Supplement to "Hypothesis Testing in High-Dimensional Linear Regression: a Normal-Reference Scale-Invariant Test"

Tianming Zhu, Liang Zhang, and Jin-Ting Zhang

Dept of Statist. and Appl. Prob., National Univ. of Singapore, Singapore

Supplementary Material

S1 Histograms of simulated T_{YS}

In Section 1, we mentioned that the null distribution of Yamada and Srivastava (2012)'s test statistic T_{YS} may not be approximately normal if it is blindly applied. To show this, in Figure S.1, we display the histograms of the simulated T_{YS} under the null hypothesis for a three-sample one-way MANOVA problem. It is seen that the shapes of the histograms are mainly controlled by the value of the tuning parameter ρ which determines the correlation among the p-variables of the generated high-dimensional data, that is, the larger the value of ρ is, the larger the correlation among the p-variables. When $\rho = 0.01$, the histograms are quite symmetric and bell-shaped, showing that a normal approximation to the null distribution of T_{YS}

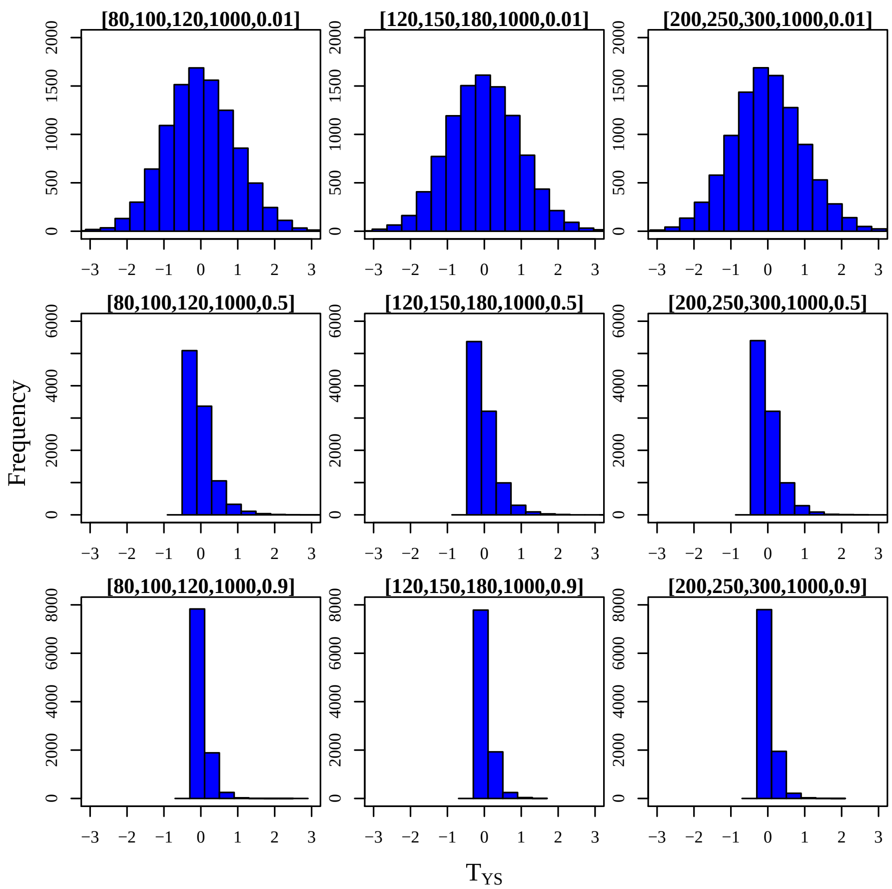

Figure S.1: *Histograms of the simulated T_{YS} under the null hypothesis for a three-sample one-way MANOVA problem. The three samples are independently generated from $N_p(\mathbf{0}, \mathbf{\Sigma})$ with $\mathbf{\Sigma} = \mathbf{D}^{1/2}\mathbf{R}\mathbf{D}^{1/2}$, $\mathbf{D} = diag(d_1^2, \ldots, d_p^2), d_k = (p-k+1)/p, k = 1, \ldots, p, \mathbf{R} = (r_{k\ell})$ with $r_{k\ell} = (-1)^{k+\ell}\rho^{0.01|k-\ell|}$, $k, \ell = 1, \ldots, p$, for various tuning parameters [n_1, n_2, n_3, p, ρ] showed in the subtitle of each panel.*

as suggested by Yamada and Srivastava (2012) is adequate. However, when $\rho = 0.5$ and 0.9, the histograms are quite skewed, showing that a normal approximation to the null distribution of T_{YS} is no longer applicable since the underlying null distribution of T_{YS} is actually skewed while a normal distribution is always symmetric and bell-shaped.

S2 A simulation study on Condition (2.3)

Scale-invariant tests are generally more powerful than non-scale-invariant tests but this is not a free lunch since scale-invariant tests often require larger sample sizes to work well than non-scale-invariant tests. This is because scale-invariant tests need to estimate the variances $\sigma_{rr}, r = 1, \ldots, p$ of the p-variables of the high-dimensional data accurately to take the variations of the p-variables into account. In Remark 1, we mentioned that Condition (2.3) is crucial for our normal-reference scale-invariant test $T_{n,p}$ to work well. It requires n should not be too small compared with $\log(p)$. If the sample size n is too small compared with $\log(p)$, we cannot estimate σ_{rr}'s with $\hat{\sigma}_{rr}$'s accurately as indicated by (2.1) and we cannot reduce studying the distribution of $T_{n,p}$ to studying the distribution of $T^*_{n,p}$ as indicated by (2.2). Therefore, if Condition (2.3) is violated, $T_{n,p}$ may not have a good size control.

A question arises naturally. How large is the sample size n, compared with $\log(p)$, needed for $T_{n,p}$ to work well? To answer this question and to address a comment from an anonymous reviewer, we conduct a small simulation study using the setup of Simulation 1 in Section 4.1. We consider five cases of $[n/\log(p)]$ as $[n/\log(p)] \in \{5, 10, 20, 30, 40\}$, where $[x]$ returns the nearest integer to x to compare the performance of $T_{n,p}$ against T_{FHW} and T_{SF}, the two non-scale-invariant tests developed by Fujikoshi et al. (2004) and Srivastava and Fujikoshi (2006), respectively, and T_{YS}, the scale-invariant test developed by Yamada and Srivastava (2012), in terms of size control.

The empirical sizes of the four tests are shown in Table S.1. For each value of $[n/\log(p)]$, we display the ARE values of the four tests associated with three values of ρ. We first investigate the case when $\rho = 0.01$, i.e., when the data are nearly uncorrelated. In this case, the null distributions of all the four tests can be adequately approximated by the normal approximation if the sample size n is sufficiently large. It is seen that when the sample size is too small compared with $\log(p)$, i.e., when $[n/\log(p)] = 5, 10$, $T_{n,p}$ is rather liberal with its empirical sizes generally larger than 5% and some of them even larger than 9% and T_{YS} is rather conservative with its empirical sizes generally smaller than 5% and some of them even smaller

