Testing Simultaneous Diagonalizability

Yuchen Xu* Marie-Christine Düker†
David S. Matteson‡
Department of Statistics and Data Science, Cornell University

November 8, 2022

Abstract

This paper proposes novel methods to test for simultaneous diagonalization of possibly asymmetric matrices. Motivated by various applications, a two-sample test as well as a generalization for multiple matrices are proposed. A partial version of the test is also studied to check whether a partial set of eigenvectors is shared across samples. Additionally, a novel algorithm for the considered testing methods is introduced. Simulation studies demonstrate favorable performance for all designs. Finally, the theoretical results are utilized to decouple multiple vector autoregression models into univariate time series, and to test for the same stationary distribution in recurrent Markov chains. These applications are demonstrated using macroeconomic indices of 8 countries and streamflow data, respectively.

Keywords: common eigenvectors, joint diagonalization, partially common eigenvectors, likelihood ratio test, vector autoregression, Markov chain, Wald test, dimension reduction.

*Corresponding author: Yuchen Xu, Email address: yx439@cornell.edu
†Email address: md2224@cornell.edu
‡Email address: matteson@cornell.edu

1 Introduction

Understanding the eigenvectors and the eigenspace of matrix-valued objects is known to be of fundamental interest in various disciplines including statistics, machine learning, and computer science. Knowledge about the eigenvectors and the eigenspace is particularly valuable in principal component analysis (PCA) [Nadler, 2008, Cai et al., 2013, Koltchinskii and Lounici, 2017] , covariance matrix estimation [Fan et al., 2013, 2015, 2018], spectral clustering [Von Luxburg, 2007, Rohe et al., 2011, Lei and Rinaldo, 2015], and network or graph theory [Tang and Priebe, 2018, Paul and Chen, 2020]. Often times it provides information for dimension reduction and clustering procedures.

This paper develops statistical tests and algorithms to check whether a set of square matrices can be diagonalized simultaneously. We are particularly interested in asymmetric square matrices with more general and flexible structural assumptions compared to symmetric ones like covariance matrices. Our work proceed from two-sample tests to multi-sample tests, and finally extends into partial cases where only a subset of eigenvectors is of interest. Besides providing the theoretical foundation and introducing practical algorithms, we motivate the usefulness of our results in several examples.

Our setting is as follows. Suppose we have a sequence of deterministic matrices $\{M_i\}_{i=1}^p \subset \mathbb{R}^{d\times d}$. Then the hypothesis testing problem we are interested in can be expressed as: the null hypothesis is H_0: $\{M_i\}_{i=1}^p$ can be jointly diagonalized, or equivalently,

$$H_0 \ : \ \exists V \in \mathbb{R}^{d\times d}, \ s.t. \ M_iV = VD_i, \ \forall i = 1,\dots,p, \ D_i \in \mathcal{M}_{\text{diag}}(d), \quad (1)$$

where $\mathcal{M}_{\text{diag}}(d)$ denotes the set of diagonal matrices in $\mathbb{R}^{d\times d}$. A modification of the problem is to discuss whether a set of matrices shares a partial set of eigenvectors. The null hypoth-

esis is then expressed as H_0^*: $\{M_i\}_{i=1}^p$ share k left eigenvectors ($k < d$), or equivalently,

$$H_0^* \ : \quad \exists V = (v_1, \ldots, v_k) \in \mathbb{R}^{d\times k} \text{ full rank}, s.t. \ M_i V = V D_i, \ D_i \in \mathcal{M}_{\text{diag}}(k). \quad (2)$$

In a series of contributions Flury [1984, 1986], Flury and Gautschi [1986] introduced common principal component analysis (CPCA) that deals with the test and calculation of simultaneous factorization among different samples of positive definite symmetric matrices. Under the same assumptions, Schott [1999] developed the terminology partial CPCA (PCPCA) with verified test methods for partially identical eigenvectors. In contrast, we do not need to impose any structural assumptions on the pool of matrices like symmetry or positive semidefinitenss which are naturally provided by considering covariance matrices.

Related to simultaneous diagonalization, previous studies mainly focused on testing whether the eigenvectors or eigenspaces of the population covariance matrix are equal to some given ones; see Tyler [1981], Koltchinskii and Lounici [2017], Silin and Spokoiny [2018], Naumov et al. [2019], Silin and Fan [2020]. Schwartzman et al. [2008] studied some related statistical tests about eigenvalues and eigenvectors of Gaussian random symmetric matrices with some pre-fixed algebraic restrictions. Especially, as stated in Schwartzman et al. [2008], the test for equality of eigenvectors with unknown eigenvalues between two sets of samples is rather difficult since no closed forms of estimations are available.

From a computational perspective, optimization routines for symmetric matrices were proposed by Fujioka [1993], Ghazi et al. [2008, 2014]. For general asymmetric matrices, some previous ideas, like '*sh-rt*' by Fu and Gao [2006], '*JUST*' by Iferroudjene et al. [2009], '*JDTM*' by Luciani and Albera [2010], and '*(W)JDTE*' by André et al. [2020] are shown to be numerically effective and ready for implementation. Colombo and Vlassis [2016a,b] focused on the joint Schur-decomposition and provided theoretical properties

of their proposed algorithms. Since joint Schur-decomposability fails to be a sufficient condition for simultaneous diagonalization, our work expands those ideas and provides an algorithm which estimates partially common eigenvectors across samples.

Possibly asymmetric matrix-valued statistics are broadly utilized in estimating the mean of random matrices, the adjacency matrices of weighted directed graphs, the coefficient matrices in linear regressions, factor models and vector autoregression (VAR) models, and transition probability matrices. However, most of the analysis has focused on studying the eigenvalues of those statistics. Analyses of eigenvalues include reduced rank estimation [Robin and Smith, 2000, Kleibergen and Paap, 2006, Donald et al., 2007], testing for cointegration [Engle and Granger, 1987, Johansen, 1991, Vogelsang, 2001, Zhang et al., 2019] and the eigenvalues of adjacency matrices [Restrepo et al., 2007, Paul and Chen, 2020]. In contrast, our applications (see Section 1.1 below) give a new perspective on the usefulness of studying the eigenvectors in various models.

The literature review shows that existing work is based on covariance matrices which are surely diagonalizable with orthogonal eigenvectors. The eigenstructure problem lacks analysis in some more general cases like asymmetric matrices in particular. The breakthrough point of our work is to design and validate efficient diagonalization test methods for possibly asymmetric matrices. Due to non-linearity of eigenproperties and the lack of closed eigensolutions, our investigation about the random eigenstructures with less restricted conditions is algebraically difficult, and our exploration is novel.

1.1 Applications

From a statistical perspective, joint diagonalizability provides valuable information. Suppose one fails to reject the null hypothesis of common eigenvectors, then, it is reasonable to only analyze the eigenvalues, which reduces the problem's complexity significantly. We will

illustrate the usefulness of our results with two relevant examples, namely the coefficient matrices in VAR models and the transition matrices of Markov chains.

The coefficient matrices in VAR models appear to be general matrices without restrictions like symmetry. For the same multivariate time series of comparable objects, simple VAR models of order one may share common components in regression which can be verified by a joint eigendecomposition of coefficient matrices. With successful verification of simultaneous diagonalizability, one can decouple multivariate time series into multiple univariate ones and conduct comparison conveniently; see Section 7.1 for more details.

Another motivating example for our tests are the transition matrices of Markov chains which are usually asymmetric. Furthermore, the leading left eigenvector of a transition matrix corresponds to eigenvalue one and represents the stationary distribution of the chain. Testing the equality of the leading eigenvectors of the transition matrices from multiple Markov chains gives information whether these chains share similar properties though differing in their transition dynamics. For instance, the Markov chains of the same object but with different time resolutions might exhibit a common stationary distribution; see also Section 7.2 for more details on this application.

1.2 Organization

The rest of this paper is organized as follows. Section 2 establishes notation and gives some preliminary results. Our main work starts in Section 3 from a two-sample test. In addition to simultaneously conducting two-sample tests pairwise for multi-sample cases, we design our test method based on the pooled estimator of common eigenvectors; see Section 4. In Section 5 we further extend our results to a partial version with a novel algorithm to estimate the subset of eigenvectors that is shared across samples. In Sections 6 and 7 we conduct a simulation study and experiment with some real data examples, respectively.

The supplementary material provides some empirical results complementary to the numerical analysis presented in the main paper; see Appendix A. Furthermore, an alternative approach for the two sample test is presented in Appendix B. We also briefly show the compatibility of our test methods with the symmetric setting in Appendix C. Finally, the proofs of the theoretical results in Sections 3 to 5 and Appendix B can be found in the supplementary material in Appendices D and E.

2 Preliminaries

2.1 Notation

Throughout this paper, p is the number of matrices to be tested, n denotes the sample size for estimation, and d is the dimension of the square matrices. Notation $\xrightarrow{\mathcal{D}}$ represents convergence in distribution, $\xrightarrow{\mathcal{P}}$ convergence in probability, and $\overset{\mathcal{D}}{\approx}$ an approximation of random distributions. The operator $\otimes$ denotes the Kronecker product between two matrices. For a matrix A, the operator $\mathrm{rk}(A)$ denotes the rank of A, $\mathrm{vec}(A)$ transforms A into a vector form by stacking all its columns, and A^+ is the Moore-Penrose general inverse of A. For a square matrix A of dimension d, we write $\mathrm{tr}(A)$ for the trace function of A and the operator $\mathrm{mat}_d(\cdot)$ is the inverse of $\mathrm{vec}(\cdot)$ such that $\mathrm{mat}_d\big(\mathrm{vec}(A)\big) = A$. The matrix I_d represents the d-dimensional identity matrix, and the function $\mathrm{blkdiag}(\{X_i\}_{i=1}^p)$ returns a block-diagonal matrix with the sub-matrices on the diagonal to be the input list of matrices $\{X_i\}_{i=1}^p$. We write $\mathcal{N}(\mu, \Sigma)$ for the multivariate normal distribution with mean vector μ and covariance matrix Σ, $\chi^2(k)$ for the chi-squared distribution with k degrees of freedom, and $\gamma(\alpha, \beta)$ for the Gamma distribution with shape parameter α and rate parameter β.

2.2 Assumptions

In this section, we give the required assumptions for future proof of the asymptotic results for our proposed test statistics.

Assumption 1. *The deterministic matrices* $\{M_i\}_{i=1}^p$ *can be estimated by mutually independent estimators* $\{A_{i,n}\}_{i=1}^p$ *from* n *samples, satisfying,*

$$c(n)\,\mathrm{vec}(A_{i,n} - M_i) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_i) \quad \textit{for} \quad i = 1, \ldots, p, \tag{3}$$

with $c(n) \to \infty$ *as* $n \to \infty$.

