Supplement to "Testing Simultaneous Diagonalizability"

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

November 8, 2022

We present here supplementary results for the article "Testing Simultaneous Diagonalizability". Appendix A provides some additional numerical results. In Appendix B, we propose and analyze an alternative to the commutator based test for the two-sample test problem. Appendix C discusses an extension for symmetric matrices. Finally, Appendices D and E contain all proofs. We adopt the notation of the article and refer to its labels.

A Complementary simulation results

We provide here some empirical results complementary to the numerical analysis presented in the main paper. Section A.1 gives empirical sizes and powers for the proposed test, Section A.2 studies sequential application of our partial tests and Section A.3 discusses application to possibly high-dimensional data.

A.1 Empirical Type I and II errors

In addition to the p-values in the main paper, we provide here tables with Type I and II errors for our tests to assess their performances. Tables A.1, A.2 and A.3 show respectively the errors for the two-sample, multi-sample and partial tests.

A.2 Sequential application of partial tests

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. The hypothesis testing problem (2) can be stated for $k \in \{1, \ldots, d\}$. The sequential testing starts with $k = d$, then $k = d - 1$ and so on, till the null hypothesis is not rejected. The performance of this procedure is accessed through a simulation study in Table A.4.

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

Test Type	Statistics Type	Sample Size	Type I Error	Type II Error	
				SNR=1000	SNR=10
Commutator-based test	Chi-test	50	0.218	0.216	0.000
		250	0.056	0.000	0.000
		1000	0.054	0.000	0.000
LLR test (Appendix B)	Oracle Chi-Test with (B.5)	50	0.182	0.000	0.000
		250	0.140	0.000	0.000
		1000	0.060	0.000	0.000
	Plugin Chi-test with (B.5)	50	1.000	0.000	0.000
		250	1.000	0.000	0.000
		1000	1.000	0.000	0.000
	Oracle Chi-Test with (B.6)	50	0.182	0.058	0.000
		250	0.140	0.000	0.000
		1000	0.060	0.000	0.000
	Plugin Chi-test with (B.6)	50	0.760	0.000	0.000
		250	0.842	0.000	0.000
		1000	0.774	0.000	0.000

Table A.1: Two-sample test results on simulated $\mathcal{M}_2(\rho, 5; 5)$ for $\rho^2 = \frac{1}{\mathrm{SNR}} \in \{0, \frac{1}{1000}, \frac{1}{10}\}$.

Statistics Type	Sample Size	Type I Error	Type II Error		
			SNR = 1000	SNR = 100	SNR = 10
Oracle Chi-test (Proposition 4.1)	100	0.230	NA	NA	NA
	1000	0.060	NA	NA	NA
	10000	0.045	NA	NA	NA
	100000	0.070	NA	NA	NA
Plugin Chi-test (Proposition 4.2)	100	0.175	0.000	0.000	0.000
	1000	0.095	0.000	0.000	0.000
	10000	0.090	0.000	0.000	0.000
	100000	0.075	0.000	0.000	0.000
Plugin Gamma-test (Corollary 4.1)	100	0.015	0.005	0.000	0.000
	1000	0.025	0.000	0.000	0.000
	10000	0.005	0.000	0.000	0.000
	100000	0.015	0.000	0.000	0.000

Table A.2: Multi-sample test results on simulated $\mathcal{M}_8(\rho, 4; 4)$ for $\rho^2 = \frac{1}{\mathrm{SNR}} \in \{0, \frac{1}{1000}, \frac{1}{100}\, \frac{1}{10}\}$.

Statistics Type	Sample Size	Type I Error	Type II Error		
			SNR = 1000	SNR = 100	SNR = 10
Chi-test (Proposition 5.2)	100	0.020	0.000	0.000	0.000
	1000	0.020	0.000	0.000	0.000
	10000	0.025	0.000	0.000	0.000
Gamma-test (Corollary 5.1)	100	0.010	0.000	0.000	0.000
	1000	0.015	0.000	0.000	0.000
	10000	0.015	0.000	0.000	0.000

Table A.3: Partial test results on simulated $\mathcal{M}_8(\rho, 2; 4)$ for $\rho = \frac{1}{\text{SNR}} \in \{0, \frac{1}{1000}, \frac{1}{100}\, \frac{1}{10}\}$.

Statistics Type	Sample Size	Rejection Rate		
		$k = 2$	$k = 3$	$k = 4$
Chi-test (Proposition 5.2)	100	0.020	1.000	1.000
	1000	0.020	1.000	1.000
	10000	0.025	1.000	1.000
Gamma-test (Corollary 5.1)	100	0.010	1.000	1.000
	1000	0.015	1.000	1.000
	10000	0.015	1.000	1.000

Table A.4: Partial test results on simulated $\mathcal{M}_8(0, 2; 4)$ and potentially mis-specified $k \in \{2, 3, 4\}$.

A.3 High-dimensional data

In this work, we consider 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 would like to point out why our methodology is not sufficient to do testing on high-dimensional data.

Table A.5 presents the empirical rejection rates and average degrees of freedom for the two-sample test in Proposition 3.1, considering different sample sizes $n = 50, 100, 500$ and letting d grow. We present results assuming that the limiting covariance matrix in (7) is estimated and known. The existence of a consistent estimator is stated in Assumption 2 and makes our procedure feasible in practice.

The results in Table A.5 show that the classical theory suffers a α test size much higher than the nominal test level once we consider high-dimensional data and estimate the limiting covariance matrix. Intuitively, the results are expected to break down once the sample size does not satisfy $n > r_1(d^2 + d^2)$. This can be easily seen by counting the degrees of freedom required to specify a rank-r_1 matrix of size $d^2 \times d^2$. Roughly speaking, we need r_1 numbers to specify the matrix's singular values, and $r_1 d^2$ and $r_1 d^2$ numbers to specify its left and right singular vectors.

The α test size much higher than the nominal test level is also due to Assumption 2 no longer being satisfied in a high-dimensional regime. In particular, the difference between estimator and true matrix is incorrectly normalized once the dimension grows with the sample size. It is expected to require results from random matrix theory to get convergence under suitable assumptions on the ratio between d and n.

