, 3)$ and $(1, 4)$. The other mixed cases can be dealt with similarly.

For $(j, k) = (2, 2)$, we have $\int \widehat{L}_{r,22} du = a_T^2 \int \text{tr}\{D' S_2 D^2 S_2 D\} du$ where S_2 is defined in (A.14). As in the proof of result in (4.2), $S_2 = O(h + (Th)^{-1})$ uniformly in u , where the latter follows by using Lemma C.1 in Sundararajan et al. [21], (ii). Then, by using the smoothness of $D = D(u)$ by Proposition C.2 in Sundararajan et al. [21], we have $\int \widehat{L}_{r,22} du = O(Th(h + (Th)^{-1})^2) = O(Th^3 + (Th)^{-1} + h)$. This is of the order $o(h^{1/2})$ since, in particular, $Th^{5/2} \rightarrow 0$ by assumption. For $(j, k) = (3, 3)$, we have similarly $\int \widehat{L}_{r,33} du = a_T^2 \int \text{tr}\{D' S_3 D^2 S_3 D\} du$ where S_3 is defined in (A.14). The term S_3 does not depend on u and its rate is $O_p(T^{-1/2})$ as shown in the proof of result in (4.2). This leads to $\int \widehat{L}_{r,33} du = O_p(Th(T^{-1/2})^2) = O_p(h)$, which is again of the order $o_p(h^{1/2})$ as desired. For $(j, k) = (4, 4)$, S_4 does not depend on u either and its rate is $O(T^{-1})$ as shown in the proof of result in (4.2). This leads to $\int \widehat{L}_{r,44} du = O(Th(T^{-1})^2) = O(h/T) = o(h^{1/2})$.

For $(j, k) = (1, 2)$, we need to consider $\int \widehat{L}_{r,12} du = a_T^2 \int \text{tr}\{D' S_1 D^2 S_2 D\} du$. After matrix multiplication and taking the trace, a general term in $\int \widehat{L}_{r,12} du$ has the form

$$R_{1,2} = a_T^2 \int D'_*(S_1)_*(D^2)_*(S_2)_*D_* du,$$

where $*$ refers to an index pair that can change from matrix to matrix. Furthermore, in view of S_4 defined in (A.14),

$$R_{1,2} = \frac{a_T^2}{T} \sum_{t=1}^T \left((X_t X'_t)_* - A^2 \left(\frac{t}{T} \right)_* \right) \int D_* D_*^2 D'_*(S_2)_* K_h \left(u - \frac{t}{T} \right) du.$$

By using the facts that $S_2 = O(h + (Th)^{-1})$ uniformly in u as above and $\int_{\mathbb{R}} K_h(u - t/T) du = 1$, it follows that $\mathbb{E}R_{1,2}^2 = O(a_T^4 T^{-1} (h + (Th)^{-1})^2) = O(Th^4 + h^2 + 1/T)$ and hence that $R_{1,2} = o_p(h^{1/2})$ since, in particular, $Th^3 \rightarrow 0$. For $(j, k) = (1, 3)$, a general term of interest is similarly,

$$\begin{aligned} R_{1,3} &= a_T^2 \int D'_*(S_1)_*(D^2)_*(S_3)_*D'_* du \\ &= \frac{a_T^2 (S_3)_*}{T} \sum_{t=1}^T \left((X_t X'_t)_* - A^2 \left(\frac{t}{T} \right)_* \right) \int D_* D_*^2 D'_* K_h \left(u - \frac{t}{T} \right) du \end{aligned}$$

and hence $\mathbb{E}R_{1,3}^2 = a_T^4 T^{-2} = h^4$. This leads to $R_{1,3} = o_p(h^{1/2})$. The case $(j, k) = (1, 4)$ can be dealt with similarly by using the fact that $S_4 = O(T^{-1})$. □