Assumption 2. *The limiting covariance matrices* Σ_i *in* (3) *can be estimated consistently by* $\widehat{\Sigma}_{i,n}$, *that is,*

$$\widehat{\Sigma}_{i,n} - \Sigma_i = O_{\mathcal{P}}\big(\sigma(n)\big) \quad \textit{for} \quad i = 1, \ldots, p, \tag{4}$$

with rate $\sigma(n) \to 0$ *as* $n \to \infty$.

Note that the estimators $A_{i,n}$ and $\widehat{\Sigma}_{i,n}$ also depend on the sample size n, but for notational simplicity we will omit it in later expressions and simplify as $A_i = A_{i,n}$ and $\widehat{\Sigma}_i = \widehat{\Sigma}_{i,n}$. In addition, we assume the sequence $c(n)$ to be the same for all $i = 1, \ldots, p$, in the following analysis but the extension to the general case is straightforward. Furthermore, we assume the following:

Assumption 3. *Each* M_i, $i = 1, \ldots, p$ *is diagonalizable with real eigenvalues.*

The theoretical results of this paper require no a priori assumptions on the rank of the limiting covariance matrices in Assumption 1. In particular, the matrices Σ_i, $i = 1, \ldots, p$, in (3) may be less than full rank. On one hand, this allows for flexibility in the choice of estimators $A_{i,n}$, on the other hand it elevates the difficulties in deriving asymptotic results

for our test statistics. In particular, we aim to give tractable versions of our test statistics in the sense that the covariance matrices can be estimated. While Assumption 2 ensures the existence of a consistent estimator, one still has to address a potential singularity. We refer to Section 2.3 for a discussion and workaround.

2.3 Covariance estimation

Our test statistics involve the inverses and ranks of the limiting covariance matrices in Assumption 1. Ideally, given exact covariance matrices, we can use the so-called Moore-Penrose inverse to address possible singularity. In order to make the statistics tractable in practice, we need to instead incorporate their consistent estimates as given by Assumption 2. However, neither the Moore-Penrose inverse nor the rank of a matrix are continuous.

To circumvent those issues, we introduce the so-called truncated singular value decomposition following Lütkepohl and Burda [1997]. For an arbitrary matrix $\Psi \in \mathbb{R}^{a\times b}$ with $\tau = \min(a, b) > 1$, its singular value decomposition (SVD) is given by $\Psi = U\Pi W'$ with orthogonal singular vectors U and W, and non-increasing, non-negative singular values $\Pi = \operatorname{diag}(\varpi_1, \ldots, \varpi_\tau)$. We define its truncated singular value decomposition with respect to a threshold $\varepsilon \geqslant 0$ as

$$\Psi(\varepsilon) = U\Pi(\varepsilon)W' \quad \text{with} \quad \Pi(\varepsilon) = \operatorname{diag}(\varpi_1\mathbb{I}(\varpi_1 > \varepsilon), \ldots, \varpi_\tau\mathbb{I}(\varpi_\tau > \varepsilon)) \tag{5}$$

with indicator function $\mathbb{I}(\cdot)$. In addition, we denote the Moore-Penrose general inverse as $\Psi^+(\varepsilon) = \big(\Psi(\varepsilon)\big)^+$ and the rank function $\operatorname{rk}(\Psi;\varepsilon) = \operatorname{rk}\big(\Psi(\varepsilon)\big)$.

With the help of the truncated SVD (5), we introduce the following lemma that gives consistent estimates for the Moore-Penrose inverse and the rank of a consistently estimated matrix.

Lemma 2.1. *Assume $\widehat{\Sigma}$ is a consistent estimator of a positive semidefinite matrix $\Sigma \in \mathbb{R}^{\tau\times\tau}$, $\varepsilon > 0$ is a constant and not an eigenvalue of Σ. Then, with $\Sigma(\varepsilon)$ and $\widehat{\Sigma}(\varepsilon)$ defined to be the corresponding truncated SVDs (5),*

$$\widehat{\Sigma}(\varepsilon) \xrightarrow{\mathcal{P}} \Sigma(\varepsilon), \;\; \widehat{\Sigma}^{+}(\varepsilon) \xrightarrow{\mathcal{P}} \Sigma^{+}(\varepsilon), \;\; \mathrm{rk}(\widehat{\Sigma};\varepsilon) \xrightarrow{\mathcal{P}} \mathrm{rk}(\Sigma;\varepsilon). \tag{6}$$

The proof of Lemma 2.1 can be found in Appendix D. Results in the manner of Lemma 2.1 are used to circumvent singularity issues which usually occur under the usage of Wald type tests; see Hadi and Wells [1990], Ratsimalahelo [2001]. The following remark comments on the choice of the threshold ε in (6).

Remark 2.1. *With the additional Assumption 2 that $\widehat{\Sigma}$ converges with rate $O_{\mathcal{P}}\big(\sigma(n)\big)$, the threshold ε that satisfies $\varepsilon = o(1)$ and $\sigma(n)/\varepsilon = o(1)$ as $n \to \infty$ can be chosen to optimize the accuracy of the generalized inverse and rank estimators. See p. 320 in Lütkepohl and Burda [1997] for a discussion on the choice of ε.*

3 Two-sample test

We start from a two-sample test $p = 2$ and design a test statistic based on the commutator of the two matrices under consideration. Under Assumption 3, matrices commute if and only if they can be diagonalized simultaneously; see Theorem 1.3.12 in Horn and Johnson [2012]. Hence, one intuitive idea to measure how far M_1 and M_2 are from being commutable is, to calculate some form of metric of their commutator $[M_1, M_2] := M_1M_2 - M_2M_1$. The following proposition introduces a statistic to test the hypothesis (1) and provides its asymptotic behavior.

Proposition 3.1. *Suppose Assumptions 1 and 3 are satisfied and denote $\boldsymbol{\eta}_n = \mathrm{vec}[A_1, A_2]$.*

Then, under H_0 *in* (1),

$$c(n)\boldsymbol{\eta}_n \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_\eta) \tag{7}$$

with $\Sigma_\eta = \Sigma_{1,2} + \Sigma_{2,1}$, *where* $\Sigma_{k,\ell} = \Lambda(M_\ell)\Sigma_k\Lambda'(M_\ell)$ *for* $k \neq \ell$, *and* $\Lambda(X) = I_d \otimes X - X' \otimes I_d$ *is a function in* $X \in \mathbb{R}^{d\times d}$. *Then,*

$$\Gamma_1 := c^2(n)\boldsymbol{\eta}_n'\Sigma_\eta^+\boldsymbol{\eta}_n \xrightarrow{\mathcal{D}} \chi^2(r_1), \tag{8}$$

where r_1 *is the rank of* Σ_η.

The proof of Proposition 3.1 can be found in Appendix D. In order to make Proposition 3.1 tractable in practice, we can obtain a consistent estimator $\widehat{\Sigma}_\eta$ by substituting Σ_i, M_i, $i = 1, 2$, in the expression of Σ_η with $A_i, \widehat{\Sigma}_i$, $i = 1, 2$, in Assumptions 1 and 2, respectively. The consistency is verified by the continuous mapping theorem and Assumptions 1 and 2 as

$$\widehat{\Sigma}_\eta - \Sigma_\eta = \Lambda(A_2)\widehat{\Sigma}_1\Lambda'(A_2) + \Lambda(A_1)\widehat{\Sigma}_2\Lambda'(A_1) - \Sigma_\eta \xrightarrow{\mathcal{P}} 0.$$

Note that Σ_η and $\widehat{\Sigma}_\eta$ are both singular matrices as there exists at least one non-trivial vector $\boldsymbol{v} = \mathrm{vec}(I_d)$ such that $\Sigma_\eta\boldsymbol{v} = \widehat{\Sigma}_\eta\boldsymbol{v} = 0$, since

$$\Lambda(X)\boldsymbol{v} = (I_d \otimes X)\boldsymbol{v} - (X' \otimes I_d)\boldsymbol{v} = 0,$$

where the last equality follows by Theorem 2 in Magnus and Neudecker [2019], p. 35. Hence r_1 and $\hat{r}_1$, the ranks of Σ_η and $\widehat{\Sigma}_\eta$ respectively, are always less than d^2. Due to the singularity issue in (8), we propose to use the truncated version (6) of $\widehat{\Sigma}_\eta$.

Proposition 3.2. *Suppose Assumptions 1, 2 and 3 are satisfied. Then, for a given threshold* $\varepsilon > 0$ *that is not an eigenvalue of* Σ_η *(defined through* (7)*), the test statistic is defined*

as and satisfies

$$\Gamma_1^{\#}(\varepsilon) := c^2(n)\boldsymbol{\eta}_n'\widehat{\Sigma}_\eta^{+}(\varepsilon)\boldsymbol{\eta}_n \xrightarrow{\mathcal{D}} \chi^2(\widehat{r}_1(\varepsilon)), \tag{9}$$

where $\widehat{r}_1(\varepsilon) = \mathrm{rk}(\widehat{\Sigma}_\eta; \varepsilon)$.

4 Multi-sample test

An extension of the topic introduced in Section 3 is to conduct the test on a larger pool of matrices ($p \geqslant 2$). In addition to testing simultaneously over all pairs of samples using the methods based on the two-sample test introduced in Section 3, we are more interested in whether the same hypothesis holds across the whole pool of matrices. With that in mind, we propose to use the estimated optimal common eigenvectors and test whether they annihilate the off-diagonal elements of the matrices after transformation.

For the simultaneous test, with the commutator-based test developed in Section 3, the approach and the test statistic are straightforward. For instance, a matrix of test statistics (or p-values) represents the pairwise test results and conclusions can then be drawn. Hence, we will omit further details here and focus on the more comprehensive approach regarding whether the same hypothesis holds across the whole pool of matrices. Specifically, we will refer to optimization algorithms for calculating the common eigenvectors that almost diagonalize a pool of matrices (see Section 4.1) and then design a test statistic (see Section 4.2).

4.1 Common eigenvectors finder

In this section, we briefly introduce the setup of the optimization problem of finding the optimal diagonalizer for a pool of matrices.

Recall our setup, with a pool of matrix-valued statistics $\mathcal{A} = \{A_i\}_{i=1}^p$ estimating the matrices $\mathcal{M} = \{M_i\}_{i=1}^p$. We refer to the '*(W)JDTE*' algorithm by André et al. [2020] due

to its performance in terms of speed and accepted accuracy. The algorithm provides the common eigenvectors of a pool of matrices by minimizing the objective function

$$\text{off}(U;\mathcal{A}) = \sum_{i=1}^{p} \text{off}_2(U^{-1}A_iU). \tag{10}$$

Here, $\text{off}_2(X) = \sum_{i\neq j} |X_{i,j}|^2$ denotes the off-diagonal sum-of-squares for $X \in \mathbb{R}^{d\times d}$. For the details of the algorithm; see André et al. [2020].