d	Sample Size=50				Sample Size=100				Sample Size = 500			
	Empirical Cov		True Cov		Empirical Cov		True Cov		Empirical Cov		True Cov	
	Size	Avg DF	Size	Avg DF	Size	Avg DF	Size	Avg DF	Size	Avg DF	Size	Avg DF
2	0.036	2.05	0.022	2.04	0.054	2.00	0.058	2.00	0.044	2.00	0.044	2.00
3	0.066	6.10	0.034	6.13	0.074	6.02	0.054	6.02	0.062	6.00	0.060	6.00
4	0.088	12.33	0.014	12.38	0.092	12.06	0.036	12.06	0.066	12.00	0.054	12.00
5	0.186	20.15	0.018	20.42	0.116	20.05	0.050	20.08	0.062	20.00	0.056	20.00
6	0.358	29.75	0.020	30.65	0.176	29.97	0.020	30.09	0.058	30.00	0.046	30.00
7	0.698	40.85	0.018	42.79	0.310	40.46	0.026	40.97	0.092	40.00	0.050	40.00
8	0.960	55.82	0.024	57.78	0.548	56.04	0.034	56.30	0.092	56.00	0.044	56.00
9	1.000	69.11	0.014	74.19	0.778	70.54	0.044	71.97	0.098	70.44	0.052	70.57
10	1.000	86.27	0.020	92.94	0.962	89.62	0.034	90.74	0.162	90.00	0.054	90.00
11	1.000	96.93	0.024	114.55	0.996	108.25	0.030	111.15	0.238	108.79	0.050	109.15
12	1.000	98.00	0.030	136.88	1.000	131.96	0.042	133.48	0.340	132.00	0.056	132.00
13	1.000	98.00	0.022	161.63	1.000	153.87	0.030	157.93	0.468	156.00	0.044	156.00
14	1.000	98.00	0.028	188.91	1.000	174.85	0.026	184.56	0.622	182.00	0.044	182.00
15	1.000	98.00	0.040	217.63	1.000	193.89	0.042	212.96	0.774	209.88	0.052	209.99
16	0.998	98.00	0.042	248.27	1.000	198.00	0.034	243.25	0.874	239.89	0.042	240.00
17	0.998	98.00	0.050	281.03	1.000	198.00	0.042	275.62	0.974	270.00	0.046	270.14
18	0.978	98.00	0.050	315.85	1.000	198.00	0.036	310.37	0.986	303.18	0.056	304.74
19	0.994	98.00	0.046	352.90	1.000	198.00	0.050	347.06	0.998	340.60	0.044	341.50
20	0.994	98.00	0.056	391.32	1.000	198.00	0.046	384.37	1.000	373.18	0.044	375.67
21	0.460	98.00	0.054	433.27	1.000	198.00	0.040	426.84	1.000	418.44	0.044	419.91
22	0.982	98.00	0.044	475.39	1.000	198.00	0.040	468.52	1.000	455.78	0.054	459.21
23	0.986	98.00	0.044	520.58	1.000	198.00	0.054	513.53	1.000	499.67	0.042	503.17
24	0.284	98.00	0.032	567.74	1.000	198.00	0.030	560.31	1.000	544.63	0.040	551.13
25	0.702	98.00	0.056	617.31	1.000	198.00	0.046	609.14	1.000	590.03	0.054	595.95

Table A.5: Two-sample test results on simulated $\mathcal{M}_2(0, d; d)$ for dimensions $d \in \{2, \ldots, 25\}$.

B Log-likelihood Ratio (LLR) test framework

In the main paper, we studied a commutator based two-sample test. In this section, we propose an alternative test based on a likelihood ratio test statistic.

Before we introduce the test statistic we state an assumption which is slightly stronger than Assumption 3 in the main paper.

Assumption 3*. *Each M_i, $i = 1, \ldots, p$, has d distinct non-zero real eigenvalues.*

According to the assumed asymptotic normality in Assumption 1, we introduce the log-likelihood type function for the estimators A_1 and A_2 as

$$L(M_1, M_2) := -\sum_{i=1}^{2} \mathrm{vec}(A_i - M_i)' \Sigma_i^+ \,\mathrm{vec}(A_i - M_i). \tag{B.1}$$

It is then possible to obtain the supremum of $L(M_1, M_2)$ within the parameter spaces H_0 and $H_0 \cup H_1$, respectively, as

$$\widetilde{L}_0 := \sup_{(M_1,M_2)\in H_0} L(M_1, M_2), \quad \widetilde{L}_1 := \sup_{(M_1,M_2)\in H_0\cup H_1} L(M_1, M_2). \tag{B.2}$$

In particular, we introduce a new version of M-estimators for (M_1, M_2) under the null hypothesis H_0 as

$$(\widehat{A}_1, \widehat{A}_2) = \operatorname*{argmax}_{(M_1,M_2)\in H_0} L(M_1, M_2), \tag{B.3}$$

and the design of the ratio-test statistic can be given by $\Gamma_2 \propto -(\widetilde{L}_0 - \widetilde{L}_1) = -\widetilde{L}_0$.

Indeed the estimators $\widehat{A}_1$ and $\widehat{A}_2$ in (B.3) can be explicitly computed given Assumption 3*. We introduce the following proposition and prove it in Appendix D.

Proposition B.1. *Suppose Assumption 3*. Then, the optimizer $(\widehat{A}_1, \widehat{A}_2)$ that maximizes (B.1) under H_0 is given by*

$$\mathrm{vec}(\widehat{A}_1) = P_2(P_2'\Sigma_1^+ P_2)^+ P_2'\Sigma_1^+ \,\mathrm{vec}(A_1), \quad \mathrm{vec}(\widehat{A}_2) = P_1(P_1'\Sigma_2^+ P_1)^+ P_1'\Sigma_2^+ \,\mathrm{vec}(A_2), \tag{B.4}$$

with P_i for $i = 1, 2$ generated from either of the two following setups:

- *Polynomial basis: with M_i^j the j-th power of M_i, for $i = 1, 2$,*

$$P_i = (\mathrm{vec}(M_i^1)/\|M_i^1\|_F, \ldots, \mathrm{vec}(M_i^d)/\|M_i^d\|_F); \tag{B.5}$$

- *Eigenvector basis: with $V = (v_1, \ldots, v_d) \in \mathbb{R}^{d\times d}$ the common eigenvectors of M_1 and M_2, $U = (u_1, \ldots, u_d)' = V^{-1}$,*

$$P_1 = P_2 = (\mathrm{vec}(v_1 u_1'), \ldots, \mathrm{vec}(v_d u_d')). \tag{B.6}$$

B.1 LLR test statistic

In this section we introduce the LLR test statistic and provide its asymptotic behavior. Under ideal conditions such that the true matrices M_1 and M_2 are known, we introduce the LLR test statistic

$$\begin{aligned}\Gamma_2 &:= c^2(n)\Big[\operatorname{vec}(A_1-\widehat{A}_1)'\Sigma_1^+\operatorname{vec}(A_1-\widehat{A}_1)+\operatorname{vec}(A_2-\widehat{A}_2)'\Sigma_2^+\operatorname{vec}(A_2-\widehat{A}_2)\Big]\\ &= c^2(n)\Big[\operatorname{vec}(A_1)'Q_{1,2}\operatorname{vec}(A_1)+\operatorname{vec}(A_2)'Q_{2,1}\operatorname{vec}(A_2)\Big],\end{aligned} \tag{B.7}$$

where $Q_{k,\ell}=\Sigma_k^+-\Sigma_k^+P_\ell(P_\ell'\Sigma_k^+P_\ell)^+P_\ell'\Sigma_k^+$ for $k,\ell=1,2$ and $k\neq\ell$, and present its asymptotic behavior in the following proposition.