Table S.1: *Empirical sizes (in %, Simulation on Condition (2.3)).*

			$\rho=0.01$				$\rho=0.55$				$\rho=0.95$			
$[\frac{n}{\log(p)}]$	Model	$[p,n_1,n_2,n_3]$	T_{FHW}	T_{SF}	T_{YS}	$T_{n,p}$	T_{FHW}	T_{SF}	T_{YS}	$T_{n,p}$	T_{FHW}	T_{SF}	T_{YS}	$T_{n,p}$
5	1	[200,7,9,10]	6.44	6.80	4.99	9.59	7.71	8.06	4.11	8.91	9.10	9.36	1.68	8.27
		[500,8,10,12]	5.77	5.98	3.17	8.77	7.03	7.27	3.22	8.73	8.65	8.84	1.41	8.42
		[1000,9,11,14]	5.44	5.69	2.34	8.06	6.37	6.60	2.38	8.32	8.52	8.71	1.20	8.24
	2	[200,7,9,10]	4.72	5.00	3.65	7.55	7.34	7.69	3.93	8.75	9.19	9.47	1.70	8.00
		[500,8,10,12]	4.86	5.15	2.42	6.95	6.08	6.33	2.76	7.87	8.93	9.18	1.29	8.28
		[1000,9,11,14]	4.51	4.66	1.64	6.34	5.96	6.21	1.99	7.39	8.11	8.34	1.19	8.46
	3	[200,7,9,10]	5.89	6.22	3.42	7.13	7.07	7.35	3.56	8.49	8.87	9.06	1.78	8.30
		[500,8,10,12]	5.24	5.50	2.04	6.14	6.83	7.09	3.00	8.41	8.11	8.29	1.37	7.85
		[1000,9,11,14]	4.94	5.16	1.34	5.32	6.45	6.73	2.11	7.70	7.88	8.08	1.19	7.85
	ARE		**10.56**	**12.98**	**44.42**	**46.33**	**35.20**	**40.73**	**39.87**	**65.71**	**71.91**	**76.29**	**71.53**	**63.71**
10	1	[200,14,18,20]	6.38	6.52	4.79	7.36	7.30	7.40	3.66	7.43	7.91	8.03	1.04	6.54
		[500,16,20,24]	5.31	5.44	3.56	6.60	7.06	7.16	3.49	7.37	8.04	8.19	0.97	6.75
		[1000,18,22,28]	5.58	5.71	3.35	6.52	6.13	6.22	2.37	6.42	7.52	7.66	0.75	6.85
	2	[200,14,18,20]	5.04	5.18	4.02	6.43	6.98	7.03	3.37	6.93	7.92	7.99	0.73	6.31
		[500,16,20,24]	4.45	4.52	3.33	6.02	6.52	6.66	3.28	6.75	7.21	7.34	1.02	6.37
		[1000,18,22,28]	4.66	4.76	2.39	5.57	5.61	5.75	2.63	6.07	7.22	7.42	0.73	6.56
	3	[200,14,18,20]	5.36	5.53	3.72	5.95	7.51	7.67	3.71	7.51	8.02	8.09	0.95	6.49
		[500,16,20,24]	5.42	5.63	3.51	6.10	6.87	7.04	3.33	6.96	7.63	7.73	0.85	6.51
		[1000,18,22,28]	4.97	5.09	2.45	5.39	6.17	6.30	2.85	6.41	7.50	7.57	1.04	6.74
	ARE		**8.91**	**10.71**	**30.84**	**24.31**	**33.67**	**36.07**	**36.24**	**37.44**	**53.27**	**55.60**	**82.04**	**31.38**
20	1	[200,28,36,40]	5.75	5.85	4.48	5.91	6.96	7.03	3.41	6.06	7.79	7.88	0.72	6.25
		[500,32,40,48]	5.83	5.93	4.25	5.67	6.48	6.57	3.36	5.97	7.47	7.52	0.80	5.89
		[1000,36,44,56]	5.59	5.68	3.93	5.91	6.04	6.10	3.00	5.48	7.38	7.41	0.85	6.03
	2	[200,28,36,40]	5.65	5.71	4.55	5.68	7.06	7.12	3.47	6.35	7.42	7.51	0.79	5.78
		[500,32,40,48]	4.64	4.75	3.79	5.32	6.19	6.26	3.53	6.20	7.53	7.61	0.81	5.87
		[1000,36,44,56]	4.97	5.02	3.53	5.35	6.18	6.23	3.34	5.83	6.94	7.01	0.61	5.82
	3	[200,28,36,40]	5.56	5.61	4.33	5.73	6.38	6.43	3.04	5.71	7.22	7.28	0.50	5.79
		[500,32,40,48]	5.65	5.70	4.39	5.92	6.51	6.55	3.11	5.48	6.92	6.97	0.65	6.03
		[1000,36,44,56]	5.24	5.31	3.60	5.09	5.97	6.05	3.32	5.69	7.18	7.25	0.79	5.78
	ARE		**10.36**	**11.24**	**18.11**	**12.40**	**28.38**	**29.64**	**34.27**	**17.27**	**46.33**	**47.64**	**85.51**	**18.31**
30	1	[200,42,54,60]	6.23	6.31	4.92	5.84	6.68	6.81	3.38	5.91	7.32	7.34	0.63	5.53
		[500,48,60,72]	5.69	5.72	4.39	5.56	6.20	6.24	3.87	6.21	7.34	7.36	0.59	6.16
		[1000,54,66,84]	4.66	4.71	3.77	5.03	5.57	5.64	3.20	5.33	6.88	6.93	0.85	5.66
	2	[200,42,54,60]	5.41	5.47	4.62	5.59	6.91	6.95	3.26	5.95	6.91	6.94	0.47	5.18
		[500,48,60,72]	5.40	5.49	4.24	5.19	6.37	6.40	3.33	5.64	7.13	7.16	0.62	5.43
		[1000,54,66,84]	4.53	4.57	3.63	4.93	5.78	5.81	3.44	5.41	6.68	6.70	0.70	5.55
	3	[200,42,54,60]	5.49	5.50	4.25	5.19	6.50	6.57	3.23	6.32	7.63	7.65	0.69	5.64
		[500,48,60,72]	5.53	5.58	4.06	4.97	5.93	5.98	3.11	5.07	6.85	6.89	0.53	5.19
		[1000,54,66,84]	5.64	5.67	4.09	5.26	6.11	6.18	3.47	5.44	6.75	6.77	0.77	5.39
	ARE		**11.56**	**12.13**	**15.62**	**6.13**	**24.56**	**25.73**	**32.69**	**13.96**	**41.09**	**41.64**	**87.00**	**10.51**
40	1	[200,56,72,80]	5.84	5.84	5.25	5.92	7.02	7.07	3.28	5.74	7.01	7.02	0.59	5.40
		[500,64,80,96]	5.65	5.69	4.75	5.48	6.77	6.80	3.71	5.87	7.18	7.18	0.56	5.57
		[1000,72,88,112]	5.50	5.51	4.42	5.31	6.05	6.07	3.69	5.33	6.68	6.72	0.74	5.41
	2	[200,56,72,80]	5.72	5.74	4.81	5.42	7.06	7.09	3.42	5.46	7.29	7.34	0.75	5.51
		[500,64,80,96]	5.32	5.34	4.33	5.07	6.12	6.15	3.31	5.42	7.35	7.38	0.82	5.73
		[1000,72,88,112]	4.69	4.72	3.76	4.62	6.25	6.30	3.60	5.54	6.92	6.92	0.70	5.31
	3	[200,56,72,80]	5.96	6.00	5.16	5.71	6.88	6.93	3.09	5.53	7.46	7.47	0.60	5.77
		[500,64,80,96]	5.35	5.42	4.52	5.32	6.00	6.03	3.42	5.37	7.46	7.48	0.56	5.89
		[1000,72,88,112]	5.47	5.50	4.34	5.37	6.08	6.12	3.73	5.51	6.53	6.58	0.53	5.32
	ARE		**11.38**	**11.82**	**9.96**	**8.84**	**29.40**	**30.13**	**30.56**	**10.60**	**41.96**	**42.42**	**87.00**	**10.91**

than 2%. However, in this case, both T_{FHW} and T_{SF} still perform well with their empirical sizes around 5% to 6%. This shows that scale-invariant tests indeed need larger sample sizes than non-scale-invariant tests to work well, as expected. It is also seen that when the sample size is sufficiently large compared with $\log(p)$, i.e., when $[n/\log(p)] = 20, 30, 40$, all the four tests perform reasonably well with their empirical sizes around 5%.