Finally, we prove the last statement of Proposition 4.1 concerning the behavior of the test statistic under the alternative. For this, note that by Proposition C.1 in Sundararajan et al. [21], $\sup_{u \in \mathcal{H}} |a_T^{-2} \xi_r(u) - \sum_{i=1}^r \gamma_{2,i}(u)| \rightarrow_p 0$. Furthermore, under the considered alternative H_1 , and by using smoothness of $\gamma_{2,i}(u)$, we have $\sup_{u \in \mathcal{H}} \sum_{i=1}^r \gamma_{2,i}(u) > 0$. It follows that $\widehat{\xi}_r = (a_T^2 \int_{\mathcal{H}} a_T^{-2} \times \xi_r(u) du - C_1)/(C_2 h^{1/2})$ with constants C_1, C_2 behaves asymptotically as $C a_T^2/h^{1/2} = C' Th^{1/2} \rightarrow \infty$ (for example, since Lemma A.2 assumes $Th^{3/2} \rightarrow \infty$). This concludes the proof of Proposition 4.1. □

Acknowledgements

We would like to thank the two anonymous referees and the associate editor for their useful comments that greatly improved the content and presentation of the paper. The second author was supported in part by the National Science Foundation grant DMS-1712966. The third author was supported by the National Science Foundation grant DMS-1612984.

Supplementary Material

Supplement to “Stationary subspace analysis of nonstationary covariance processes: Eigenstructure description and testing” (DOI: [10.3150/20-BEJ1243SUPP](https://doi.org/10.3150/20-BEJ1243SUPP); .pdf). Section A contains the proof of the result in (4.2), Section B contains a consistent estimator for the fourth moment and Section C includes the key technical results for the proof of Proposition 4.1. In Section D, we provide supporting plots for the models discussed in Section 6. Finally, in Section E, we include additional figures and tables from the real data application in Section 7.

References

- [1] Blythe, D.A.J., von Bunau, P., Meinecke, F.C. and Muller, K.R. (2012). Feature extraction for change-point detection using stationary subspace analysis. *IEEE Trans. Neural Netw. Learn. Syst.* **23** 631–643.
- [2] Bunse-Gerstner, A., Byers, R., Mehrmann, V. and Nichols, N.K. (1991). Numerical computation of an analytic singular value decomposition of a matrix valued function. *Numer. Math.* **60** 1–39. [MR1131497 https://doi.org/10.1007/BF01385712](https://doi.org/10.1007/BF01385712)
- [3] Cardinali, A. and Nason, G.P. (2010). Costationarity of locally stationary time series. *J. Time Ser. Econom.* **2** Art. 1, 33. [MR2915633 https://doi.org/10.2202/1941-1928.1074](https://doi.org/10.2202/1941-1928.1074)
- [4] Dahlhaus, R. (1997). Fitting time series models to nonstationary processes. *Ann. Statist.* **25** 1–37. [MR1429916 https://doi.org/10.1214/aos/1034276620](https://doi.org/10.1214/aos/1034276620)
- [5] Dahlhaus, R. (2012). 13 – locally stationary processes. In *Time Series Analysis: Methods and Applications* (Subba Rao, T., Subba Rao, S. and Rao, C.R., eds.). *Handbook of Statistics* **30** 351–413. Amsterdam: Elsevier. [MR3295420](https://doi.org/10.1016/B978-0-12-385183-9.00013-1)
- [6] de Jong, P. (1987). A central limit theorem for generalized quadratic forms. *Probab. Theory Related Fields* **75** 261–277. [MR0885466 https://doi.org/10.1007/BF00354037](https://doi.org/10.1007/BF00354037)
- [7] Donald, S.G., Fortuna, N. and Pipiras, V. (2007). On rank estimation in symmetric matrices: The case of indefinite matrix estimators. *Econometric Theory* **23** 1217–1232. [MR2397198 https://doi.org/10.1017/S0266466607070478](https://doi.org/10.1017/S0266466607070478)
- [8] Donald, S.G., Fortuna, N. and Pipiras, V. (2011). Local and global rank tests for multivariate varying-coefficient models. *J. Bus. Econom. Statist.* **29** 295–306. [MR2807882 https://doi.org/10.1198/jbes.2010.07303](https://doi.org/10.1198/jbes.2010.07303)
- [9] Düker, M., Pipiras, V. and Sundararajan, R. (2019). Cotrending: Testing for common deterministic trends in varying means model. Preprint.
- [10] Fortuna, N. (2008). Local rank tests in a multivariate nonparametric relationship. *J. Econometrics* **142** 162–182. [MR2408734 https://doi.org/10.1016/j.jeconom.2007.03.001](https://doi.org/10.1016/j.jeconom.2007.03.001)
- [11] Golub, G.H. and Van Loan, C.F. (2013). *Matrix Computations*, 4th ed. *Johns Hopkins Studies in the Mathematical Sciences*. Baltimore, MD: Johns Hopkins Univ. Press. [MR3024913](https://doi.org/10.1017/C9781107071132)
- [12] Kaplan, A.Y., Fingelkurts, A.A., Fingelkurts, A.A., Borisov, S.V. and Darkhovsky, B.S. (2005). Nonstationary nature of the brain activity as revealed by EEG/MEG: Methodological, practical and conceptual challenges. *Signal Process.* **85** 2190–2212.
- [13] Kato, T. (1966). *Perturbation Theory for Linear Operators*. *Die Grundlehren der Mathematischen Wissenschaften, Band 132*. New York: Springer. [MR0203473](https://doi.org/10.1007/978-1-4939-9829-9)