Proposition B.2 (LLR test statistic)**.** *Suppose Assumptions 1 and 3* are satisfied. Then, under H_0 in (1), the test statistic Γ_2 in (B.7) satisfies*

$$\Gamma_2 \xrightarrow{\mathcal{D}} \chi^2(r_2), \tag{B.8}$$

where $r_2=r_{1,2}+r_{2,1}$ and $r_{k,\ell}=\operatorname{rk}(\Sigma_k)-\operatorname{rk}(P_\ell'\Sigma_k^+P_\ell)$ for $k,\ell=1,2$, and $k\neq\ell$.

Note that when Σ_1 and Σ_2 are non-singular, $r_2=2d^2-2d$. With our loose constraints on the covariance matrices, we may encounter singularity issues when computing (B.7) with Σ_1^+ and Σ_2^+. To have a tractable version of Proposition B.2 with respect to the limiting covariance matrices, we propose to use the truncated version (6). Note that the generalized inverse of $P_\ell'\Sigma_k^+P_\ell$ is a part of a projection matrix hence will not have the same discontinuity concerns.

Proposition B.3. *Suppose Assumptions 1, 2 and 3* are satisfied. Let $\varepsilon>0$ be a threshold that is not an eigenvalue of Σ_1 and Σ_2. Define the test statistic*

$$\Gamma_2^{\#}(\varepsilon) := c^2(n)\Big[\operatorname{vec}(A_1)'\widehat{Q}_{1,2}[\varepsilon]\operatorname{vec}(A_1)+\operatorname{vec}(A_2)'\widehat{Q}_{2,1}[\varepsilon]\operatorname{vec}(A_2)\Big] \tag{B.9}$$

with

$$\widehat{Q}_{k,\ell}[\varepsilon] := \widehat{\Sigma}_k^+(\varepsilon)-\widehat{\Sigma}_k^+(\varepsilon)P_\ell\big(P_\ell'\widehat{\Sigma}_k^+(\varepsilon)P_\ell\big)^+P_\ell'\widehat{\Sigma}_k^+(\varepsilon) \tag{B.10}$$

for $k,\ell=1,2$, and $k\neq\ell$. Then,

$$\Gamma_2^{\#}(\varepsilon) \xrightarrow{\mathcal{D}} \xi, \quad \text{with} \quad \xi\sim\chi^2\big(\widehat{r}_2(\varepsilon)\big), \tag{B.11}$$

where $\widehat{r}_2(\varepsilon)=\widehat{r}_{1,2}(\varepsilon)+\widehat{r}_{2,1}(\varepsilon)$ and $\widehat{r}_{k,\ell}(\varepsilon)=\operatorname{rk}(\widehat{\Sigma}_k;\varepsilon)-\operatorname{rk}\big(P_\ell'\widehat{\Sigma}_k^+(\varepsilon)P_\ell\big)$ for $k,\ell=1,2$, and $k\neq\ell$. Furthermore, note that $\widehat{r}_2^l(\varepsilon)\leqslant\widehat{r}_2(\varepsilon)\leqslant\widehat{r}_2^u(\varepsilon)$ with

$$\widehat{r}_2^l(\varepsilon) := \operatorname{rk}(\widehat{\Sigma}_1;\varepsilon)+\operatorname{rk}(\widehat{\Sigma}_2;\varepsilon)-2d, \quad \widehat{r}_2^u(\varepsilon) := \operatorname{rk}(\widehat{\Sigma}_1;\varepsilon)+\operatorname{rk}(\widehat{\Sigma}_2;\varepsilon).$$

Then, with $\xi^l\sim\chi^2(\widehat{r}_2^l(\varepsilon))$, $\xi^u\sim\chi^2(\widehat{r}_2^u(\varepsilon))$, the p-value based on (B.11) can be bounded by

$$\mathbb{P}\big(\xi^l>\Gamma_2^{\#}(\varepsilon)\mid H_0\big)\leqslant\mathbb{P}\big(\xi>\Gamma_2^{\#}(\varepsilon)\mid H_0\big)\leqslant\mathbb{P}\big(\xi^u>\Gamma_2^{\#}(\varepsilon)\mid H_0\big). \tag{B.12}$$

We include (B.12) to deal with the potentially inconsistent rank estimators of $P_\ell'\widehat{\Sigma}_k^+(\varepsilon)P_\ell$, $k,\ell=1,2$ with $k\neq\ell$, and state the following proposition to justify the effectiveness of the relaxed test based on (B.12). In particular, the proposition indicates that the hypothesis gets rejected with high probability within the hypothesis space H_1.

Proposition B.4. *Under the alternative hypothesis H_1 in (2), set*

$$\boldsymbol{m}_1 = \operatorname{vec}(M_1) - P_2\big(P_2'\Sigma_1^+(\varepsilon)P_2\big)^+ P_2'\Sigma_1^+(\varepsilon)\operatorname{vec}(M_1),$$
$$\boldsymbol{m}_2 = \operatorname{vec}(M_2) - P_1\big(P_1'\Sigma_2^+(\varepsilon)P_1\big)^+ P_1'\Sigma_2^+(\varepsilon)\operatorname{vec}(M_2),$$

with ε chosen by Proposition B.3 and $\boldsymbol{m}_i \in \mathbb{R}^{d^2}$ for $i = 1, 2$. If $\widehat{\Sigma}_i^+(\varepsilon)\boldsymbol{m}_i \neq 0$ for $i = 1, 2$, then the test statistic (B.9) satisfies

$$\lim_{n\to\infty} \Gamma_2^{\#}(\varepsilon) \to \infty \;\Rightarrow\; \lim_{n\to\infty} \mathbb{P}\big(\xi > \Gamma_2^{\#}(\varepsilon) \mid H_1\big) = 0.$$

Note that under the null hypothesis H_0, it might also be true that $\boldsymbol{m}_i \neq 0$ when $\widehat{\Sigma}_i^+(\varepsilon)$ is singular, but $\widehat{\Sigma}_i^+(\varepsilon)\boldsymbol{m}_i = 0$ for $i = 1, 2$ always hold.