We next investigate the case when $\rho = 0.55$ and 0.95, i.e., when the data are moderately or highly correlated. In these two cases, the null distributions of all the four tests cannot be well approximated by the normal approximation even when the sample size n is sufficiently large. Therefore, in terms of size control, it is expected that $T_{\text{FHW}}, T_{\text{SF}}$ and T_{YS} will not perform well even when the sample size is sufficiently large, i.e., when $[n/\log(p)] = 20, 30, 40$. This is actually confirmed by the empirical sizes of the three tests presented in Table S.1 from which, it is seen that both T_{FHW} and T_{SF} are liberal with their empirical sizes larger than 5% and many of them around 7%, while T_{YS} is conservative with its empirical sizes smaller than 5% and many of them even smaller than 1%. Nevertheless, $T_{n,p}$ performs differently: when $[n/\log(p)] = 5, 10$, $T_{n,p}$ is rather liberal with its empirical sizes generally larger than 5% and some of them even larger than 8% and when $[n/\log(p)] = 20, 30, 40$, $T_{n,p}$ performs well with its empirical sizes around 5%.

The above simulation study partially demonstrates that scale-invariant tests require larger sample sizes than non-scale-invariant tests to work well and when the sample size n is large enough compared with $\log(p)$, i.e., when Condition (2.3) holds approximately, in terms of size control, $T_{n,p}$ performs well regardless of whether the data are nearly uncorrelated, moderately correlated, or highly correlated and it outperforms T_{FHW}, T_{SF} and T_{YS} substantially. This latter conclusion is consistent with those drawn from the simulation studies presented in Section 4.

S3 Comparing $T_{n,p}, T^*_{n,p}, T^*_{n,p,0}$ and $T^*_{p,0}$ by simulation

To address a comment from an anonymous reviewer, in this section, we compare $T_{n,p}, T^*_{n,p}, T^*_{n,p,0}$, and $T^*_{p,0}$ via their kernel density estimates (KDE) to assess if the theories established in Section 2 are valid, i.e., if the probability density functions (pdfs) of $T_{n,p}$ and $T^*_{n,p}$ are close to each other and if the pdfs of $T^*_{n,p,0}$ and $T^*_{p,0}$ are close to each other. Under Conditions C1–C4 and (2.3), as $n, p \to \infty$, we have (2.2), meaning that for large samples, the KDEs of $T_{n,p}$ and $T^*_{n,p}$ should be close to each other and under the conditions of Theorem 1, for large samples, the KDEs of $T^*_{n,p,0}$ and $T^*_{p,0}$ should be close to each other. In real data analysis, $T^*_{n,p}, T^*_{n,p,0}$ and $T^*_{p,0}$ are not

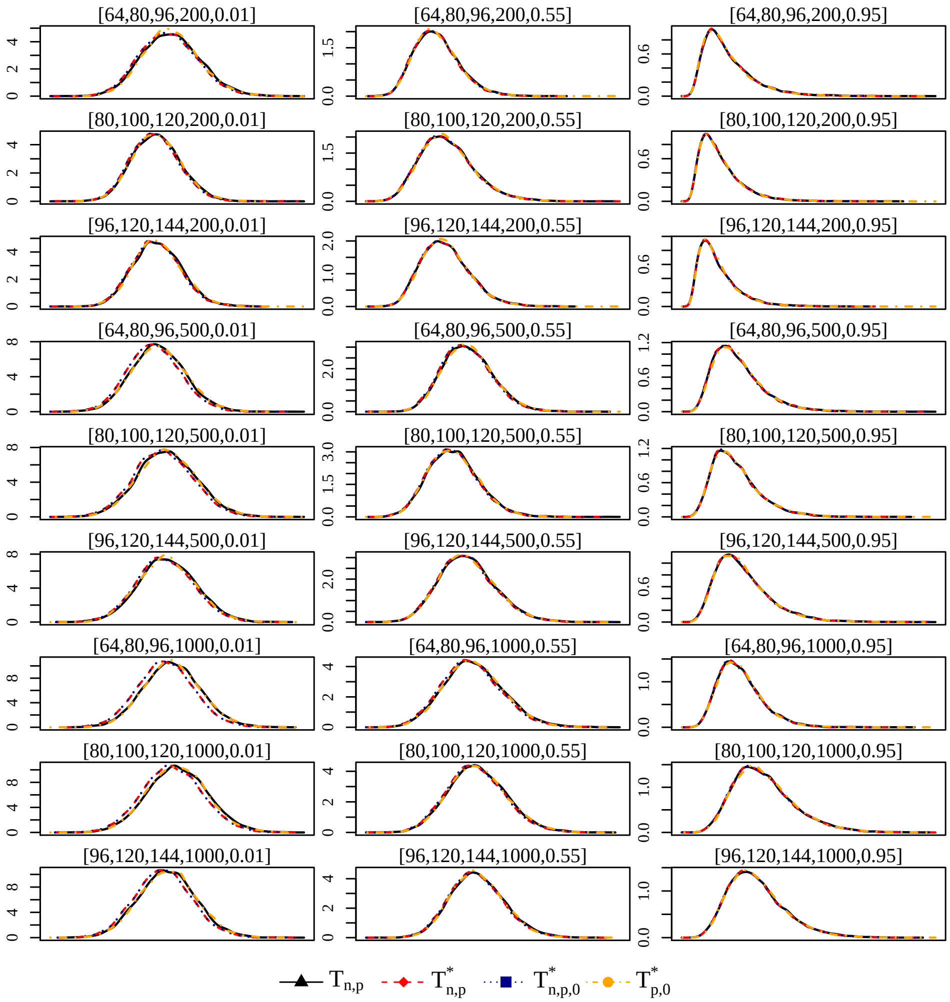

Figure S.2: *KDEs of the simulated* $T_{n,p}$*,* $T^*_{n,p}$*,* $T^*_{n,p,0}$ *and* $T^*_{p,0}$ *(*$T_{n,p}$*: black solid curves with triangles,* $T^*_{n,p}$*: red dashed curves with diamonds,* $T^*_{n,p,0}$*: dark blue dotted curves with squares,* $T^*_{p,0}$*: orange dot-dashed curves with circles) under the null hypothesis associated with parameters* $[n_1, n_2, n_3, p, \rho]$ *from the settings under Model 1 in Simulation 1 of Section 4.1.*