4.2 Eigenvector test

In this section, we propose a test for H_0 in (1) allowing the number of matrices p to be larger than two. The test is based on the assumption that an invertible matrix $V \in \mathbb{R}^{d\times d}$ is given as a guess for the common eigenvector matrix. We acknowledge that the assumption of knowing the matrix V is quite restrictive. For all practical purposes, we refer to the optimization problem in Section 4.1 which provides an optimal diagonalizer based on the matrix estimators $\mathcal{A} = \{A_i\}_{i=1}^p$. From a theoretical perspective, existing literature does not suggest any formal testing procedure based on explicit estimators for V. In particular, there are no asymptotic results, neither under the assumption that the matrices to be tested are asymmetric nor symmetric. For this reason, we see the below stated theoretical results under the assumption that V is known as a starting point and leave a more rigorous investigation for future work.

Define the function $\text{offvec}_d : \mathbb{R}^{d\times d} \to \mathbb{R}^{d(d-1)}$, $\text{offvec}_d(X) = S_d\,\text{vec}(X)$, which stacks all off-diagonal elements of a square matrix X with dimension d columnwise. We will always use S_d as the off-diagonal selection matrix for square matrices of dimension d. The test can then be designed as follows.

Proposition 4.1. *Suppose Assumptions 1 and 3, and V is given as a guess for the common*

eigenvector matrix. Let

$$\boldsymbol{\zeta}_i = \mathrm{offvec}_d(V^{-1}A_iV), \;\; i = 1, \ldots, p, \;\; \textit{and} \;\; S_{V,d} = S_d(V' \otimes V^{-1}). \tag{11}$$

Then, under H_0 *in* (1), $c(n)\boldsymbol{\zeta}_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Theta_i)$ *with* $\Theta_i = S_{V,d}\Sigma_i S'_{V,d}$ *for* $i = 1, \ldots, p$. *The test statistic is defined as and satisfies*

$$\Gamma_3 := c^2(n) \sum_{i=1}^{p} \boldsymbol{\zeta}_i' \Theta_i^+ \boldsymbol{\zeta}_i \xrightarrow{\mathcal{D}} \chi^2(r_3), \tag{12}$$

where $r_3 = \sum_{i=1}^{p} \mathrm{rk}(\Theta_i)$.

Note that in practice, the eigenvector matrix to be tested is always obtained from optimization, and hence this idea highly depends on the accuracy of such algorithms. In order to reduce the influence of estimation errors, we develop the following analogous test that tolerates relatively larger errors while maintaining acceptable efficiency.

Corollary 4.1. *Under the assumptions of Proposition 4.1, let* $\Theta = \mathrm{blkdiag}(\{\Theta_i\}_{i=1}^p)$ *be a block-diagonal matrix, and* $\boldsymbol{\zeta} := \left(\boldsymbol{\zeta}_1', \ldots, \boldsymbol{\zeta}_p'\right)' \in \mathbb{R}^{p(d^2-d)}$, *with* $\boldsymbol{\zeta}_i$ *as in* (11) *such that* $c(n)\boldsymbol{\zeta} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Theta)$. *Then, the test statistic is defined as and satisfies*

$$\Gamma_3^* := c^2(n) \sum_{i=1}^{p} \|\boldsymbol{\zeta}_i\|^2 \xrightarrow{\mathcal{D}} \psi_3^* := \sum_{r=1}^{p(d^2-d)} \lambda_r(\Theta)\chi^2(1), \tag{13}$$

where $\lambda_r(\Theta)$ *denotes the* r*th eigenvalue of* Θ. *Furthermore, the p-value based on* (13) *can be approximated by*

$$\mathbb{P}(\psi_3^* > \Gamma_3^* \mid H_0) \approx \mathbb{P}(\gamma_3^* > \Gamma_3^* \mid H_0), \;\; \textit{where} \;\; \gamma_3^* \sim \mathit{Gamma}\Big(\frac{\mathrm{tr}(\Theta)^2}{2\,\mathrm{tr}(\Theta^2)}, \frac{\mathrm{tr}(\Theta)}{2\,\mathrm{tr}(\Theta^2)}\Big). \tag{14}$$

The approximation in (14) between ψ_3^*, a weighted sum of chi-squared distributed ran-

dom variables, and γ_3^*, a gamma distribution, was introduced in Box [1954, Theorem 3.1] based on matching first- and second-order moments of the two distributions. For accuracy analysis, see Bodenham and Adams [2016].

Note, that Proposition 4.1 requires to calculate the generalized inverse of Θ. In contrast, Corollary 4.1 only needs the trace of Θ. Calculating the trace of Θ is expected to be more robust towards estimation errors of the eigenvector matrix V than finding the generalized inverse of Θ.

In practice, only the estimated $\widehat{\Sigma}_i$ are accessible. Due to possible singularity issues with the general inverse Θ_i^+ in Proposition 4.1, we use results in Appendix E to introduce the following proposition.

Proposition 4.2. *Suppose Assumptions 1, 2 and 3. Then, for a given threshold $\varepsilon > 0$ that is not an eigenvalue of Θ_i for $i = 1, \ldots, p$, the test statistic is defined as and satisfies*

$$\Gamma_3^{\#}(\varepsilon) := c^2(n) \sum_{i=1}^{p} \boldsymbol{\zeta}_i' \widehat{\Theta}_i^+(\varepsilon) \boldsymbol{\zeta}_i \xrightarrow{\mathcal{D}} \chi^2(\widehat{r}_3(\varepsilon)), \tag{15}$$

where $\widehat{\Theta}_i = S_{V,d} \widehat{\Sigma}_i S_{V,d}$ and $\widehat{r}_3(\varepsilon) = \sum_{i=1}^{p} \mathrm{rk}(\widehat{\Theta}_i; \varepsilon)$.

5 Partial test

In this section, we focus on the hypothesis H_0^* in (2). We first reformulate the hypothesis testing problem (Section 5.1) and then design the corresponding test statistics and analyze their asymptotic behavior (Section 5.2). Meanwhile, an optimization algorithm is also proposed to approximate the matrix V in the statement of H_0^* (Section 5.3).

5.1 Problem representation

Suppose $\{M_i\}_{i=1}^p$ satisfy H_0^*, then for V specified in the hypothesis, there exists an orthogonal $d \times k$ matrix Q_k (i.e. $Q_k' Q_k = I_k$), such that $V = Q_k R$, where R is a $k \times k$ upper-triangular matrix. By orthogonally spanning $Q_k = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_k)$ to $Q = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_d) \in \mathbb{R}^{d \times d}$, we have

$$Q' M_i Q = \begin{pmatrix} R D_i R^{-1} & * \\ 0 & ** \end{pmatrix} =: \begin{pmatrix} \Phi_i & * \\ 0 & ** \end{pmatrix} \tag{16}$$

with D_i, $i = 1, \ldots, p$, as in (2). We use the symbols $*$ and $**$ to denote non-zero block matrices with proper dimensions which are not relevant. Note that the set of upper-triangular matrices $\{\Phi_i\}_{i=1}^p$ shares the common eigenvectors R. For the estimators $\mathcal{A} = \{A_i\}_{i=1}^p$ given by Assumption 1, suppose a matrix $\widehat{Q}$ is given as a guess for Q, such that

$$\widehat{Q}' A_i \widehat{Q} = \begin{pmatrix} B_i & * \\ C_i & ** \end{pmatrix}, \tag{17}$$

where $B_i \in \mathbb{R}^{k \times k}$ and $C_i \in \mathbb{R}^{(d-k) \times k}$. The matrix $\widehat{Q}$ can either be given by knowing the ground-truth Q or by estimation; see Section 5.3 for an algorithm to estimate Q. The procedure of our test starts from finding the orthogonal matrix Q that contains the first k columns as the common invariant subspace, and designing tests on the transformed matrices $\{B_i, C_i\}_{i=1}^p$. The test is a combination of:

1. Whether the transformation by the orthogonal matrix $\widehat{Q}$ introduces the all-zero lower-left block in (16) by testing on $\{C_i\}_{i=1}^p$.

2. Whether the $k \times k$ matrices $\{\Phi_i\}_{i=1}^p$ defined in (16) satisfy H_0 in (1) by testing on $\{B_i\}_{i=1}^p$.

To introduce our test statistic, suppose $\widehat{Q}$ is given, and let

$$\widehat{Q}_{BC} = \begin{pmatrix} S_B \\ S_C \end{pmatrix} (\widehat{Q}' \otimes \widehat{Q}') \quad \text{with} \quad \begin{aligned} S_B \text{vec}(\widehat{Q}' A_i \widehat{Q}) &= \text{vec}(B_i), \\ S_C \text{vec}(\widehat{Q}' A_i \widehat{Q}) &= \text{vec}(C_i), \end{aligned}$$

where S_B and S_C are selection matrices defined according to (17). Hence, the partially common eigenvectors of $\{M_i\}_{i=1}^p$ can be estimated as $\widehat{V} = \widehat{Q}_k \widetilde{V}$, where $\widehat{Q}_k$ is a matrix of the first k columns of $\widehat{Q}$ and $\widetilde{V}$ the estimated common eigenvectors of $\{B_i\}_{i=1}^p$. Referring to Section 4.1, $\widetilde{V}$ can be received from '*(W)JDTE*'. So far, we have supposed that $\widehat{Q}_k$ is a guess. However, it can be estimated by an algorithm proposed in Section 5.3 below.

5.2 Partial eigenvector test

In this section, we introduce our test statistic to test for partially common eigenvectors and present their asymptotic behavior.

Proposition 5.1. *Suppose Assumptions 1 and 3 are satisfied and let*

$$\boldsymbol{w}_i = P_w \, \text{vec}(A_i) \quad \textit{with} \quad P_w := \begin{pmatrix} S_{\widetilde{V},k} & 0 \\ 0 & I_{k(d-k)} \end{pmatrix} Q_{BC}, \tag{18}$$

where $S_{\widetilde{V},k}$ is defined as in (11) and Q_{BC} is analogous to $\widehat{Q}_{BC}$ except replacing the guess $\widehat{Q}$ with true Q. Then, under H_0 in (1), $c(n)\boldsymbol{w}_i \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Omega_i)$ with $\Omega_i = P_w \Sigma_i P_w'$ for $i = 1, \ldots, p$. The test statistic is defined as and satisfies

$$\Gamma_4 := c^2(n) \sum_{i=1}^{p} \boldsymbol{w}_i' \Omega_i^+ \boldsymbol{w}_i \xrightarrow{\mathcal{D}} \chi^2(r_4), \tag{19}$$

where $r_4 = \sum_{i=1}^p \text{rk}(\Omega_i)$.

And similar to Corollary 4.1, we design the following test statistic with p-value approx-

imated by a gamma distribution.