B.2 Error analysis

In this section, we study the effects of replacing the matrices P_1 and P_2 in (B.4) by their estimators in our proposed test. In particular, we start from the expression (B.5) of polynomial basis. We define the estimators (B.5) for P_1 and P_2 as

$$\widehat{P}_i = (\operatorname{vec}(A_i^1)/\|A_i^1\|_F, \dots, \operatorname{vec}(A_i^d)/\|A_i^d\|_F)$$

for $i = 1, 2$. Then, under Assumption 1, $\widehat{P}_1$ and $\widehat{P}_2$ are consistent estimators for P_1 and P_2 with the same convergence rate $1/c(n)$. However, even with extra care about the covariance singularity, replacing P_i by $\widehat{P}_i$, for $i = 1, 2$, in $\Gamma_2^{\#}(\varepsilon)$ in (B.9) makes the asymptotic distribution of the test statistic (B.11) inaccurate. For this reason, one thing remains to be discussed is whether the statistical order of the error introduced from this approximation step is negligible in testing. To be more precise, for Proposition B.3, the error for the first summand in $\Gamma_2^{\#}(\varepsilon)$ is

$$\Delta_\varepsilon := c^2(n)\operatorname{vec}(A_1)'(\widehat{Q}_{1,2}[\varepsilon] - \widehat{\mathcal{Q}}_{1,2}[\varepsilon])\operatorname{vec}(A_1) \tag{B.13}$$

with $\widehat{Q}_{1,2}[\varepsilon]$ as in (B.10) and $\widehat{\mathcal{Q}}_{1,2}[\varepsilon]$ is defined by replacing the matrices P_ℓ in (B.10) by their sample counterparts $\widehat{P}_\ell$ such that

$$\widehat{\mathcal{Q}}_{k,\ell}[\varepsilon] := \widehat{\Sigma}_k^+(\varepsilon) - \widehat{\Sigma}_k^+(\varepsilon)\widehat{P}_\ell\big(\widehat{P}_\ell'\widehat{\Sigma}_k^+(\varepsilon)\widehat{P}_\ell\big)^+\widehat{P}_\ell'\widehat{\Sigma}_k^+(\varepsilon) \tag{B.14}$$

for $k, \ell = 1, 2$, $k \neq \ell$. The following proposition provides information about the asymptotic behavior of Δ_ε in (B.13). The proof can be found in Appendix D.

Proposition B.5. *Assume the choice of ε satisfies $\Sigma_1(\varepsilon) = \Sigma_1$. Then, under Assumption 1, there exists an $r_{1,2} \times r_{1,2}$ positive semi-definite matrix $\check{\Sigma} = \check{\Sigma}(M_1, M_2, \Sigma_1, \Sigma_2)$ such that the error term in (B.13) satisfies*

$$\Delta_\varepsilon \xrightarrow{\mathcal{D}} Z, \quad \textit{with} \;\; Z = \sum_{i=1}^{r_{1,2}} 2\sqrt{\lambda_i}(\nu_{i,1} - \nu_{i,2}),$$

where λ_i are the eigenvalues of $\check{\Sigma}$, and $\nu_{i,j} \overset{i.i.d.}{\sim} \chi^2(1)$ for $i \in \{1, \dots, r_{1,2}\}$, $j = 1, 2$. Furthermore, the variance of the limit is $\operatorname{Var}(Z) = 16\operatorname{tr}(\check{\Sigma})$.

According to Proposition B.5, the error term Δ_ε in (D.7) is still asymptotically unbiased. However, with a mild choice of matrix dimension d, its asymptotic variance, which represents the perturbation range, is comparable with the magnitude of the test statistic $\Gamma_2^{\#}(\varepsilon)$ in (B.11), as the matrix $\check{\Sigma} \in \mathbb{R}^{r_{1,2} \times r_{1,2}}$ is generated by well-conditioned matrices $(M_1, M_2, \Sigma_1, \Sigma_2)$. Hence, even with the relaxed test introduced in Proposition B.3, there are no guarantees that the test statistic is valid in real applications. The weighted projections $\widehat{A}_i$, however, could sometimes be useful while problem setup or interests change.

On the other hand, due to the lack of stochastic convergence results for optimization with respect to the common eigenvectors V, the consistency rate of plugging $\widehat{V}$ from '*(W)JDTE*' along with its inverse $\widehat{U} = \widehat{V}^{-1}$ into (B.6) for $\widehat{P}_i$ remains unclear. However, as long as the optimization procedure fails to improve the original $1/c(n)$ rate in Assumption 1 with positive probability, the analogous derivations will lead to a similar conclusion as Proposition B.5.

On the contrary, if one has confident prior knowledge of common eigen-structures, one can simply define the space matrices P_1 and P_2 using such prior information to make this particular approach applicable with reasonably strong test power. In addition to the direct access to the common eigenvectors for constructing (B.6), knowledge of common eigen-structures could also be that, when defining (B.5), there is a reference square matrix which shares eigenvectors with the matrices to be tested.

B.3 Summary of two-sample test

The test methods developed in Section 3 and Appendix B could be applied in different settings. For example, if the estimators are available with reasonable asymptotic normality, only the commutator-based test design would guarantee acceptable effectiveness; and if exact eigen-information is given with certainty, the LLR test could be a good choice. However, cases with such strong restrictions and adequate information could be rather rare in real applications. In the simulations and applications later, the commutator-based test is conducted.

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, 250 for instance, the test performs well with acceptable type I error and excellent type II error.

For the LLR test, we conduct (i) Proposition B.2 given either the exact $\mathcal{M}_2(\rho, d; d)$ for (B.5) or the exact eigenvectors V for (B.6), and (ii) Proposition B.3 with estimators $\mathcal{A}_2(\rho)$ plugged-in. By comparing the first and the third rows of Figure B.1 with Figure 1, Proposition B.2 has better performance in terms of type II error, especially with small sample sizes. With the plugin version Proposition B.3 (the second and the fourth rows), however, one frequently fails to distinguish the null hypothesis even under the ideal case when SNR $= \infty$ or $\rho = 0$. It is compatible with our error analysis given in Section B.2. To summarize, though with the well-behaved p-values of Proposition B.2 under H_0, such idea from Appendix B may rarely be applicable unless the test is against some deterministic reference eigen-structure.