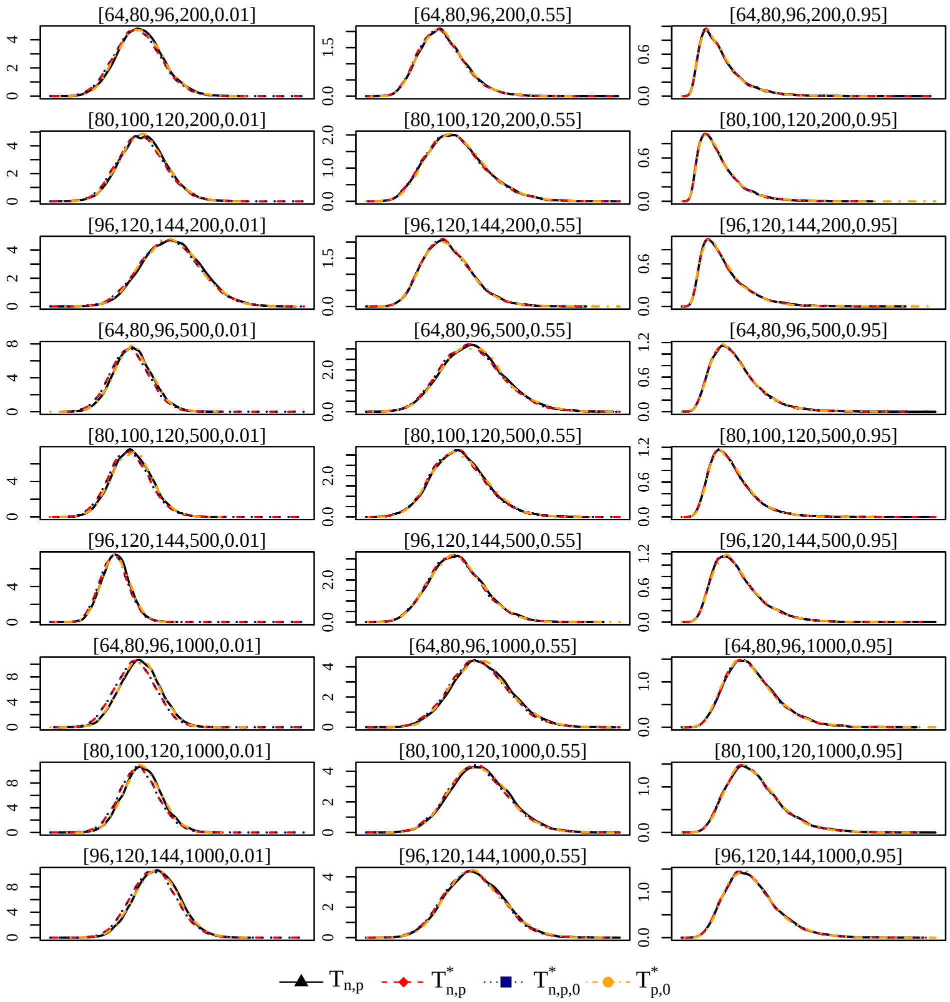

Figure S.3: *Same caption as that of Figure S.2 except under Model 2 in Simulation 1 of Section 4.1.*

computable but they are used to simplify the deriving process of the main results presented in Section 2.

We first generate the high-dimensional data using the setup of Simulation 1 in Section 4.1. Each time, based on the simulated data, we calculate

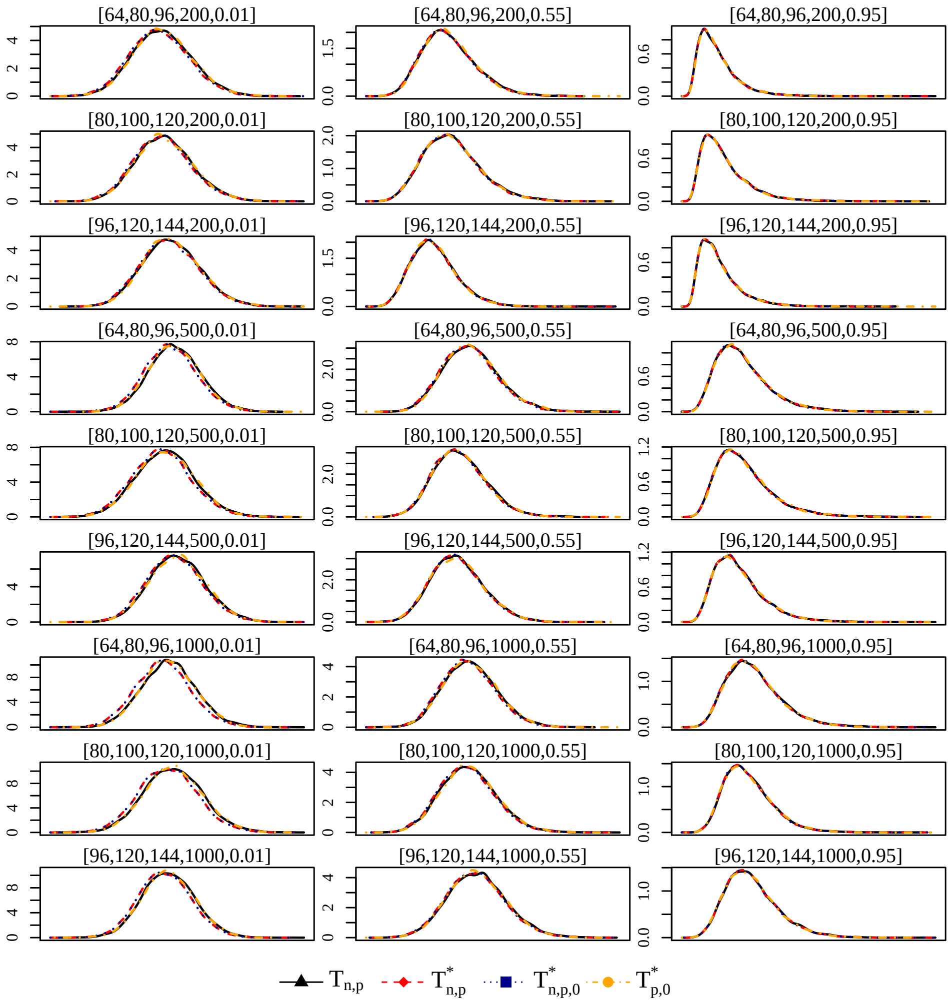

Figure S.4: *Same caption as that of Figure S.2 except under Model 3 in Simulation 1 of Section 4.1.*

the values of $T_{n,p}, T^*_{n,p}$, and $T^*_{n,p,0}$ using the formulas in (1.6), (2.2), and (2.6), respectively. That is,

$$T_{n,p} = \frac{n-k-2}{(n-k)pq}\mathrm{tr}(\mathbf{S}_h\hat{\mathbf{D}}^{-1}),\ T^*_{n,p} = \frac{n-k-2}{(n-k)pq}\mathrm{tr}(\mathbf{S}_h\mathbf{D}^{-1}),\ \text{and}$$
$$T^*_{n,p,0} = (pq)^{-1}\mathrm{tr}(\boldsymbol{\epsilon}^\top\mathbf{H}\boldsymbol{\epsilon}\mathbf{D}^{-1}). \tag{S3.1}$$

At the same time, to compute $T^*_{p,0} \overset{d}{=} (pq)^{-1}\sum_{r=1}^{p}\lambda_{p,r}A_r$, $A_r \overset{i.i.d.}{\sim} \chi^2_q$, we randomly generate $z_r \sim \chi^2_q, r = 1\ldots,p$ and compute $T^*_{p,0}$ as $(pq)^{-1}\sum_{r=1}^{p}\lambda_{p,r}z_r$. Note that in the above computations, we use the true values of $\mathbf{D}, \lambda_{p,r}, r = 1,\ldots,p$ which are not available in real data applications but they are available in the simulation studies. We repeat the above process for $10,000$ times so that for each of $T_{n,p}, T^*_{n,p}, T^*_{n,p,0}$, and $T^*_{p,0}$, we have an independent sample of size $10,000$ which can be used to compute the KDEs of $T_{n,p}, T^*_{n,p}, T^*_{n,p,0}$, and $T^*_{p,0}$ respectively. To compute the KDEs, we use the R-function "density" directly, i.e., the default kernel and bandwidth are used directly without adjustment.