- [14] Lotte, F. and Guan, C.T. (2011). Regularizing common spatial patterns to improve BCI designs: Unified theory and new algorithms. *IEEE Trans. Biomed. Eng.* **58** 355–362.
- [15] Magnus, J.R. and Neudecker, H. (1999). *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics. Chichester: Wiley. MR1698873
- [16] Naem, M., Brunner, C., Leeb, R., Graimann, B. and Pfurtscheller, G. (2006). Seperability of four-class motor imagery data using independent components analysis. *J. Neural Eng.* **3** 208.
- [17] Ombao, H., von Sachs, R. and Guo, W. (2005). SLEX analysis of multivariate nonstationary time series. *J. Amer. Statist. Assoc.* **100** 519–531. MR2160556 <https://doi.org/10.1198/016214504000001448>
- [18] Pons, P. and Latapy, M. (2005). Computing communities in large networks using random walks. In *Computer and Information Sciences – ISCIS 2005* (P. Yolum, T. Güngör, F. Gürgen and C. Özturan, eds.) 284–293. Berlin, Heidelberg: Springer.
- [19] Robin, J.-M. and Smith, R.J. (2000). Tests of rank. *Econometric Theory* **16** 151–175. MR1763430 <https://doi.org/10.1017/S0266466600162012>
- [20] Sundararajan, R.R., Palma, M.A. and Pourahmadi, M. (2017). Reducing brain signal noise in the prediction of economic choices: A case study in neuroeconomics. *Front. Neurosci.* **11** 704.
- [21] Sundararajan, R.R., Pipiras, V. and Pourahmadi, M. (2021). Supplement to “Stationary subspace analysis of nonstationary covariance processes: Eigenstructure description and testing.” <https://doi.org/10.3150/20-BEJ1243SUPP>
- [22] Sundararajan, R.R. and Pourahmadi, M. (2018). Stationary subspace analysis of nonstationary processes. *J. Time Series Anal.* **39** 338–355. MR3796523 <https://doi.org/10.1111/jtsa.12274>
- [23] von Bahr, B. and Esseen, C.-G. (1965). Inequalities for the r th absolute moment of a sum of random variables, $1 \leq r \leq 2$. *Ann. Math. Stat.* **36** 299–303. MR0170407 <https://doi.org/10.1214/aoms/1177700291>
- [24] von Bünau, P., Meinecke, F.C., Király, F.C. and Müller, K.-R. (2009). Finding stationary subspaces in multivariate time series. *Phys. Rev. Lett.* **103** 214101.
- [25] von Bünau, P., Meinecke, F.C., Scholler, S. and Müller, K.R. (2010). Finding stationary brain sources in EEG data. In *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology* 2810–2813.

Received April 2019 and revised May 2020