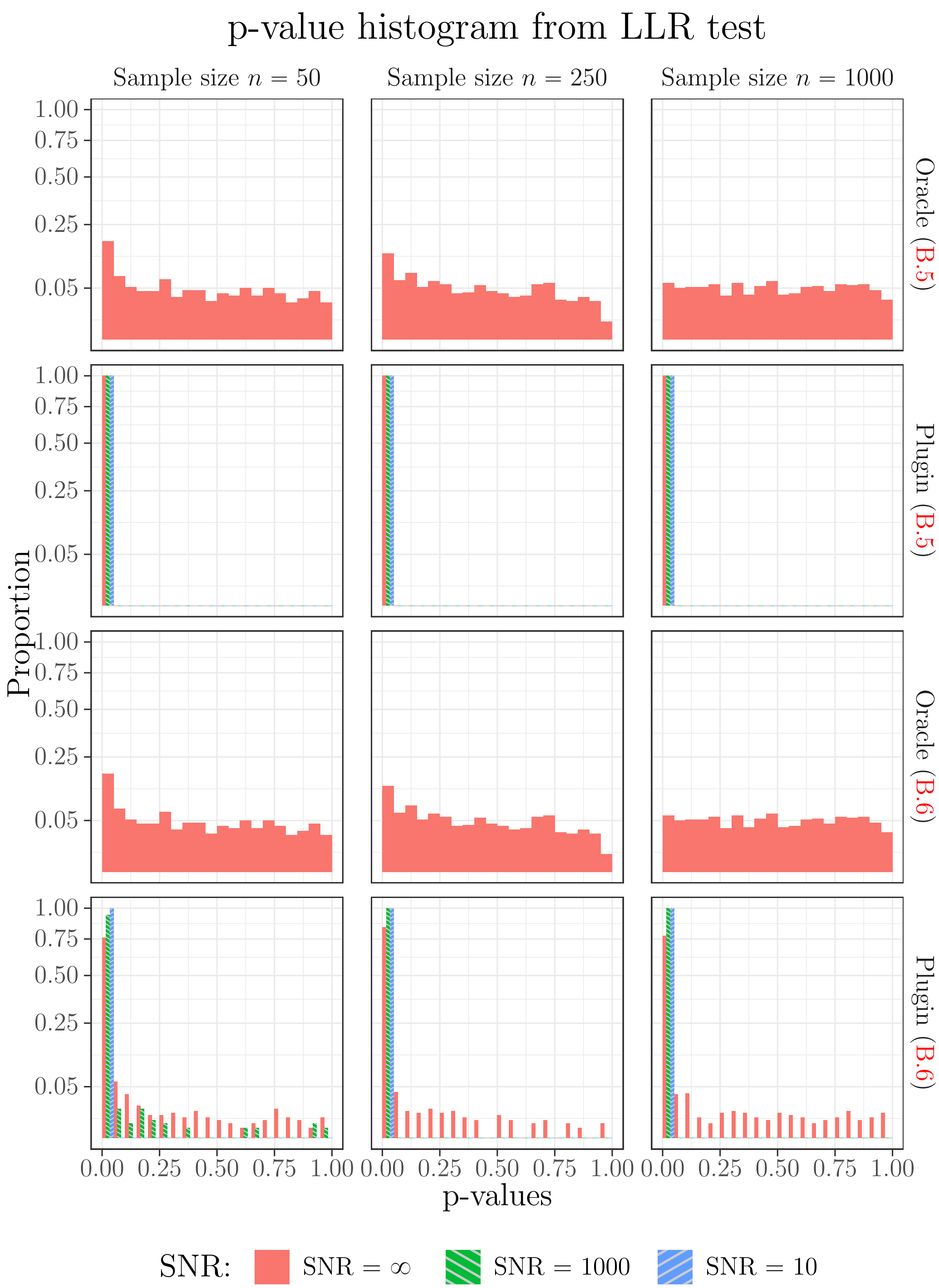

Figure B.1: The histograms of p-values for the LLR test. The first / third row are results based on Proposition B.3 given the exact polynomial basis (B.5) / eigenvector basis (B.6), while the second / fourth row are obtained by plugging estimated $\mathcal{A}_2(\rho)$ / optimized $\widehat{V}$ into (B.5) / (B.6) for implementations of the first / third row.

C Extension to symmetric matrices

As mentioned in the introduction, the analysis of common eigenvectors has many applications for symmetric matrices, for example, CPCA. The test methods introduced in this work can be implemented for symmetric matrices as well if we take additional care of the assumptions in Section 2.2.

Suppose the matrices M_i, $i = 1, \ldots, p$, and their respective estimators A_i, $i = 1, \ldots, p$, are symmetric matrices. Denote the function $\operatorname{vech} : \mathbb{R}^{d\times d} \to \mathbb{R}^{d(d+1)/2}$ that converts a symmetric matrix A into a vector stacking only distinct elements columnwise. Then, from estimations for symmetric M_i, the available consistency statements are of the form

$$c(n) \operatorname{vech}(A_i - M_i) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \widetilde{\Sigma}_i)$$

with positive semi-definite $\widetilde{\Sigma}_i \in \mathbb{R}^{d(d+1)/2\times d(d+1)/2}$. There exists the duplication matrix $G_d \in \{0,1\}^{d^2\times d(d+1)/2}$ such that $G_d \operatorname{vech}(A) = \operatorname{vec}(A)$ for any symmetric $A \in \mathbb{R}^{d\times d}$; see Magnus and Neudecker [2019] for more details on such operations. Hence, we can obtain exactly the same setup as Assumption 1, since

$$c(n) \operatorname{vec}(A_i - M_i) = c(n)G_d \operatorname{vech}(A_i - M_i) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_i)$$

with $\Sigma_i = G_d\widetilde{\Sigma}_i G_d'$. It is then straightforward to implement the above test designs directly, except that we may require the input eigenvector matrix V (or $\widehat{V}$) to be orthogonal. Such orthogonal matrices can be obtained referring to existing optimization schemes like FG-algorithm by Flury and Gautschi [1986].

As we focus on the general asymmetric setting, our simulation study as well as the application section do not cover symmetric extensions.

D Proofs

We provide here the proofs of most theoretical results except Propositions 3.2, 4.2, 5.2 and B.3. The proofs of those results can be found in Appendix E since they are based on a generic result.

Proof of Lemma 2.1. Under the conditions that given $\varepsilon > 0$ and ε is not an eigenvalue of Σ, the mappings $\Sigma \mapsto \Sigma(\varepsilon)$, $\Sigma \mapsto \Sigma^+(\varepsilon)$ and $\Sigma \mapsto \operatorname{rk}(\Sigma;\varepsilon)$ are all at least locally continuous at Σ. Hence (6) follows by the continuous mapping theorem. □

Proof of Proposition 3.1. Introduce the two vectors

$$\boldsymbol{a}_0 = \big(\operatorname{vec}(A_1)', \operatorname{vec}(A_2)'\big)', \ \boldsymbol{m}_0 = \big(\operatorname{vec}(M_1)', \operatorname{vec}(M_2)'\big)'.$$

By Assumption 1,

$$c(n)(\boldsymbol{a}_0 - \boldsymbol{m}_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_0) \tag{D.1}$$

with $\Sigma_0 = \operatorname{blkdiag}(\Sigma_1, \Sigma_2)$. Recall that $[M_1, M_2] = M_1M_2 - M_2M_1$ and define the function $g : \mathbb{R}^{2d^2} \to \mathbb{R}^{d^2}$ such that for $\boldsymbol{x} = \big(\boldsymbol{x}_1', \boldsymbol{x}_2'\big)'$ with $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^{d^2}$,

$$g(\boldsymbol{x}) = \operatorname{vec}[\operatorname{mat}_d(\boldsymbol{x}_1), \operatorname{mat}_d(\boldsymbol{x}_2)].$$

Under H_0, we know that $g(\boldsymbol{m}_0) = \mathbf{0}$ and $g(\boldsymbol{a}_0) = \boldsymbol{\eta}_n = \mathrm{vec}[A_1, A_2]$, hence the asymptotic distribution of $\boldsymbol{\eta}_n$ can be derived via delta method.

Define

$$\nabla_g := \nabla g(\boldsymbol{m}_0) = \begin{pmatrix} \Lambda(M_2) \\ \Lambda(M_1) \end{pmatrix}.$$

Then, by delta method and (D.1),

$$c(n)\big(g(\boldsymbol{a}_0) - g(\boldsymbol{m}_0)\big) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma_\eta),$$

where $\Sigma_\eta = \nabla_g' \Sigma_0 \nabla_g$. Then (8) follows directly by the continuous mapping theorem. □

Proof of Proposition 3.2. This is a direct corollary from Theorem E.1 in Appendix E. □

Before we prove Proposition B.1, we introduce the following lemma.