Corollary 5.1. *Under the assumptions in Proposition 5.1, let* $\Omega = \mathrm{blkdiag}(\{\Omega_i\}_{i=1}^p)$ *be a block-diagonal matrix, and* $\boldsymbol{w} := \left(\boldsymbol{w}_1', \ldots, \boldsymbol{w}_p'\right)' \in \mathbb{R}^{pk(d-1)}$, *with* $\boldsymbol{w}_i$ *as in (18) such that* $c(n)\boldsymbol{w} \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Omega)$. *Then the test statistic is defined as and satisfies*

$$\Gamma_4^* := c^2(n)\|\boldsymbol{w}\|^2 \xrightarrow{\mathcal{D}} \psi_4^* := \sum_{r=1}^{pk(d-1)} \lambda_r(\Omega)\chi^2(1), \tag{20}$$

where $\lambda_r(\Omega)$ *denotes the* r*th eigenvalue of* Ω. *Furthermore, the p-value based on (20) can be approximated by*

$$\mathbb{P}(\psi_4^* > \Gamma_4^* \mid H_0) \approx \mathbb{P}(\gamma_4^* > \Gamma_4^* \mid H_0), \quad \textit{where } \gamma_4^* \sim \mathit{Gamma}\Big(\frac{\mathrm{tr}(\Omega)^2}{2\,\mathrm{tr}(\Omega^2)}, \frac{\mathrm{tr}(\Omega)}{2\,\mathrm{tr}(\Omega^2)}\Big). \tag{21}$$

Similar as Proposition 4.2, we use results in Appendix E for the following tractable version of Proposition 5.1 that takes care of possible singularity issues with Ω_i^+ in (19).

Proposition 5.2. *Assume Assumptions 1, 2 and 3 are satisfied. Then, for a given threshold* $\varepsilon > 0$ *that is not an eigenvalue of* Ω_i *for* $i = 1, \ldots, p$, *the test statistic is defined as and satisfies*

$$\Gamma_4^{\#}(\varepsilon) := c^2(n) \sum_{i=1}^{p} \boldsymbol{w}_i' \widehat{\Omega}_i^+(\varepsilon) \boldsymbol{w}_i \xrightarrow{\mathcal{D}} \chi^2(\widehat{r}_4(\varepsilon)), \tag{22}$$

where $\widehat{\Omega}_i = P_w \widehat{\Sigma}_i P_w'$ *for* $i = 1, \ldots, p$, *and* $\widehat{r}_4(\varepsilon) = \sum_{i=1}^p \mathrm{rk}(\widehat{\Omega}_i; \varepsilon)$.

5.3 Optimization algorithm

This section is dedicated to finding an estimator $\widehat{Q}$ for Q in (16). In implementation, $\widehat{Q}$ can be obtained by minimizing the following objective function

$$f(Q; \mathcal{A}, k) = \sum_{i=1}^{p} \sum_{r=k+1}^{d} \sum_{s=1}^{k} (\boldsymbol{q}_r' A_i \boldsymbol{q}_s)^2 \quad \text{subject to } \boldsymbol{q}_r' \boldsymbol{q}_s = \delta_{rs}, \ \forall\ r, s = 1, \ldots, d,$$

where $Q = (\boldsymbol{q}_1, \ldots, \boldsymbol{q}_d) \in \mathbb{R}^{d\times d}$. Colombo and Vlassis [2016b] introduced a version of Gauss-Newton algorithm for the joint Schur decomposition based on matrix exponential, and showed its global minimum guarantees if the initial value is sufficiently close to the ground-truth Q. In our work, we inherit the idea of this algorithm with slight revision, where the major difference is to substitute the selection matrix from lower-triangular indicator to the $(d-k)\times k$ lower-left block indicator.

We also introduce a warm-up algorithm to supply the initial values for this Gauss-Newton approach. Since the matrices $\{\Phi_i\}_{i=1}^p$ in (16) are upper-triangular, we can split the minimization with respect to Q into sequentially optimizing each column $\boldsymbol{q}_r$ with r from 1 to k based on the following objective function

$$f^*(Q;\mathcal{A},k) = \sum_{i=1}^{p}\sum_{s=1}^{k}\sum_{r=s+1}^{d}(\boldsymbol{q}_r' A_i \boldsymbol{q}_s)^2, \text{ with } Q = (\boldsymbol{q}_1,\ldots,\boldsymbol{q}_d) \in \mathbb{R}^{d\times d}.$$

More precisely, we introduce the following Algorithm 1 as the whole process for optimizing orthogonal $\widehat{Q}$, including the initialization warm-up before line 7.

Algorithm 1: Estimation of Q

Input: Estimators $\mathcal{A} = \{A_i\}_{i=1}^p$, number of common eigenvectors k.
Output: Estimator $\widehat{Q}$ for Q.

1 Initialize $\mathcal{O}_0 = I_d$ as an identity matrix.
2 **for** $j = 1 : k$ **do**
3 Optimize $\mathcal{O}_j = \operatorname{argmin}_Q f(Q;\mathcal{A},1)$ and write $\mathcal{O}_j = (\boldsymbol{p}_j, P_j)$, where $\boldsymbol{p}_j \in \mathbb{R}^{d-j+1}$ and $P_j \in \mathbb{R}^{(d-j+1)\times(d-j)}$.
4 Update $\mathcal{O}_{j-1} = \mathcal{O}_{j-1}\mathcal{O}_j$.
5 Update $\mathcal{A}$ by $A_i = P_j' A_i P_j$ for $i = 1,\ldots,p$.
6 **end**
7 Input $\mathcal{O}_k$ into the modified version of Gauss-Newton approach by Colombo and Vlassis [2016b] as the initial value and get the final output $\widehat{Q}$.

For realization of the optimization on line 3, we refer to the FG-algorithm by Flury and Gautschi [1986].

6 Simulation study

Simulation studies are run for two-sample, multi-sample, and partial tests. The barplots of p-values from multiple replicates, especially, the type I and type II errors which are readable from the plots of histograms, are given as evidence for the effectiveness of our test designs. Complementary to the plots presented in this section, we refer the reader to Section A.1 in the supplementary material for tables providing type I and II errors for all our simulation studies.

In addition, instead of implementing Propositions 3.1, 4.1 and 5.1, we always turn to their truncated versions (Propositions 3.2, 4.2 and 5.2 respectively) in Sections 6 and 7, in order to take care of possibly singular covariance matrices. Furthermore, since the following examples with sample size n all have the same rate of convergence $\sigma(n) = n^{-1/2}$ for their limiting covariance matrix estimators, we will not further specify but choose $\varepsilon = n^{-1/3}$ by default whenever applicable in implementations. This choice of ε is also in accordance with the discussion in Remark 2.1 and maintains the test power.

6.1 Data generating process (DGP)

Prior to our testing analysis, we first introduce the data generating process (DGP) for our targeted matrices $\mathcal{M}_p(\rho, k; d) = \{M_i(\rho, k; d)\}_{i=1}^p$ with parameters ρ accounting for the signal-to-noise variance ratio (SNR $:= \frac{1}{\rho^2}$) and k the number of common eigenvectors.

With the set of target matrices $\mathcal{M}_p(\rho, k; d)$, we can then proceed to design different sampling setup for their corresponding consistent estimators $\mathcal{A}_p(\rho) = \{A_i(\rho)\}_{i=1}^p$. Note that we define SNR $= \frac{1}{\rho^2}$ to quantify the similarity among the underlying eigenvectors $\{V_i(\rho, k; d)\}_{i=1}^p$ (see line 4 of Algorithm 2) rather than the estimation accuracy for $\mathcal{A}_p(\rho)$, i.e., if ideally SNR $= \infty$ or $\rho = 0$, the matrices in $\mathcal{M}_p(\rho, k; d)$ can be perfectly diagonalized with common eigenvectors (partially when $k < d$), while the randomness of our test statistics

Algorithm 2: DGP for $\mathcal{M}_p(\rho, k; d)$

Input: Number of matrices $p \in \mathbb{N}$, matrix dimension $d \in \mathbb{N}$, number of common eigenvectors $k \in \{1, \ldots, d\}$, noise measure $\rho \geqslant 0$.

Output: Pool of matrices $\mathcal{M}_p(\rho, k; d)$ with common eigenvectors.

Randomly generate an eigenvector matrix $V(k; d) \in \mathbb{R}^{d \times k}$ $(k \leqslant d)$.

Set $V_i(k; d) = V(k; d)$ if $k = d$, otherwise span $V_i(k; d) = (V(k; d), \widetilde{V}_i) \in \mathbb{R}^{d \times d}$ with random but sufficiently distinct $\widetilde{V}_i \in \mathbb{R}^{d \times (d-k)}$ for $i = 1, \ldots, p$.

for $i = 1, \ldots, p$ **do**

    perturb the i-th eigenvector matrix $V_i(k; d)$ as $V_i(\rho, k; d) = V_i(k; d) + \rho E_i$ with noise E_i to be independent and standard normal, element-wise.

end

Randomly generate non-singular diagonal matrices $D_i \in \mathbb{R}^{d \times d}$ for $i = 1, \ldots, p$.

Generate the target matrices $\mathcal{M}_p(\rho, k; d) = \{V_i(\rho, k; d) D_i V_i^{-1}(\rho, k; d)\}_{i=1}^{p}$.

still exists due to the subsequent estimation for $\mathcal{A}_p(\rho)$.

6.2 Two-sample test

With $\mathcal{M}_2(\rho, d; d) = \{M_1(\rho, d; d), M_2(\rho, d; d)\}$ generated following Algorithm 2 in Section 6.1, we treat them as the mean matrices for two pools of n normally distributed observations and propose the averages of samples as the corresponding estimators $\mathcal{A}_2(\rho)$. The covariance matrices are consistently estimated as well. According to the central limit theorem (CLT), the convergence rate in Assumption 1 satisfies $c(n) = \sqrt{n}$. If not further specified, we stick to the classical p-value test framework and always reject the null H_0 (or H_0^*) when the p-value is below a 0.05 significance level.

We set the dimension of the matrices as $d = 5$, SNR $\in \{10, 1000, \infty\}$, and the sample size $n \in \{50, 250, 1000\}$. For each data-generating setting determined by (SNR, n), the Algorithm 2 is repeated 500 times independently to get the data of 500 pairs of estimators for the commutator-based test; see Figure 1 for performance. Note that the first bar in these figures represents the proportion of the 500 p-values lower than the critical threshold 0.05, i.e., the proportion of simulations that one rejects the null hypothesis H_0 based on the test design. Ideally for SNR $= \infty$, this proportion, also known as type I error rate, should

be close to 0.05, which is a typical choice of significance level. For SNR $\neq \infty$, the proportion of p-values outside the first bar, also known as type II error rate, should approach 0 if the test has high power.