We first compare the KDEs of $T_{n,p}, T^*_{n,p}, T^*_{n,p,0}$, and $T^*_{p,0}$ when the null hypothesis is valid. Figures S.2, S.3, and S.4 display the KDEs of the simulated $T_{n,p}, T^*_{n,p}, T^*_{n,p,0}$, and $T^*_{p,0}$ ($T_{n,p}$: black solid curves with triangles, $T^*_{n,p}$: red dashed curves with diamonds, $T^*_{n,p,0}$: dark blue dotted curves with squares, $T^*_{p,0}$: orange dot-dashed curves with circles) associated with parameters $[n_1, n_2, n_3, p, \rho]$ from the settings under Models 1, 2, and 3, respectively of Simulation 4.1. It is seen that the KDEs of the simulated $T_{n,p}, T^*_{n,p}, T^*_{n,p,0}$, and $T^*_{p,0}$ are nearly the same under various configurations. This is consistent with those theories established in Section 2. This is reasonable since we have $n/\log(p) = 35$ so that Condition (2.3) is approximately valid.

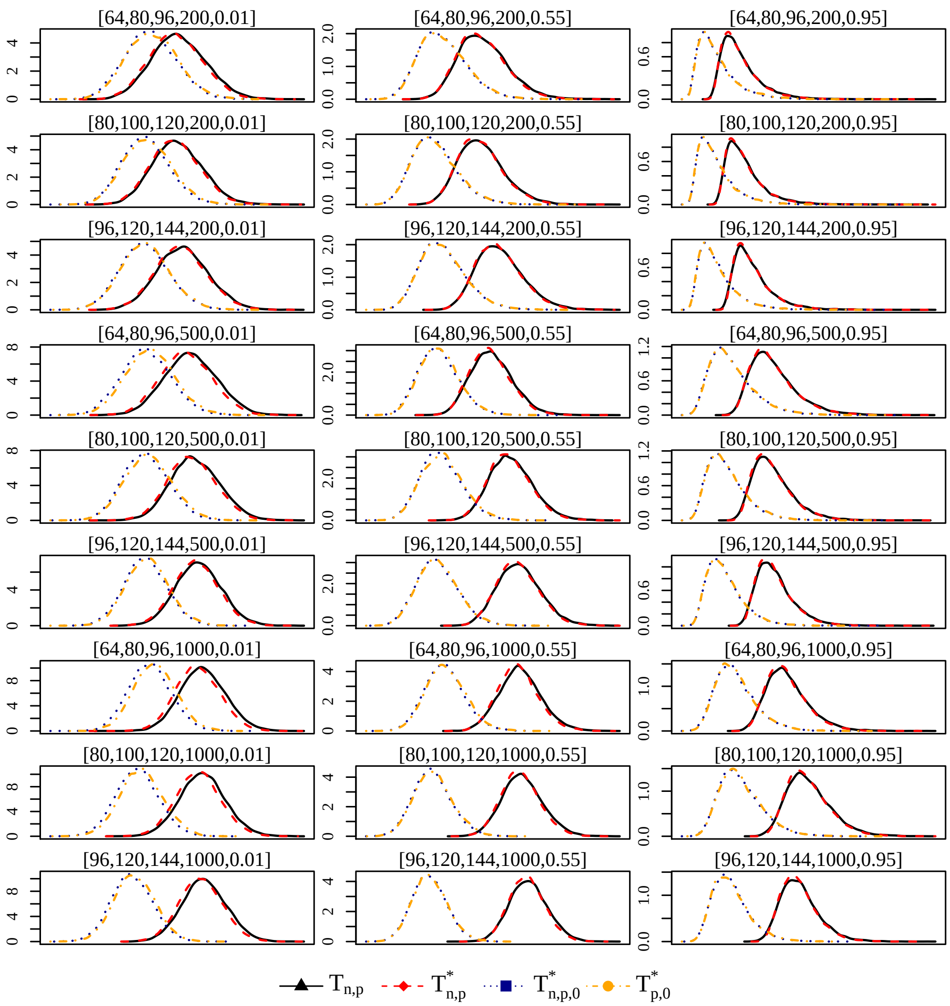

Figure S.5: *KDEs of* $T_{n,p}$, $T^*_{n,p}$, $T^*_{n,p,0}$ *and* $T^*_{p,0}$ *under an alternative hypothesis associated with parameters* $[n_1, n_2, n_3, p, \rho]$ *from the settings under Model 1 in Simulation 1 of Section 4.1.*

We next compare the KDEs of $T_{n,p}, T^*_{n,p}, T^*_{n,p,0}$, and $T^*_{p,0}$ when an alternative hypothesis is valid. Figures S.5 $\sim$ S.7 display the KDEs of the simulated $T_{n,p}$, $T^*_{n,p}$, $T^*_{n,p,0}$ and $T^*_{p,0}$ under an alternative hypothesis. It is

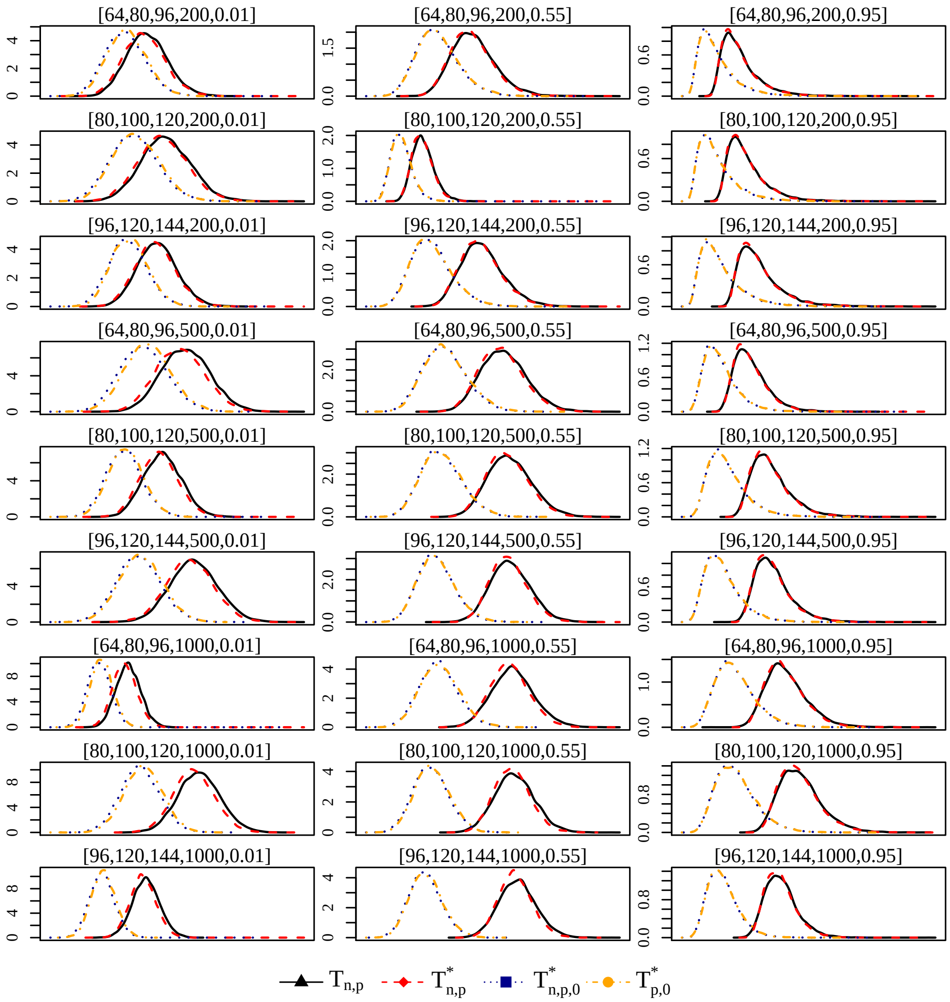

Figure S.6: *Same caption as that of Figure S.5 except under Model 2 in Simulation 1 of Section 4.1.*

seen that the KDEs of $T_{n,p}$ and $T^*_{n,p}$ are close to each other, and the KDEs $T^*_{n,p,0}$ and $T^*_{p,0}$ are close to each other. However, the KDEs of $T_{n,p}, T^*_{n,p}$ and those of $T^*_{n,p,0}, T^*_{p,0}$ are very different. This is not surprise since under an alternative hypothesis, $T^*_{n,p}$ and $T^*_{n,p,0}$ do not have the same distribution.