Lemma D.1. *Suppose a matrix $C \in \mathbb{R}^{d\times d}$ has distinct real non-zero eigenvalues. Then any square matrix with the same eigenvectors as C can be expressed by polynomials of C with order less than d.*

Proof of Lemma D.1. Assume $C = VD_CU$ where $V = (v_1, \ldots, v_d)$ is the eigenvector matrix, $U = (u_1, \ldots, u_d)' = V^{-1}$, and D_C is the diagonal matrix whose diagonal elements are the corresponding eigenvalues. Then the matrices that also have V as the eigenvector matrix form the following linear space:

$$\mathfrak{S} = \{X : X = V(\sum_{i=1}^{d} \boldsymbol{a}_i E_i)U = \sum_{i=1}^{d} \boldsymbol{a}_i v_i u_i',\ \boldsymbol{a} = (\boldsymbol{a}_1, \ldots, \boldsymbol{a}_d) \in \mathbb{R}^d\},$$

where the matrix $E_i \in \mathbb{R}^{d\times d}$ has only one non-zero element 1 at the ith diagonal entry. Then the space $\mathfrak{S}$ has dimension d and linear basis $\{v_i u_i'\}_{i=1}^d$.

Since C has distinct real non-zero eigenvalues, according to Cayley-Hamilton theorem [Horn and Johnson, 2012, Theorem 2.4.3.2], the characteristic polynomial and the minimal polynomial of C coincide and have degree d. Hence the matrices $\{C^0 = I_d, C^1, \ldots, C^{d-1}\}$ form an independent basis and the polynomial space

$$\mathfrak{S}_P = \{X : X = \sum_{i=1}^{d} \boldsymbol{b}_i C^{i-1},\ \boldsymbol{b} = (\boldsymbol{b}_1, \ldots, \boldsymbol{b}_d) \in \mathbb{R}^d\}$$

has dimension d as well. In addition, since $\mathfrak{S}_P \subset \mathfrak{S}$ and both spaces have dimension d, it follows $\mathfrak{S}_P = \mathfrak{S}$. □

Proof of Proposition B.1. Lemma D.1 and its proof immediately imply that under the null hypothesis H_0, there exist coordinate vectors $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \mathbb{R}^d$, such that

$$\mathrm{vec}(M_1) = P_2 \boldsymbol{x}_1,\ \ \mathrm{vec}(M_2) = P_1 \boldsymbol{x}_2, \tag{D.2}$$

with matrix P_i, $i = 1, 2$, defined by either (B.5) or (B.6).

Then the maximization problem in (B.3) can be transformed to optimizing with respect to free vectors $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ in $\mathbb{R}^d$. By setting the first-order derivative of function L in (B.1) to zero, the estimators can then be solved explicitly as

$$\begin{aligned}\operatorname{vec}(\widehat{A}_1) &= P_2(P_2'\Sigma_1^+ P_2)^+ P_2'\Sigma_1^+ \operatorname{vec}(A_1),\\ \operatorname{vec}(\widehat{A}_2) &= P_1(P_1'\Sigma_2^+ P_1)^+ P_1'\Sigma_2^+ \operatorname{vec}(A_2).\end{aligned}$$

□

Proof of Proposition B.2. Since Σ_1 is a positive semidefinite matrix, we can find the low-rank square root $\Sigma_1^{+/2} \in \mathbb{R}^{r\times d^2}$ of its general inverse, such that $\Sigma_1^+ = (\Sigma_1^{+/2})'\Sigma_1^{+/2}$, where r is the rank of Σ_1. Then, by Assumption 1,

$$c(n)\Sigma_1^{+/2}\operatorname{vec}(A_1 - M_1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I_r).$$

Set $G = I_r - \Sigma_1^{+/2}P_2(P_2'\Sigma_1^+P_2)^+P_2'(\Sigma_1^{+/2})'$, then under H_0, the first summand of Γ_2 in (B.7) is

$$\begin{aligned}&c^2(n)\operatorname{vec}(A_1)'Q_{1,2}\operatorname{vec}(A_1)\\ &= c^2(n)\operatorname{vec}(A_1)'\Big(\Sigma_1^+ - \Sigma_1^+P_2(P_2'\Sigma_1^+P_2)^+P_2'\Sigma_1^+\Big)\operatorname{vec}(A_1)\\ &= c^2(n)\big(\Sigma_1^{+/2}\operatorname{vec}(A_1)\big)'G\big(\Sigma_1^{+/2}\operatorname{vec}(A_1)\big)\\ &= c^2(n)\big(\Sigma_1^{+/2}\operatorname{vec}(A_1 - M_1)\big)'G\big(\Sigma_1^{+/2}\operatorname{vec}(A_1 - M_1)\big).\end{aligned} \tag{D.3}$$

The equality in (D.3) uses the fact that $G\Sigma_1^{+/2}\operatorname{vec}(M_1) = G\Sigma_1^{+/2}P_2\boldsymbol{x}_1 = 0$. In addition, the matrices G and $I_r - G$ are both projection matrices, and $I_r - G$ projects matrices onto the column space of $\Sigma_1^{+/2}P_2$, hence $\operatorname{rk}(I_r - G) = \operatorname{rk}(P_2'\Sigma_1^+P_2)$ and $\operatorname{rk}(G) = r - \operatorname{rk}(P_2'\Sigma_1^+P_2) = r_{1,2}$ with $r_{1,2}$ defined in Proposition B.2. Then the first summand in (B.7) satisfies

$$c^2(n)\big(\Sigma_1^{+/2}\operatorname{vec}(A_1 - M_1)\big)'G\big(\Sigma_1^{+/2}\operatorname{vec}(A_1 - M_1)\big) \xrightarrow{\mathcal{D}} \chi^2(r_{1,2}).$$

Similarly, the second summand in Γ_2, $c^2(n)\operatorname{vec}(A_2)'Q_{2,1}\operatorname{vec}(A_2)$, converges to a chi-square distribution with $r_{2,1}$ degrees of freedom, and since the two summands are independent, the result (B.8) follows. □

Proof of Proposition B.4. Denote

$$\widehat{\Sigma}_1^+(\varepsilon) = (\widehat{\Sigma}_{1,\varepsilon}^{+/2})'\widehat{\Sigma}_{1,\varepsilon}^{+/2}, \quad \text{with} \quad \widehat{\Sigma}_{1,\varepsilon}^{+/2} \in \mathbb{R}^{\hat{r}\times d^2},\ \hat{r}(\varepsilon) = \operatorname{rk}(\widehat{\Sigma}_1;\varepsilon) \tag{D.4}$$

and

$$G[\varepsilon] = I_{\hat{r}(\varepsilon)} - \widehat{\Sigma}_{1,\varepsilon}^{+/2}P_2\big(P_2'\widehat{\Sigma}_1^+(\varepsilon)P_2\big)^+P_2'(\widehat{\Sigma}_{1,\varepsilon}^{+/2})'.$$