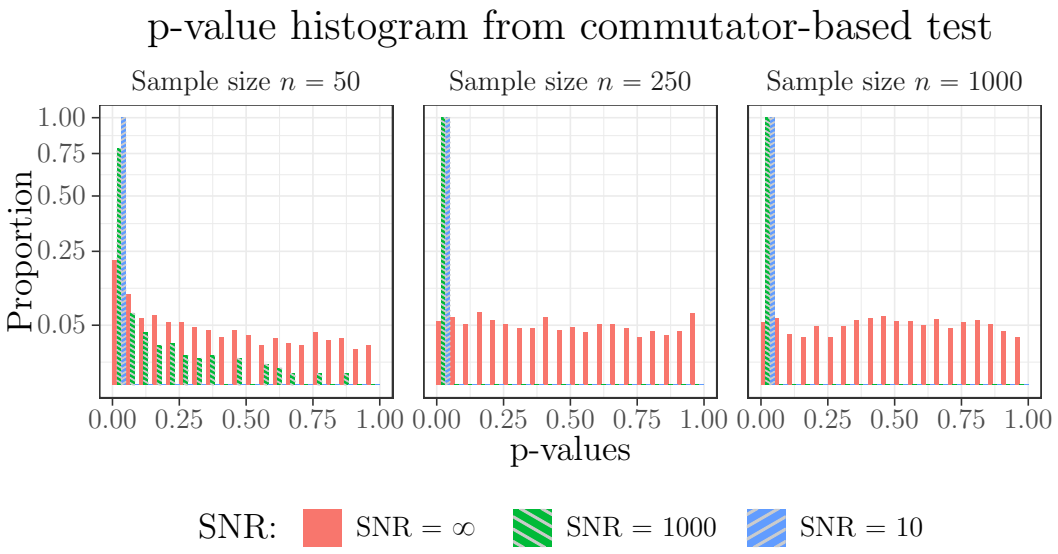

Figure 1: The histograms of p-values for the commutator-based test in Proposition 3.2 with simulated data.

For the commutator-based test Proposition 3.2, we see from Figure 1 that with sample size increasing, the p-values of samples from null space (SNR $= \infty$) tend to be uniformly distributed on the interval $[0, 1]$, and the p-values of samples from alternative spaces start to concentrate in the interval $[0, 0.05)$. When the sample size n exceeds a certain level, 200 for instance, the test performs well with acceptable type I error and excellent type II error.

6.3 Multi-sample test

For the simulation study of the multi-sample test introduced in Section 4, we follow Algorithm 2 to set the mean matrices of normally distributed observations as $\mathcal{M}_p(\rho, d; d)$. We consider here the setting that dimension $d = 4$, $p = 8$, SNR $= \frac{1}{\rho^2} \in \{10, 100, 1000, \infty\}$, and sample size $n \in \{10^2, 10^3, 10^4, 10^5\}$. Each consistent estimate in $\mathcal{A}(\rho)$ is the empirical average from the n observations, hence $c(n) = \sqrt{n}$ by the CLT. Note that when $\rho = 0$,

equivalently SNR $= \infty$, the matrices in $\mathcal{A}_p(\rho)$ satisfy H_0, i.e., share common eigenvectors.

To verify the testing efficiency, we (i) use the exact $V = V(d; d)$ (see line 1 of Algorithm 2) to test on $\mathcal{A}_p(0)$ according to Proposition 4.2, and (ii) use the optimized $\widehat{V}$ from algorithm '*(W)JDTE*' and implement both Proposition 4.2 and Corollary 4.1. We present the test power versus SNR and sample size n through histogram plots; see Figure 2.

With evidence that the p-values are almost uniformly distributed by using exact V, the test of Proposition 4.2 is shown to be effective when the supplied common eigenvectors are sufficiently accurate. However, for estimated V, one can observe that for small p-values the proportion of rejections exceeds the nominal test size of 0.05 for the chi-squared test in Proposition 4.2. While the gamma test in Corollary 4.1 admits almost the opposite behavior subceeding the nominal test level. From a theoretical perspective that might also be due to the singular sensitivities when inverting the estimated covariance matrices to calculate the test statistic in Proposition 4.2.

As for Corollary 4.1, it is a reasonably efficient test method as the false positive rate (type I error) remains at a relatively low level regardless of the sample size n, while the false negative rate (type II error) is following a reasonable pattern that with higher SNR, i.e. less perturbations to the shared eigenvectors, the test is ultimately able to reject the H_0 as the sample size n increases to be sufficiently large.

6.4 Partial test

For the simulations regarding the partial test in Section 5, we fix $k < d$, and generate the mean matrices as $\mathcal{M}_p(\rho, k; d)$ following Algorithm 2. Consistent estimates $\mathcal{A}_p(\rho)$ are obtained similarly from the average of the random observations to follow ordinary CLT. We set $d = 4$, $p = 8$, $k = 2$, SNR $= \frac{1}{\rho^2} \in \{10, 100, 1000, \infty\}$, and sample size $n \in \{10^2, 10^3, 10^4\}$.

The histograms in Figure 3 show that both Proposition 5.2 and Corollary 5.1 have

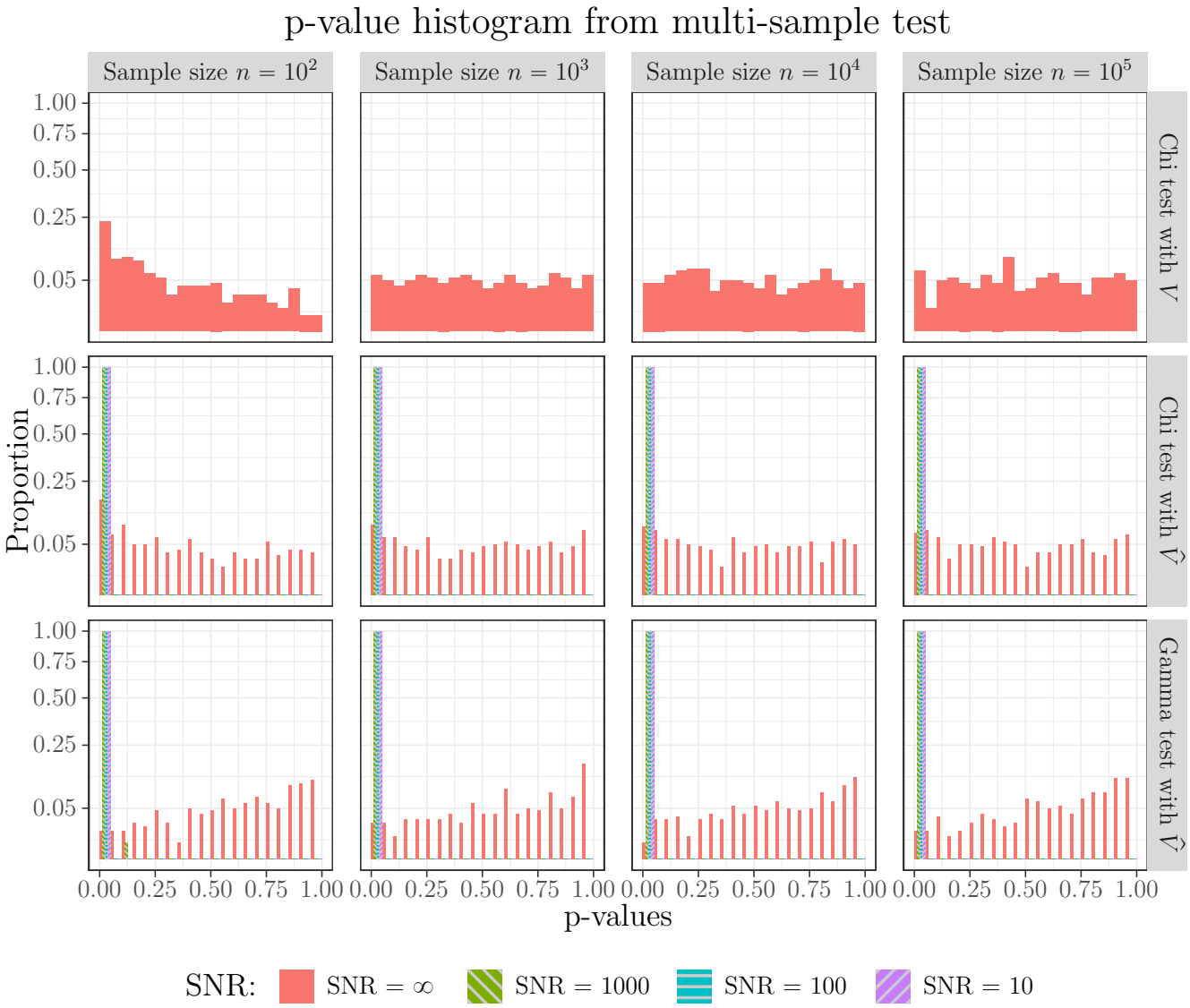

Figure 2: The histograms of p-values for the multi-sample test. We randomly sampled 200 independent replicates for each data-generating setting of (SNR, n). The first row of plots are the test results of Proposition 4.2 on $\mathcal{A}_p(0)$ supposing the exact common eigenvectors matrix V is known, hence p-values are only available with $\text{SNR} = \infty$. The simulations of the next two rows use the estimated $\widehat{V}$ from '*(W)JDTE*' as input for Proposition 4.2 and Corollary 4.1, respectively.

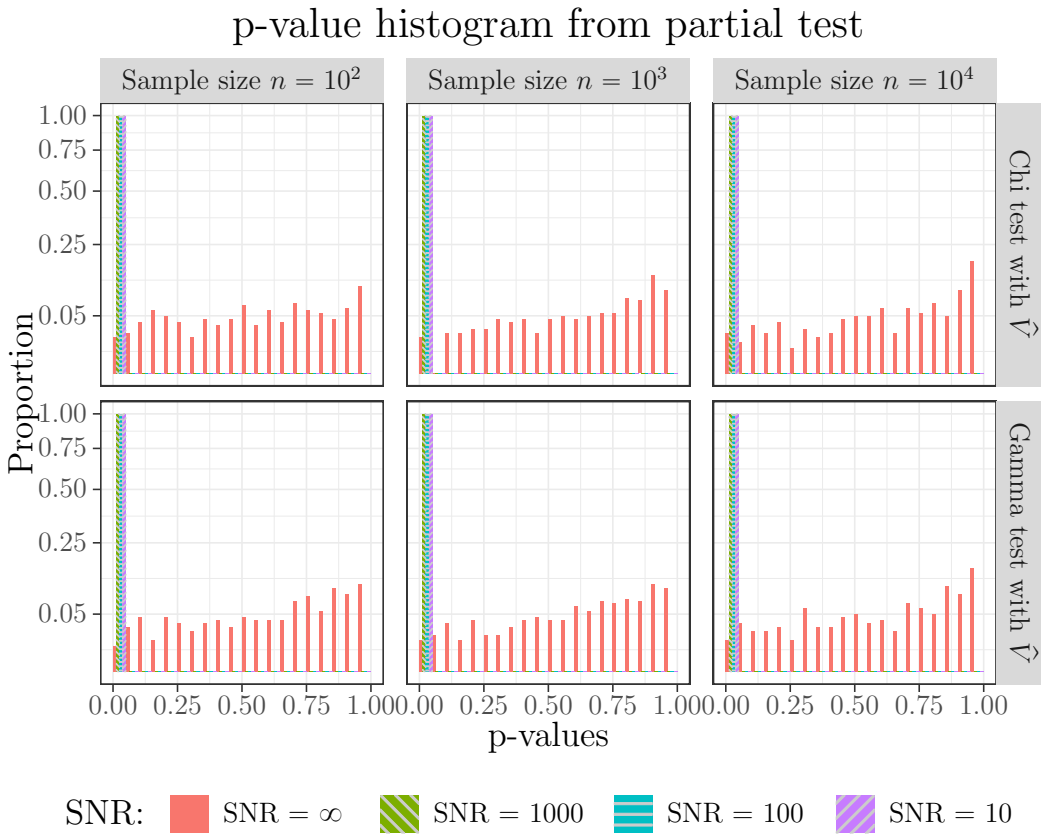

Figure 3: The histograms of p-values for the partial test. 200 independent replicates are sampled for each data-generating setting of (SNR, n). Testing with estimated $\widehat{V}$, the first and second rows are results from Proposition 5.2 and Corollary 5.1, respectively.

similar testing power. In addition, even if SNR increases, i.e. the perturbation on common eigenvectors V becomes subtle, the type II error can maintain at almost zero even with small sample size n.