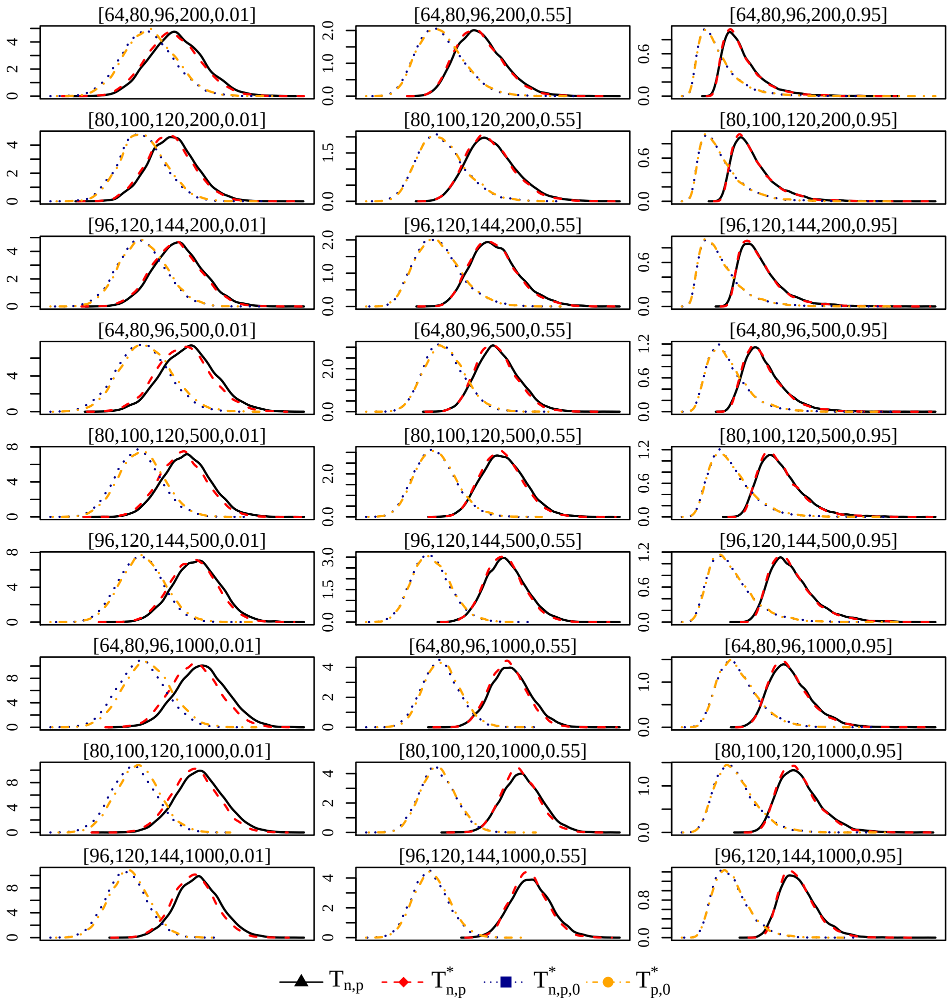

Figure S.7: *Same caption as that of Figure S.5 except under Model 3 in Simulation 1 of Section 4.1.*

From the above simulation study, we can see that for large samples, under null or alternative hypotheses, it is reasonable to approximate the distribution of $T_{n,p}$ using that of $T^*_{n,p}$ as guaranteed theoretically by (2.2) and it is reasonable to approximate the distribution of $T^*_{n,p,0}$ using that of

$T^*_{p,0}$ as guaranteed by Theorem 1.

S4 Comparison with some tests by Li et al. (2020)

To address a comment from an anonymous reviewer, we examine the performance of $T_{n,p}$ against the tests proposed by Li et al. (2020) for the GLHT problem (1.2) under the three-sample one-way MANOVA framework considered in Simulation 1. Li et al. (2020) proposed several tests with different shrinkage methods to regularize the spectrum of $\hat{\mathbf{\Sigma}}$. To save space, we only compare $T_{n,p}$ against the composite ridge-regularized tests of Algorithm 4.2 in Li et al. (2020), denoted respectively as LR_c, LH_c, and BNP_c, which are implemented via the R codes kindly provided by the first author of Li et al. (2020). Figures 1 $\sim$ 3 of Li et al. (2020) indicate that these three tests outperform or perform not worse than other tests proposed in their paper. Without loss of generality, we consider the contrast test $\mathrm{H}_0: \boldsymbol{\mu}_1 + 2\,\boldsymbol{\mu}_2 - 3\,\boldsymbol{\mu}_3 = \mathbf{0}$, which can be written in the form of the GLHT problem (1.2) with $\mathbf{\Theta} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \boldsymbol{\mu}_3)^\top$ and $\mathbf{C} = (1, 2, -3)$. Other tuning parameters are set to be the same as those specified in Simulation 1 except we now take $\delta = 0.03, 0.045, 0.07$ for $\rho = 0.01, 0.55, 0.95$ respectively.

Table S.2 displays the empirical sizes of the four tests under various configurations. It is seen that in terms of size control, $T_{n,p}$ performs well

Table S.2: *Empirical sizes (in %) of LR_c, LH_c, BNP_c and $T_{n,p}$.*

			$\rho = 0.01$				$\rho = 0.55$				$\rho = 0.95$			
Model	p	n_0	LR_c	LH_c	BNP_c	$T_{n,p}$	LR_c	LH_c	BNP_c	$T_{n,p}$	LR_c	LH_c	BNP_c	$T_{n,p}$
1	200	80	7.53	8.83	6.38	5.06	8.53	9.35	7.55	5.36	8.74	9.23	8.19	4.98
		100	7.58	8.54	6.71	5.32	7.99	8.62	7.39	5.37	8.89	9.32	8.58	4.57
		120	7.65	8.33	6.85	5.55	8.24	8.78	7.73	5.05	8.76	9.07	8.43	5.27
	500	80	7.66	8.80	6.61	5.73	7.97	9.22	6.85	5.39	8.65	9.49	8.06	5.39
		100	7.13	8.29	6.00	5.47	7.66	8.60	6.62	5.43	9.38	9.94	8.55	5.90
		120	6.66	7.84	5.61	5.09	7.63	8.46	6.75	5.35	8.41	8.99	7.98	5.12
	1000	80	7.67	8.40	7.00	5.33	7.04	8.18	5.98	5.02	8.19	9.12	7.21	5.51
		100	7.42	8.21	6.66	5.23	7.54	8.60	6.52	5.54	8.39	9.39	7.58	5.45
		120	6.62	7.38	5.90	4.98	6.75	7.82	5.93	4.94	7.83	8.47	7.05	5.41
2	200	80	6.99	8.34	6.03	5.19	8.33	9.32	7.58	5.55	8.41	8.87	7.90	4.80
		100	6.81	7.85	5.99	5.23	8.55	9.20	8.00	5.59	8.79	9.15	8.37	5.24
		120	7.12	7.81	6.42	5.09	8.08	8.69	7.50	5.70	8.19	8.35	7.90	5.18
	500	80	7.14	8.25	6.14	5.22	7.76	9.00	6.48	5.29	7.98	8.60	7.17	4.88
		100	6.76	8.21	5.61	5.46	7.87	9.08	6.86	5.36	8.24	8.97	7.64	5.36
		120	6.50	7.46	5.49	5.01	7.15	8.08	6.43	5.15	8.60	9.15	8.06	5.39
	1000	80	7.28	8.10	6.57	5.33	7.48	8.61	6.36	5.67	8.71	9.77	7.82	5.49
		100	6.53	7.34	5.89	5.24	6.72	8.01	5.86	5.23	8.37	9.09	7.42	5.13
		120	6.27	7.19	5.65	4.71	7.14	8.04	6.28	5.25	8.04	8.79	7.31	5.32
3	200	80	7.57	8.67	6.58	5.10	8.52	9.50	7.78	6.18	8.46	9.10	7.98	4.90
		100	7.16	7.90	6.36	4.82	7.71	8.40	7.11	5.21	8.61	8.92	8.22	5.21
		120	7.14	7.74	6.48	5.20	7.67	8.22	7.14	5.38	8.93	9.30	8.64	5.14
	500	80	7.00	8.09	6.15	5.42	7.79	9.17	6.74	5.41	8.59	9.34	7.91	5.34
		100	6.89	8.10	5.97	5.31	7.80	8.96	6.82	5.88	8.32	8.90	7.76	5.35
		120	6.66	7.89	5.69	5.14	7.82	8.80	6.93	5.55	8.44	9.00	7.93	5.70
	1000	80	7.75	8.66	7.00	5.24	7.42	8.71	6.29	5.78	8.42	9.46	7.46	5.83
		100	7.35	8.13	6.62	5.44	7.16	8.37	6.16	5.62	8.30	9.23	7.72	5.08
		120	6.85	7.53	6.22	4.96	7.08	8.01	6.24	5.27	8.41	9.06	7.74	5.51
	ARE		**41.99**	**61.39**	**24.87**	**5.13**	**53.63**	**73.19**	**36.21**	**8.62**	**69.67**	**82.27**	**57.47**	**6.81**