Then, the first summand of $\Gamma_2^{\#}(\varepsilon)$ can be written as

$$\begin{aligned}&c^2(n)\big(\widehat{\Sigma}_{1,\varepsilon}^{+/2}\operatorname{vec}(A_1)\big)'G[\varepsilon]\big(\widehat{\Sigma}_{1,\varepsilon}^{+/2}\operatorname{vec}(A_1)\big)\\ &= c^2(n)\big(\widehat{\Sigma}_{1,\varepsilon}^{+/2}\operatorname{vec}(A_1 - M_1)\big)'G[\varepsilon]\big(\widehat{\Sigma}_{1,\varepsilon}^{+/2}\operatorname{vec}(A_1 - M_1)\big)\\ &\quad + 2c^2(n)\big(\widehat{\Sigma}_{1,\varepsilon}^{+/2}\operatorname{vec}(A_1 - M_1)\big)'G[\varepsilon]\widehat{\Sigma}_{1,\varepsilon}^{+/2}\boldsymbol{m}_1\\ &\quad\quad + c^2(n)\boldsymbol{m}_1'(\widehat{\Sigma}_{1,\varepsilon}^{+/2})'G[\varepsilon]\widehat{\Sigma}_{1,\varepsilon}^{+/2}\boldsymbol{m}_1.\end{aligned} \tag{D.5}$$

The equality in (D.5) uses the fact that

$$G[\varepsilon]\widehat{\Sigma}_{1,\varepsilon}^{+/2}\operatorname{vec}(M_1) = G[\varepsilon]\widehat{\Sigma}_{1,\varepsilon}^{+/2}\boldsymbol{m}_1.$$

In addition, it can be verified that $\boldsymbol{m}_1'\widehat{\Sigma}_1^{+}(\varepsilon)P_2 = 0$, then

$$\boldsymbol{m}_1'(\widehat{\Sigma}_{1,\varepsilon}^{+/2})'G[\varepsilon]\widehat{\Sigma}_{1,\varepsilon}^{+/2}\boldsymbol{m}_1 = \boldsymbol{m}_1'\widehat{\Sigma}_1^{+}(\varepsilon)\boldsymbol{m}_1 > 0,$$

since $\widehat{\Sigma}_1^{+}(\varepsilon)\boldsymbol{m}_1 \neq 0$. Hence, the deterministic leading term in (D.5), $c^2(n)\boldsymbol{m}_1'\widehat{\Sigma}_1^{+}(\varepsilon)\boldsymbol{m}_1$, goes to infinity as $c(n)$ goes to infinity. Similar arguments can be used for the second summand of $\Gamma_2^{\#}(\varepsilon)$ and the proposition is proved. □

Proof of Proposition B.5. For Δ_ε in (B.13), use the notation of $\widehat{\Sigma}_{1,\varepsilon}^{+/2}$ introduced in (D.4) to simplify

$$\widehat{Q}_{1,2}[\varepsilon] - \widehat{\mathcal{Q}}_{1,2}[\varepsilon] = \big(\widehat{\Sigma}_{1,\varepsilon}^{+/2}\big)'(\Delta\widehat{Q}_\varepsilon)\widehat{\Sigma}_{1,\varepsilon}^{+/2}, \tag{D.6}$$

where

$$\Delta\widehat{Q}_\varepsilon = \widehat{Q}_\varepsilon(P_2) - \widehat{Q}_\varepsilon(\widehat{P}_2),$$

with $\widehat{Q}_\varepsilon : \mathbb{R}^{d^2\times d} \to \mathbb{R}^{\hat{r}(\varepsilon)\times\hat{r}(\varepsilon)}$ a matrix-valued projection function given by

$$\widehat{Q}_\varepsilon(P) := \widehat{\Sigma}_{1,\varepsilon}^{+/2}P\big(P'\widehat{\Sigma}_1^{+}(\varepsilon)P\big)^{+}P'\big(\widehat{\Sigma}_{1,\varepsilon}^{+/2}\big)'.$$

Then, the first-order-perturbation approximation can be calculated as

$$\begin{aligned}
\Delta_\varepsilon &= c^2(n)\operatorname{vec}(A_1)'(\widehat{Q}_{1,2}[\varepsilon] - \widehat{\mathcal{Q}}_{1,2}[\varepsilon])\operatorname{vec}(A_1)\\
&= c^2(n)\operatorname{vec}(A_1)'\big(\widehat{\Sigma}_1^{+/2}(\varepsilon)\big)'(\Delta\widehat{Q}_\varepsilon)\widehat{\Sigma}_1^{+/2}(\varepsilon)\operatorname{vec}(A_1)\\
&= 2c^2(n)\operatorname{vec}(A_1)'\widehat{\Sigma}_1^{+}(\varepsilon)P_2\big(P_2'\widehat{\Sigma}_1^{+}(\varepsilon)P_2\big)^{+}(\Delta P_2)'(\widehat{\Sigma}_{1,\varepsilon}^{+/2})'\times\\
&\qquad \big(I_{\hat{r}} - \widehat{Q}_\varepsilon(P_2)\big)\widehat{\Sigma}_{1,\varepsilon}^{+/2}\operatorname{vec}(A_1 - M_1) + o_{\mathcal{P}}\big(c^2(n)\Delta P_2\big)\\
&= 2c^2(n)\operatorname{vec}(A_1)'\Sigma_1^{+}P_2\big(P_2'\Sigma_1^{+}P_2\big)^{+}(\Delta P_2)'(\Sigma_1^{+/2})'\times\\
&\qquad G\Sigma_1^{+/2}\operatorname{vec}(A_1 - M_1) + o_{\mathcal{P}}\big(c^2(n)\Delta P_2\big)
\end{aligned} \tag{D.7}$$

where $\Delta P_2 = \widehat{P}_2 - P_2$. Here in the last equality, we use the assumption that $\Sigma_1(\varepsilon) = \Sigma_1$, hence $\widehat{\Sigma}_1^{+}(\varepsilon) = \Sigma_1^{+} + o_{\mathcal{P}}(1)$, $\widehat{\Sigma}_{1,\varepsilon}^{+/2} = \Sigma_1^{+/2} + o_{\mathcal{P}}(1)$, and $I_{\hat{r}} - \widehat{Q}_\varepsilon(P_2) = G + o_{\mathcal{P}}(1)$.

From now on we omit the $o_{\mathcal{P}}\big(c^2(n)\Delta P_2\big)$ term of Δ_ε in (D.7) for the following proof since we are only interested in its stochastic limit.