As pointed out in Section 5, we assume that the number of partial common eigenvectors in known. Since this assumption is not feasible in practice, we propose a sequential testing procedure. We refer to Section A.2 in the supplementary material for a detailed description of the testing procedure and a corresponding simulation study to access its performance.

7 Applications

In Section 7.1 we consider VAR models and analyze their dynamic structure in terms of our test methods. In Section 7.2, we introduce the application of our partial test on identical stationary distributions of different Markov chains.

7.1 VAR models

We consider gross domestic product (GDP), money supply (M2), and real effective exchange rate (REER) for eight of the most influential countries distributed across three different continents; see Figure 4. GDP and M2 data are available through CEIC [1992], and REER data through OECD [2015]. The Bayesian information criterion (BIC) favors VAR models of order one for all eight countries. Therefore, for each country, we fit

$$\boldsymbol{y}_t^c = \boldsymbol{\mu}^c + \Phi^c \boldsymbol{y}_{t-1}^c + \boldsymbol{e}_t^c, \;\; t = 1, \ldots, n+1,$$

to the three different economic indices (GDP, M2, REER), where the superscript c distinguishes the eight countries. We may compare the growth tendency among subjects if the eigendecomposition guarantees common components. If our test fails to reject the null hypothesis for the matrices Φ^c, it provides us confidence to conduct the transformation $\boldsymbol{z}_t^c = V\boldsymbol{y}_t^c$, $\boldsymbol{\mu}_*^c = V\boldsymbol{\mu}^c$ with $V\Phi^c V^{-1} = D^c$ diagonal. The lagged cross-dependence cancels out for each coordinate of $\boldsymbol{z}_t^c$, i.e. $\mathbb{E}[\boldsymbol{z}_t^c] = \boldsymbol{\mu}_*^c + D^c\mathbb{E}[\boldsymbol{z}_{t-1}^c]$. Hence, with $\boldsymbol{z}_t^c$ as the new set of variables, comparison between countries can then be done on each variable separately.

The quarterly data of seasonally adjusted time series span from the first quarter of 1992 to the first quarter of 2020, with length $n+1 = 113$, and are pre-processed by taking the log-difference and standardization. The least square estimators of the coefficient matrices are then obtained, and they follow asymptotic normal distributions with rate $n^{-1/2}$; see

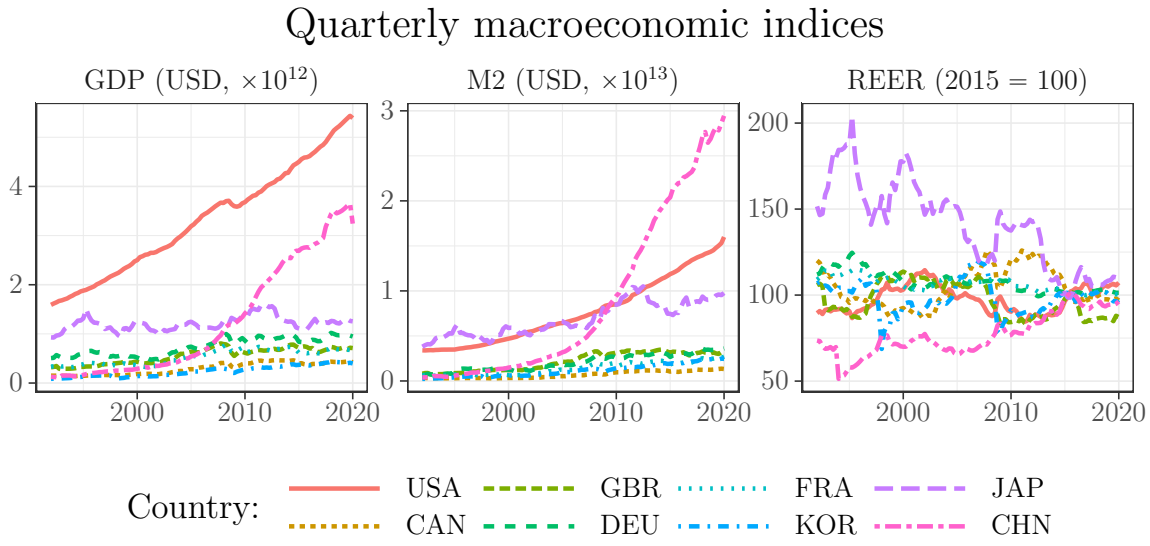

Figure 4: The time plots of the macro-economic indices of eight countries from the first quarter of 1992 to the first quarter of 2020. The data set is collected from CEIC [1992] and OECD [2015].

Chapter 3 in Lütkepohl [2005] for more details on estimating VAR models. We conduct our multi-sample tests Corollary 4.1 and Proposition 4.2 on those coefficient matrices.

Implementing Corollary 4.1, we get a p-value 0.991 which indicates that we fail to reject H_0 concerning all eight countries. Meanwhile Proposition 4.2 yields a p-value of 0.045 at the margin of rejection level. These results are not surprising based on our simulation study in Section 6.3. Figure 2 indicates that the chi-squared test (Proposition 4.2) tends to reject the hypothesis more often than suggested by the nominal test size and the gamma test (Corollary 4.1) tends to reject the hypothesis less often.

In addition, if we look at the p-value table (Figure 5) for simultaneous commutator-based tests according to Proposition 3.2, we see that the United States and China both share quite evident similarities with all other countries, and there are reasonable similarity structures within each continental group, except between Korea and Japan. Splitting the eight countries into three continental groups and repeating our test based on Corollary 4.1 group-wisely, we successfully get the conclusion that H_0 holds within the continental groups.

The corresponding p-values are, 0.530 for North America (United States and Canada), 0.978 for Europe (France, Germany, and United Kingdom), and 0.900 for Asia (China, Japan, and Korea). In addition, even the unstable Proposition 4.2 gives p-values 0.076 and 0.657 for North America and Europe, respectively.

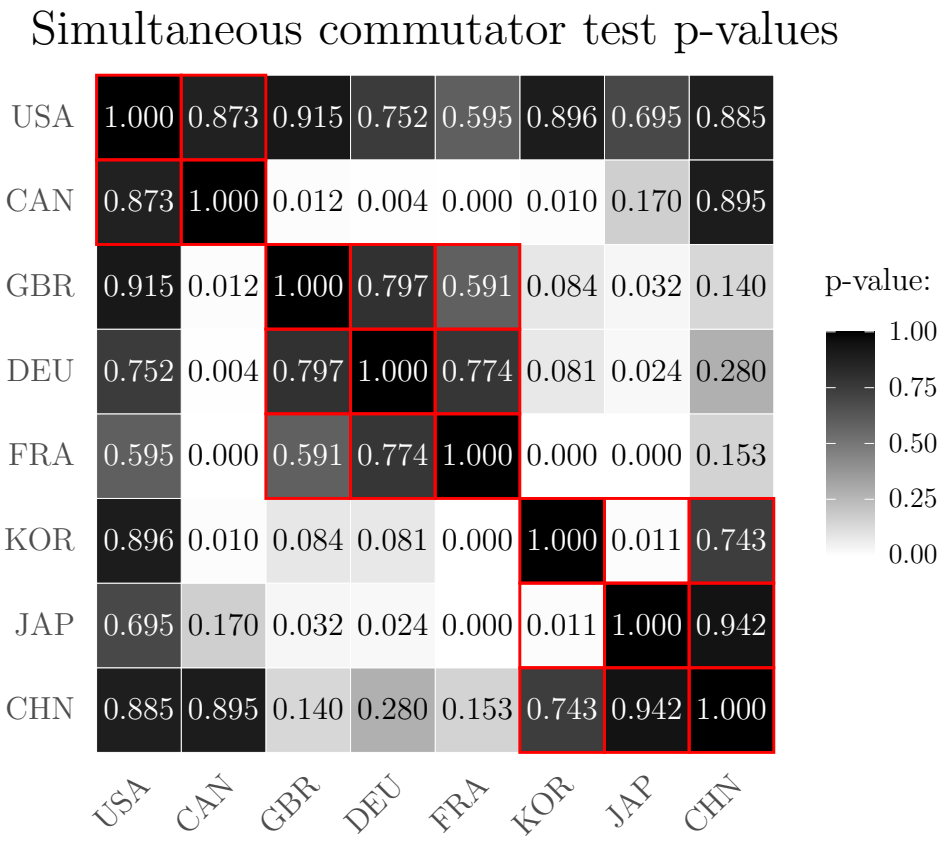

Figure 5: The heatmap of the p-values of the pairwise commutator-based test.

Since the test statistic in Corollary 4.1 gives a relatively large p-value and taking into consideration the test results of our pairwise test as well as the test performances in the simulation study, we may conclude that one fails to reject H_0. Our test results provide evidence that it is reasonable to assume that the coefficient matrices share the same eigen-

vectors. The approximated common left eigenvectors V and the new shared variables are

$$\boldsymbol{z}_t = V \begin{pmatrix} \text{GDP}_t \\ \text{M2}_t \\ \text{REER}_t \end{pmatrix} = \begin{pmatrix} 0.66 & 0.41 & -0.10 \\ 0.62 & 0.04 & -0.81 \\ 0.72 & -1.03 & -0.10 \end{pmatrix} \begin{pmatrix} \text{GDP}_t \\ \text{M2}_t \\ \text{REER}_t \end{pmatrix}. \tag{23}$$

In addition, we conduct the partial test for one common eigenvector ($k = 1$) on all eight coefficient matrices, and get that the p-values equal 0.737 from the chi-squared test (19) and 0.964 from the gamma approximation (21). In addition, when considering $k = 2$, (21) gives the p-value 0.948. Recall that the transformed variables have the following notation $\boldsymbol{z}_t^c = (\boldsymbol{z}_t[1], \boldsymbol{z}_t[2], \boldsymbol{z}_t^c[3])$, where the first two variables (without superscript c) are shared across all eight countries. The two common variables $\boldsymbol{z}_t[1]$ and $\boldsymbol{z}_t[2]$ that only depend on itself in expectation are

$$\boldsymbol{z}_t[1] = 0.67 * \text{GDP}_t + 0.34 * \text{M2}_t - 0.68 * \text{REER}_t,$$

$$\boldsymbol{z}_t[2] = 0.58 * \text{GDP}_t - 0.93 * \text{M2}_t - 0.03 * \text{REER}_t.$$

Note that $\boldsymbol{z}_t[1]$ and $\boldsymbol{z}_t[2]$ correspond to the first and third row of V in (23) respectively, with a bit rescaling and fluctuations.