and generally outperforms the other three tests regardless of whether the simulated data are nearly uncorrelated, moderately correlated, or highly correlated while LR_c, LH_c, and BNP_c are in general rather liberal as most of their empirical sizes are larger than 7%. The empirical powers of the four considered tests are displayed in Figure S.8 (LH_c: black solid curves with triangles, LR_c: red dashed curves with diamonds, BNP_c: green dotted curves with squares, and $T_{n,p}$: blue dot-dashed curves with circles). It is seen that when $\rho = 0.01$ and 0.55, LR_c, LH_c, and BNP_c almost have no

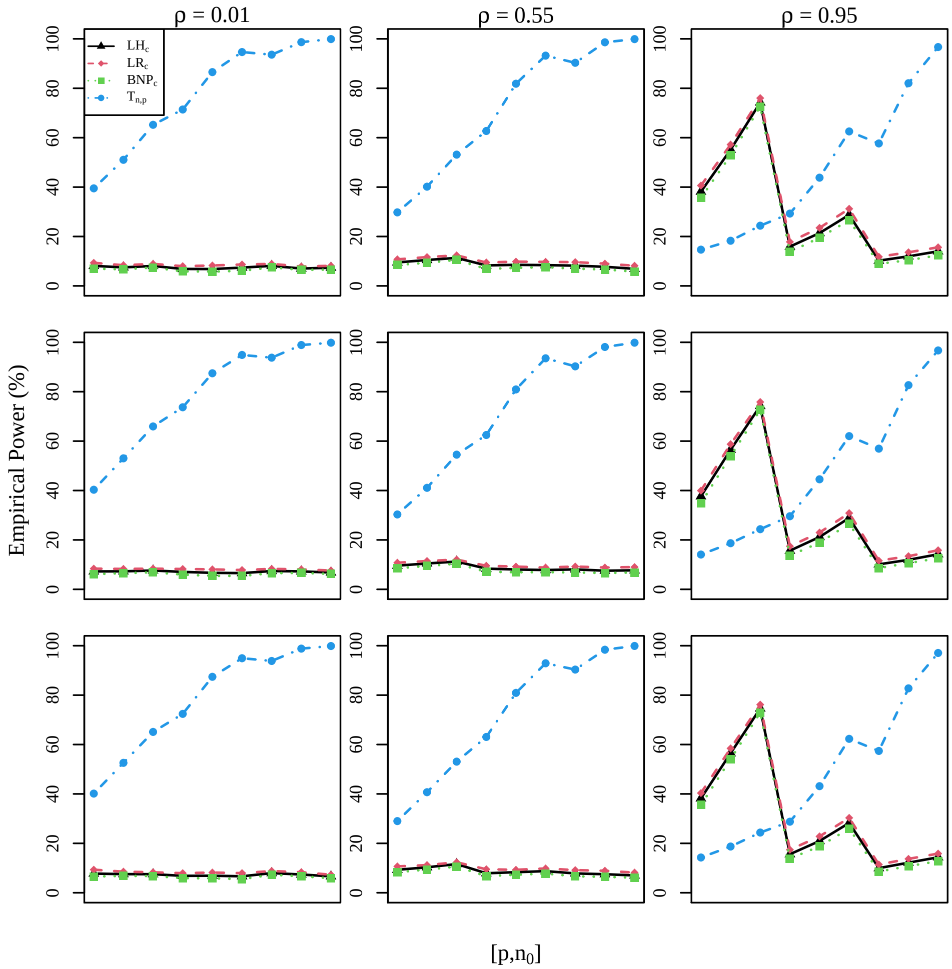

Figure S.8: *Empirical powers (in %) of* LR_c, LH_c, BNP_c *and* $T_{n,p}$ *(*LH_c*: black solid curves with triangles,* LR_c*: red dashed curves with diamonds,* BNP_c*: green dotted curves with squares, and* $T_{n,p}$*: blue dot-dashed curves with circles) associated with parameters* $[p, n_0]$ *from the settings under Model 1 (1st row), Model 2 (2nd row), and Model 3 (3rd row).*

powers under the various configurations. When $\rho = 0.95$, LR_c, LH_c, and BNP_c have higher powers than $T_{n,p}$ only in the first three settings when $p = 200$ and they have much lower powers than $T_{n,p}$ in the remaining settings when $p = 500, 1000$.

S5 Some asymptotical properties of $c_{n,p}$

To address a comment from an anonymous reviewer, we derive some asymptotical properties of the adjustment coefficient $c_{n,p}$ used in the test statistic T_{YS} (1.4) of Yamada and Srivastava (2012). By (1.5), (2.16) and (2.17), some simple algebra yields

$$c_{n,p} = 1 + \frac{(n-k-1)(n-k+2)}{(n-k-2)^2}\frac{q\sqrt{p}}{\hat{d}} + \frac{\sqrt{p}}{n-k},$$

where $\hat{d}$ is the estimated approximate degrees of freedom defined in (2.17). Under Conditions C1–C4 and (2.4), by Theorem 2, as $n, p \to \infty$, we have

$$c_{n,p} = \left(1 + \frac{q\sqrt{p}}{d}\right)[1 + o_p(1)]. \tag{S5.2}$$

According to the discussion presented in Section 2.2, when d is bounded for all p, the normal-reference distribution of $T^*_{n,p,0}$, i.e., the distribution of $T^*_{p,0}$ will not tend to normal and in this case, by (S5.2), we have $c_{n,p} \to \infty$ in probability. That is to say, when $\hat{d}$ is small and $c_{n,p}$ is large, as in the corneal surface data example presented in Section 5 (see Table 4), the underlying null distributions of $T_{n,p}, T_{\text{YS}}$ and T^*_{YS} are unlikely to be normal and hence the normal approximation to the null distributions of $T_{n,p}, T_{\text{YS}}$ and T^*_{YS} is no longer applicable.