Then, assume $G = U'U$ where $U \in \mathbb{R}^{r_{1,2}\times r}$, $UU' = I_{r_{1,2}}$, define

$$\boldsymbol{w} = (\boldsymbol{y}', \boldsymbol{z}')', \quad \boldsymbol{y} := U\Sigma_1^{+/2}\Delta P_2\big(P_2'\Sigma_1^{+}P_2\big)^{+}P_2'\Sigma_1^{+}\operatorname{vec}(A_1), \quad \boldsymbol{z} := U\Sigma_1^{+/2}\operatorname{vec}(A_1 - M_1).$$

By Assumption 1, A_1 converges in probability to the true matrix M_1, and $\widehat{P}_2$ is a consistent estimator of P_2 with rate $c(n)$, then according to delta method and Slutsky's theorem, we can define the matrix $\check{\Sigma} = \check{\Sigma}(M_1, M_2, \Sigma_1, \Sigma_2)$ in Proposition B.5 to satisfy

$$c(n)\boldsymbol{y} = c(n)U\Sigma_1^{+/2}\Delta P_2\big(P_2'\Sigma_1^{+}P_2\big)^{+}P_2'\Sigma_1^{+}\operatorname{vec}(A_1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \check{\Sigma}).$$

In addition, we have

$$c(n)\boldsymbol{z} = c(n)U\Sigma_1^{+/2}\mathrm{vec}(A_1 - M_1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, I_{r_{1,2}}).$$

And note that due to the independence between A_1 and ΔP_2, the following expectation can be decomposed to

$$\mathbb{E}\Big[\big(c(n)\boldsymbol{z}\big)\big(c(n)\boldsymbol{y}\big)'\Big] = U\Sigma_1^{+/2}\mathbb{E}\Big[c(n)f(A_1)\Big]\mathbb{E}\big[c(n)\Delta P_2'\big](\Sigma_1^{+/2})'U', \tag{D.8}$$

where $f(A_1) = \mathrm{vec}(A_1 - M_1)\big(\mathrm{vec}(A_1)\big)'\Sigma_1^+ P_2\big(P_2'\Sigma_1^+ P_2\big)^+$ is a matrix whose only randomness is from A_1. And since $c(n)\Delta P_2'$ has asymptotically zero mean, the expectation (D.8) is asymptotically zero, which indicates that $c(n)\boldsymbol{y}$ and $c(n)\boldsymbol{z}$ are asymptotically uncorrelated. Hence we have

$$c(n)\boldsymbol{w} = c(n)\begin{pmatrix}\boldsymbol{y}\\ \boldsymbol{z}\end{pmatrix} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0,\ \begin{pmatrix}\check{\Sigma} & 0\\ 0 & I_{r_{1,2}}\end{pmatrix}\right),$$

and consequently by the continuous mapping theorem,

$$\Delta_\varepsilon = 2c^2(n)\boldsymbol{w}'\begin{pmatrix}0 & I_{r_{1,2}}\\ I_{r_{1,2}} & 0\end{pmatrix}\boldsymbol{w} \xrightarrow{\mathcal{D}} Z, \quad \text{with} \quad Z = \sum_{i=1}^{r_{1,2}} 2\sqrt{\lambda_i}(\nu_{i,1} - \nu_{i,2}),$$

where λ_i are the eigenvalues of $\check{\Sigma}$ and $\nu_{i,j} \overset{i.i.d.}{\sim} \chi^2(1)$ for $i \in \{1, \dots, r_{1,2}\}$, and $j = 1, 2$. In addition, note that the variance of a $\chi^2(1)$ distributed random variable is 2. For this reason,

$$\mathrm{Var}(Z) = \sum_{i=1}^{r_{1,2}} 4\lambda_i\big(\mathrm{Var}(\nu_{i,1}) + \mathrm{Var}(\nu_{i,2})\big) = 16\sum_{i=1}^{r_{1,2}}\lambda_i = 16\,\mathrm{tr}(\check{\Sigma}).$$

□

E Generalized Wald test

In this section we provide a generic result to prove Propositions 3.2, 4.2, 5.2 and B.3. In fact, Theorem E.1 below combined together with the asymptotic normality results proved in Propositions 3.1, 4.1, 5.1 and B.2 gives Propositions 3.2, 4.2, 5.2 and B.3.

Consider an estimated vector $\widehat{\boldsymbol{v}}$ that depends on the sample size n and satisfies the asymptotic normality statement

$$c(n)(\widehat{\boldsymbol{v}} - \boldsymbol{v}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Sigma) \tag{E.1}$$

with deterministic $\boldsymbol{v}$ and positive semidefinite limiting covariance matrix $\Sigma \in \mathbb{R}^{\tau\times\tau}$. Then, a Wald-type test statistic can be defined as

$$\Gamma = c^2(n)(\widehat{\boldsymbol{v}} - \boldsymbol{v})'\Sigma^+(\widehat{\boldsymbol{v}} - \boldsymbol{v}) \xrightarrow{\mathcal{D}} \chi^2(\mathrm{rk}(\Sigma)). \tag{E.2}$$

If Σ is singular, the continuous mapping theorem is no longer applicable to justify replacing Σ by a consistent estimator $\widehat{\Sigma}$ in (E.2). In order to achieve a convergence result for Γ in (E.2) based on an estimator for Σ, we introduce the following theorem.

Theorem E.1. *Assume (E.1) holds and $\widehat{\Sigma}$ is a consistent estimator of Σ, threshold $\varepsilon > 0$ is not an eigenvalue of Σ. Then, alternative to (E.2), we propose statistics with truncated SVD introduced in Section 2.3,*

$$\Gamma(\varepsilon) = c^2(n)(\widehat{\boldsymbol{v}} - \boldsymbol{v})'\Sigma^+(\varepsilon)(\widehat{\boldsymbol{v}} - \boldsymbol{v}) \xrightarrow{\mathcal{D}} \chi^2\big(\operatorname{rk}(\Sigma;\varepsilon)\big), \tag{E.3}$$

and

$$\Gamma^{\#}(\varepsilon) = c^2(n)(\widehat{\boldsymbol{v}} - \boldsymbol{v})'\widehat{\Sigma}^+(\varepsilon)(\widehat{\boldsymbol{v}} - \boldsymbol{v}) \xrightarrow{\mathcal{D}} \chi^2\big(\operatorname{rk}(\widehat{\Sigma};\varepsilon)\big). \tag{E.4}$$

Indeed if the threshold ε is less than the smallest non-zero eigenvalue of Σ, especially when at the extreme case $\varepsilon = 0$, the limiting distribution in (E.3) is exactly the same as (E.2).

Proof of Theorem E.1. Assume SVD gives $\Sigma = Z\Pi Z'$ with orthogonal Z, then by definition of the truncated SVD,

$$\Sigma \times \Sigma^+(\varepsilon) = \Sigma^+(\varepsilon) \times \Sigma = Z \begin{pmatrix} I_{\operatorname{rk}(\Sigma;\varepsilon)} & 0 \\ 0 & 0 \end{pmatrix} Z'$$

is an idempotent matrix with rank $\operatorname{rk}(\Sigma;\varepsilon)$. Then, (E.3) follows immediately by the continuous mapping theorem. According to (6) in Lemma 2.1, we can find consistent estimators for the generalized inverse as well as for the rank of Σ. Lemma 2.1, (E.3) and the continuous mapping theorem prove (E.4). $\square$

References

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.

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

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