7.2 Stationary distribution of Markov chains

Consider p recurrent Markov chains $\{X_i = \{X_{i,t}\}_{t=1}^{n+1}\}_{i=1}^{p}$ with time length $n+1$ and labels $X_{i,t}$ from a finite discrete state space $\{1, \ldots, d\}$. Within each chain the transition probability matrix P_i can be estimated consistently to follow asymptotic normality. For example, P_i

can be estimated by

$$\widehat{P}_i = \left(\widehat{P}_{i,rs}\right)_{d\times d},\ \widehat{P}_{i,rs} = \frac{\sum_{t=1}^{n}\mathbb{I}(X_{i,t}=r)\mathbb{I}(X_{i,t+1}=s)}{\sum_{t=1}^{n}\mathbb{I}(X_{i,t}=r)}; \tag{24}$$

see (3) in Barsotti et al. [2014]. The estimates $\widehat{P}_i$ in (24) can be deduced to follow asymptotic normality in the sense of Assumption 1 with limiting covariance Σ_i given by

$$\lim_{n\to\infty} n\,\mathrm{Cov}(\widehat{P}_{i,rs},\widehat{P}_{i,uv}) = \begin{cases} P_{i,rs}(1-P_{i,rs})/\boldsymbol{\pi}_{i,r}, & r=u, s=v, \\ -P_{i,rs}P_{i,uv}/\boldsymbol{\pi}_{i,r}, & r=u, s\neq v, \\ 0, & \text{otherwise}, \end{cases} \tag{25}$$

where $\boldsymbol{\pi}_i = (\boldsymbol{\pi}_{i,1},\ldots,\boldsymbol{\pi}_{i,d}) \in \mathbb{R}^d$ is the stationary distribution of chain X_i; see Lemma 3.1 in Barsotti et al. [2014]. Note that $\boldsymbol{\pi}_{i,r}$ is strictly positive for any i and r since all chains are recurrent. In practice, $\boldsymbol{\pi}_i$ can be estimated either from $\widehat{P}_i$ or directly from chain X_i.

Applying the Perron-Frobenius theorem [Horn and Johnson, 2012, Theorem 8.4.4], it is possible to conduct our partial test to check identical stationary distributions if we could find a common non-negative eigenvector. Using the fact that each P_i (and $\widehat{P}_i$) has stationary distribution as an eigenvector corresponding to eigenvalue 1, we can optimize the common non-negative eigenvector associated with fixed eigenvalue 1, which must be proportional to the common distribution vector if it exists. We aim to find the non-negative vector $\widehat{\boldsymbol{\pi}}$ which optimizes the problem

$$\text{minimize } f(\boldsymbol{x}) = \sum_{i=1}^{p} \boldsymbol{x}'(\widehat{P}_i - I_d)(\widehat{P}_i' - I_d)\boldsymbol{x}, \quad \text{subject to } \sum_{i=1}^{d}\boldsymbol{x}_i = 1,\ \text{and } \boldsymbol{x}_i \geqslant 0,\ \forall i = 1,\ldots,d. \tag{26}$$

It can be solved via quadratic programming with explicitly given constraints.

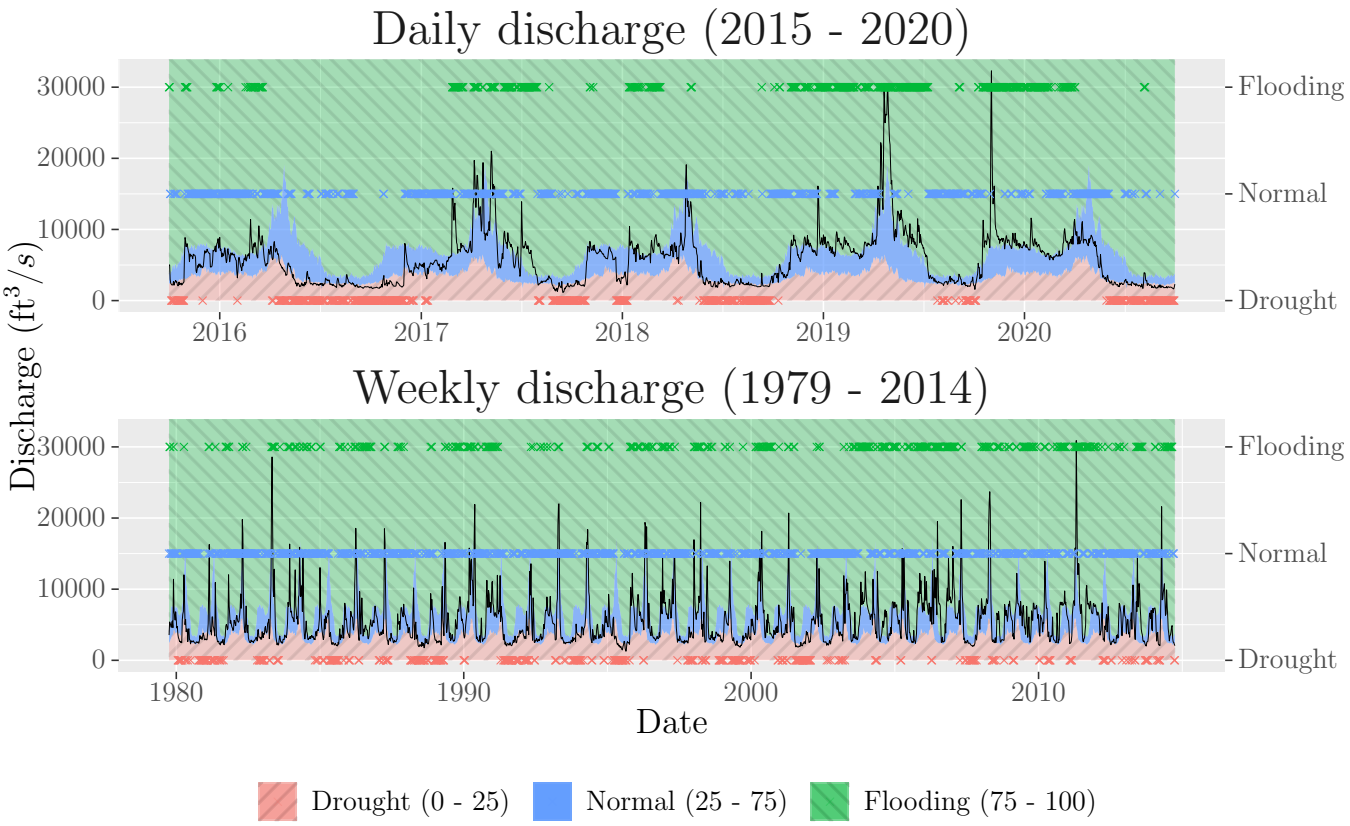

Figure 6: The discharge dataset for Hudson river with different time resolutions and disjoint time coverage. The unit is ft^3/s. The discharges are classified as 'drought' if in the lower region (below 25 percentile), as 'flooding' if in the upper region (above 75 percentile), and as 'normal' if in the middle. In the plot, the (colored) crosses represent the classified states, with axis on the right.

We consider the streamflow discharge data of Hudson river collected at Fort Edward NY; see Figure 6. The dataset is available at U.S. Geological Survey [2016]. We use the historical weekly data (from Oct. 1st 1979 to Sept. 30th 2014) and the more recent daily data (from Oct. 1st 2015 to Sept. 30th 2020). Both series have the same length $n = 1827$. We classify the discharge records to 3 levels according to the percentile statistics based on data from Oct. 1st 1977 to Sept. 30th 2019. An observation is referred to as 'drought' if it is below the 25th percentile, as 'flooding' if above the 75th percentile, and as 'normal' otherwise. We assume that this 3-state sequence satisfies the Markov property,

and estimate the transition probability matrices (see Figure 7) for our partial test. Note that the two time series are disjoint, hence it is reasonable to view these estimators as independent.

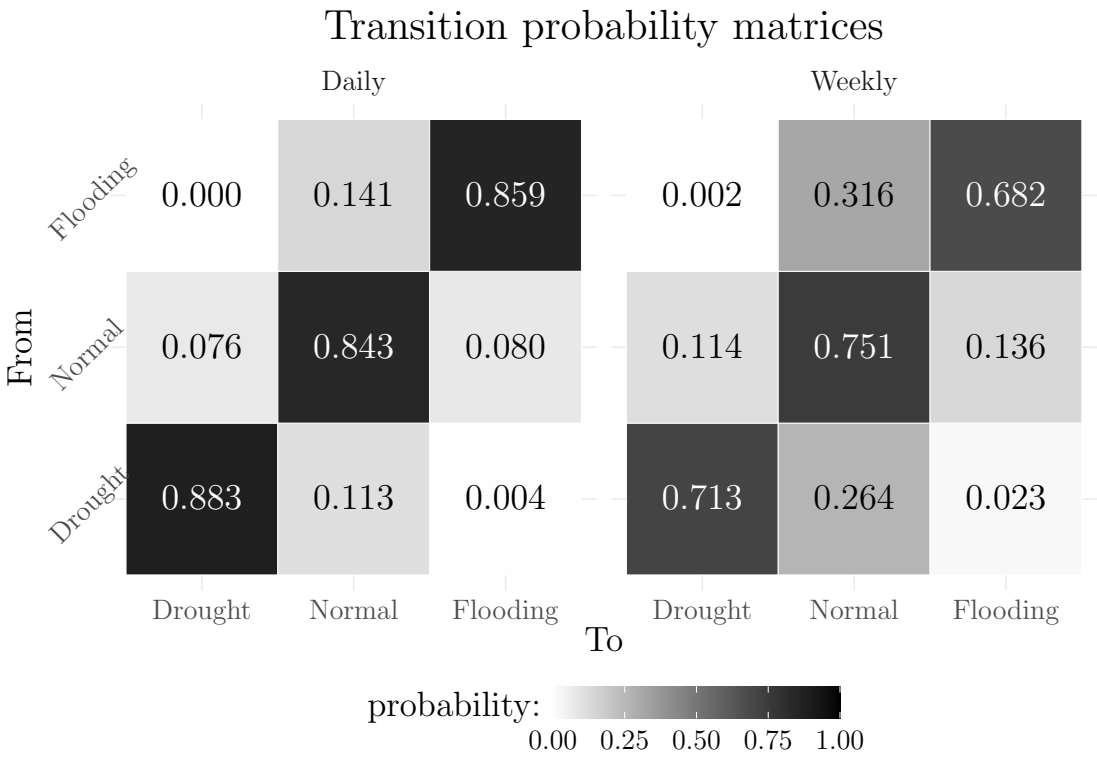

Figure 7: The 2 transition matrices are estimated according to (24) and they are statistically different according to asymptotic Wald test. Numerically speaking, at least the diagonal entries of the 2 matrices are obviously unequal.

We implement the partial test according to both Proposition 5.2 and Corollary 5.1, with parameter $k = 1$ and $V = \widehat{\boldsymbol{\pi}}_i$ given by optimization (26). The two tests give the p-values of 0.106 and 0.102 respectively, which indicate that one fails to reject H_0^* in (2) and may conclude the two series share the same stationary distribution. The estimated common stationary distribution is $\mathbb{P}(\text{Drought}) = 0.230$, $\mathbb{P}(\text{Normal}) = 0.519$, $\mathbb{P}(\text{Flooding}) = 0.251$. The results indicate that for time series with different time resolutions, even though the transition probability matrices can be different, the stationary distribution might still be verified to be statistically equivalent.

8 Conclusions

In this work, we focused on general asymmetric matrix-valued population quantities and developed effective tests on simultaneous diagonalizability under both two-sample and multi-sample settings. We also generalized our designs to test on partially shared eigenvectors, with introduction of a supplemental optimization algorithm that retrieves those common eigenvectors. Finally, we applied our test to real data examples and successfully revealed interesting structural properties of the underlying models.

In this work, we considered the classical "fixed d, large n" regime. However, many contemporary data go beyond the low dimensional setting and require the dimension d to be of the same order as, or possibly even larger than, the sample size n. While the high-dimensional setting goes beyond the scope of this work, we added a discussion and simulation study in Section A.3 of the supplementary material to emphasize that the methodology in this paper is not sufficient to do testing on high-dimensional data.

Other possible future directions include adjacency matrices of weighted directed networks which are possibly asymmetric square matrices. Our partial test could be of special interest to test on their leading eigenvectors. Combined with the high-dimensional setting, one could even study expanding networks.

In our multi-sample (Section 4) and partial tests (Section 5), the estimated $\widehat{V}$ is treated as given and deterministic in the test statistics. It might be possible to pursue more rigorous results by deriving the stochastic properties of the optimizer $\widehat{V}$.

Supplementary material The supplementary material includes all appendices. In particular, it includes an alternative approach for the two-sample test, all proofs and some additional discussions.

Acknowledgments The authors thank the anonymous Reviewer and the Associate editor

for their comments which helped improving the paper. The authors gratefully acknowledge financial support from a Xerox PARC Faculty Research Award, National Science Foundation Awards 1455172, 1934985, 1940124, and 1940276, USAID, and Cornell University Atkinson Center for a Sustainable Future. This research was conducted with support from the Cornell University Center for Advanced Computing, which receives funding from Cornell University, the National Science Foundation, and members of its Partner Program.

Conflict of interest The authors report there are no competing interests to declare.

References

André, R., Luciani, X., and Moreau, E. Joint eigenvalue decomposition algorithms based on first-order Taylor expansion. *IEEE Transactions on Signal Processing*, 68:1716–1727, 2020. doi: 10.1109/TSP.2020.2976580.

Barsotti, F., De Castro, Y., Espinasse, T., and Rochet, P. Estimating the transition matrix of a Markov chain observed at random times. *Statistics & Probability Letters*, 94:98–105, 2014.

Bodenham, D. A. and Adams, N. M. A comparison of efficient approximations for a weighted sum of chi-squared random variables. *Statistics and Computing*, 26(4):917–928, 2016.

Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics*, 25(2):290–302, 1954.

Cai, T. T., Ma, Z., and Wu, Y. Sparse PCA: Optimal rates and adaptive estimation. *The Annals of Statistics*, 41(6):3074–3110, 2013.

CEIC. Global database, 1992. URL https://www.ceicdata.com/en/products/global-economic-database[Accessed: Nov. 2020].

Colombo, N. and Vlassis, N. Approximate joint matrix triangularization. *arXiv preprint arXiv:1607.00514*, 2016a.

Colombo, N. and Vlassis, N. Tensor decomposition via joint matrix Schur decomposition. In *International Conference on Machine Learning*, pages 2820–2828, 2016b.

Donald, S. G., Fortuna, N., and Pipiras, V. On rank estimation in symmetric matrices: the case of indefinite matrix estimators. *Econometric Theory*, 23(6):1217–1232, 2007.

Engle, R. and Granger, C. Co-integration and error correction: Representation, estimation, and testing. *Econometrica*, 55(2):251–76, 1987.

Fan, J., Liao, Y., and Mincheva, M. Large covariance estimation by thresholding principal orthogonal complements. *Journal of the Royal Statistical Society. Series B, Statistical methodology*, 75(4), 2013.

Fan, J., Rigollet, P., and Wang, W. Estimation of functionals of sparse covariance matrices. *The Annals of Statistics*, 43(6):2706, 2015.

Fan, J., Wang, W., and Zhong, Y. An ℓ_∞ eigenvector perturbation bound and its application to robust covariance estimation. *Journal of Machine Learning Research*, 18(207):1–42, 2018.

Flury, B. N. Asymptotic theory for common principal component analysis. *The Annals of Statistics*, 14(2):418–430, 1986.

Flury, B. N. Common principal components in k groups. *Publications of the American Statistical Association*, 79(388):892–898, 1984.

Flury, B. N. and Gautschi, W. An algorithm for simultaneous orthogonal transformation of several positive definite symmetric matrices to nearly diagonal form. *Siam Journal on Scientific & Statistical Computing*, 7(1):169–184, 1986.

Fu, T. and Gao, X. Simultaneous diagonalization with similarity transformation for non-defective matrices. In *2006 IEEE International Conference on Acoustics Speech and Signal Processing Proceedings*, volume 4, pages IV–IV. IEEE, 2006.

Fujioka, T. An approximate test for common principal component subspaces in two groups. *The Annals of the Institute of Statistical Mathematics*, 45(1):147–158, 1993.

Ghazi, A. E., Hajji, S. E., Giraud, L., and Gratton, S. Newton's method for the common eigenvector problem. *Journal of Computational & Applied Mathematics*, 219(2):398–407, 2008.

Ghazi, A. E., Hajji, S. E., Giraud, L., and Gratton, S. A short note on backward errors for the common eigenvector problem, 2014.

Hadi, A. and Wells, M. A note on generalized Wald's method. *Metrika*, 37(1):309–315, 1990.

Horn, R. A. and Johnson, C. R. *Matrix analysis*. Cambridge University Press, 2012.

Iferroudjene, R., Abed Meraim, K., and Belouchrani, A. A new Jacobi-like method for joint diagonalization of arbitrary non-defective matrices. *Applied Mathematics and Computation*, 211(2):363–373, 2009.

Johansen, S. Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models. *Econometrica*, 59(6):1551–1580, 1991.

Kleibergen, F. and Paap, R. Generalized reduced rank tests using the singular value decomposition. *Journal of Econometrics*, 133(1):97–126, 2006.

Koltchinskii, V. and Lounici, K. New asymptotic results in principal component analysis. *Sankhya A*, 79(2):254–297, 2017.

Lei, J. and Rinaldo, A. Consistency of spectral clustering in stochastic block models. *The Annals of Statistics*, 43(1):215–237, 2015.

Luciani, X. and Albera, L. Joint eigenvalue decomposition using polar matrix factorization. In *Latent Variable Analysis and Signal Separation*, pages 555–562. Springer Berlin Heidelberg, 2010. ISBN 978-3-642-15995-4.

Lütkepohl, H. *New introduction to multiple time series analysis*. Springer, Berlin, 2005.

Lütkepohl, H. and Burda, M. M. Modified Wald tests under nonregular conditions. *Journal of Econometrics*, 78(2):315–332, 1997.

Magnus, J. R. and Neudecker, H. *Matrix differential calculus with applications in statistics and econometrics*. John Wiley & Sons, 2019.

Nadler, B. Finite sample approximation results for principal component analysis: A matrix perturbation approach. *The Annals of Statistics*, 36(6):2791–2817, 2008.

Naumov, A., Spokoiny, V., and Ulyanov, V. Bootstrap confidence sets for spectral projectors of sample covariance. *Probability Theory and Related Fields*, 174(3-4):1091–1132, 2019.

OECD. Main economic indicators - complete database, 2015. URL https://www.oecd-ilibrary.org/content/data/data-00052-en[Accessed: Nov. 2020].

Paul, S. and Chen, Y. Spectral and matrix factorization methods for consistent community detection in multi-layer networks. *The Annals of Statistics*, 48(1):230–250, 2020.

Ratsimalahelo, Z. Rank test based on matrix perturbation theory. Technical report, EERI Research Paper Series, 2001.

Restrepo, J. G., Ott, E., and Hunt, B. R. Approximating the largest eigenvalue of network adjacency matrices. *Physical Review E*, 76(5):056119, 2007.

Robin, J.-M. and Smith, R. J. Tests of rank. *Econometric Theory*, 16(2):151–175, 2000.

Rohe, K., Chatterjee, S., and Yu, B. Spectral clustering and the high-dimensional stochastic blockmodel. *The Annals of Statistics*, 39(4):1878–1915, 2011.

Schott, J. R. Partial common principal component subspaces. *Biometrika*, 86(4):899–908, 1999.

Schwartzman, A., Mascarenhas, W. F., and Taylor, J. E. Inference for eigenvalues and eigenvectors of Gaussian symmetric matrices. *The Annals of Statistics*, 36(6):2886–2919, 2008.

Silin, I. and Fan, J. Hypothesis testing for eigenspaces of covariance matrix. *arXiv preprint arXiv:2002.09810*, 2020.

Silin, I. and Spokoiny, V. Bayesian inference for spectral projectors of the covariance matrix. *Electronic Journal of Statistics*, 12(1):1948–1987, 2018.

Tang, M. and Priebe, C. E. Limit theorems for eigenvectors of the normalized Laplacian for random graphs. *The Annals of Statistics*, 46(5):2360–2415, 2018.

Tyler, D. E. Asymptotic inference for eigenvectors. *The Annals of Statistics*, pages 725–736, 1981.

U.S. Geological Survey. USGS water data for the nation. *National Water Information System Data*, 2016. doi: 10.5066/F7P55KJN. URL https://nwis.waterdata.usgs.gov/nwis/[Accessed: Nov. 2020].

Vogelsang, T. J. Unit roots, cointegration, and structural change. *Journal of the American Statistical Association*, 96(453):346–346, 2001.

Von Luxburg, U. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

Zhang, R., Robinson, P., and Yao, Q. Identifying cointegration by eigenanalysis. *Journal of the American Statistical Association*, 114(526):916–927, 2019.