Recall that under Conditions C1–C5, by Theorem 1(a), as $n, p \to \infty$, both $\tilde{T}_{n,p,0}$ and $\tilde{T}^*_{p,0}$ will tend to ζ, a non-normal random variable and under Conditions C1–C4, and C6, by Theorem 1(b), as $n, p \to \infty$, both $\tilde{T}_{n,p,0}$ and $\tilde{T}^*_{p,0}$ will tend to $N(0,1)$. Then by (2.2) and Remark 1, under Conditions C1–C5 and (2.4), as $n, p \to \infty$, the limiting null distribution of $T_{n,p}$ after normalization is not normal and under Conditions C1–C4, C6, and (2.4), as $n, p \to \infty$, the limiting null distribution of $T_{n,p}$ after normalization is normal. In these two cases, we can show that $c_{n,p}$ will tend to ∞ and 1 respectively.

In fact, under Condition C5 and by (2.15), as $n, p \to \infty$, we have

$$d = \frac{(n-k)^2 p^2 q}{(n-k-2)^2 \text{tr}(\mathbf{R}^2)} \to q \left(\sum_{r=1}^{\infty} \rho_r \right)^2 < \infty. \tag{S5.3}$$

Then under Conditions C1–C5 and (2.4), we have (S5.2) so that as $n, p \to \infty$, we have $c_{n,p} \to \infty$ in probability. Further, under Condition C6 and by (2.15), as $n, p \to \infty$, we have

$$d = \frac{(n-k)^2 p^2 q}{(n-k-2)^2 \text{tr}(\mathbf{R}^2)} = \frac{qp}{a}[1 + o(1)] = O(p), \tag{S5.4}$$

where a is a finite constant defined in Condition C6. Then under Conditions

C1–C4, C6, and (2.4), we have (S5.2) so that as $n, p \to \infty$, we have $c_{n,p} = \left(1 + \frac{a}{\sqrt{p}}\right)[1 + o_p(1)] \to 1$ in probability.

The above theoretical results, together with the relationship (4.1), i.e., $T_{\text{YS}} = T^*_{\text{YS}}/\sqrt{c_{n,p}}$, clearly explain why when Conditions C1–C4, C6 and (2.4) are satisfied, i.e., when the distributions of T_{YS} and T^*_{YS} are nearly normal, the empirical sizes and powers of T_{YS} are close to those of T^*_{YS} and when Conditions C1–C5 and (2.4) are satisfied, i.e., when the distributions of T_{YS} and T^*_{YS} are unlikely normal, the empirical sizes and powers of T_{YS} are much smaller than those of T^*_{YS}, as seen from Table 1 and Figure 1 presented in Section 4.

From the above analysis, we may loosely suggest that the tests T_{YS} and T^*_{YS} can be used only when the value of $c_{n,p}$ is very close to 1 and they should not be used when the value of $c_{n,p}$ is much larger than 1. However, according to the simulation results presented in Section 4 and in Section S4, the proposed test $T_{n,p}$ can work well regardless of what value the adjustment coefficient $c_{n,p}$ takes.

S6 Technical proofs

S6.1 Proofs of (2.1) and (2.2)

Recall that $\hat{\sigma}_{rr}, r = 1, \ldots, p$ and $\sigma_{rr}, r = 1, \ldots, p$ are the diagonal entries of $\hat{\mathbf{\Sigma}}$ and $\mathbf{\Sigma}$, respectively. Let s_{rr} denote the r-th diagonal entry of $\mathbf{S}_h$ in (1.3). Then we have

$$\begin{aligned} T_{n,p} &= \frac{n-k-2}{(n-k)pq}\mathrm{tr}(\mathbf{S}_h\hat{\mathbf{D}}^{-1}) = \frac{n-k-2}{(n-k)pq}\sum_{r=1}^{p}\hat{\sigma}_{rr}^{-1}s_{rr} \\ &= \frac{n-k-2}{(n-k)pq}\sum_{r=1}^{p}\frac{s_{rr}}{\sigma_{rr}}\left[\left(\frac{\sigma_{rr}}{\hat{\sigma}_{rr}}-1\right)+1\right] \\ &= T_{n,p}^{*} + \frac{n-k-2}{(n-k)pq}\sum_{r=1}^{p}\frac{s_{rr}}{\sigma_{rr}}\left(\frac{\sigma_{rr}}{\hat{\sigma}_{rr}}-1\right). \end{aligned}$$

It follows that

$$\begin{aligned} |T_{n,p} - T_{n,p}^{*}| &= \left|\frac{n-k-2}{(n-k)pq}\sum_{r=1}^{p}\frac{s_{rr}}{\sigma_{rr}}\left(\frac{\sigma_{rr}}{\hat{\sigma}_{rr}}-1\right)\right| \\ &\leq \frac{n-k-2}{(n-k)pq}\sum_{r=1}^{p}\frac{s_{rr}}{\sigma_{rr}}\left|\frac{\sigma_{rr}}{\hat{\sigma}_{rr}}-1\right| \leq \max_{1\leq r\leq p}\left|\frac{\sigma_{rr}}{\hat{\sigma}_{rr}}-1\right| T_{n,p}^{*}. \end{aligned} \tag{S6.5}$$

Next, we aim to study $\max_{1\leq r\leq p}\left|\frac{\sigma_{rr}}{\hat{\sigma}_{rr}}-1\right|$. Under Condition C1, we have $\boldsymbol{\epsilon}_i = (\epsilon_{i1}, \ldots, \epsilon_{ip})^\top = \mathbf{\Gamma}^\top\mathbf{v}_i, i = 1, \ldots, n$ which are i.i.d. p-vectors with $\mathrm{E}(\boldsymbol{\epsilon}_i) = \mathbf{0}$ and $\mathrm{Cov}(\boldsymbol{\epsilon}_i) = \mathbf{\Sigma}$. The usual unbiased estimator of $\mathbf{\Sigma}$ is $\hat{\mathbf{\Sigma}} = (n-k)^{-1}\mathbf{S}_e$, where $\mathbf{S}_e$ is defined in (1.3) and also can be written as $\mathbf{S}_e = \boldsymbol{\epsilon}^\top(\mathbf{I}_n - \mathbf{P}_X)\boldsymbol{\epsilon}$, where $\mathbf{I}_n$ is the usual $n \times n$ identity matrix, and $\mathbf{P}_X = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top$. Let p_{ij} denote the (i,j)-th entry of $\mathbf{P}_X$, then we have $\hat{\sigma}_{rr} = (n-k)^{-1}\left(\sum_{i=1}^{n}\epsilon_{ir}^2 - \sum_{i=1}^{n}\sum_{j=1}^{n}\epsilon_{ir}\epsilon_{jr}p_{ij}\right)$, where ϵ_{ir} is the r-th entry of $\boldsymbol{\epsilon}_i$ with $\mathrm{E}(\epsilon_{ir})=0$ and $\mathrm{Var}(\epsilon_{ir})=\sigma_{rr}$. It follows that

$$\frac{\hat{\sigma}_{rr}}{\sigma_{rr}}=(n-k)^{-1}\left(\sum_{i=1}^{n}\frac{\epsilon_{ir}^2}{\sigma_{rr}}-\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{\epsilon_{ir}\epsilon_{jr}p_{ij}}{\sigma_{rr}}\right).$$

Then we have

$$\sqrt{n}\left(\frac{\hat{\sigma}_{rr}}{\sigma_{rr}}-1\right)=\frac{n}{n-k}\left[I_{r1}-I_{r2}/(\sqrt{n}\sigma_{rr})\right],$$

where

$$\begin{aligned} I_{r1} &= \sqrt{n}\left(n^{-1}\textstyle\sum_{i=1}^{n}\frac{\epsilon_{ir}^2}{\sigma_{rr}}-1\right), \text{ and}\\ I_{r2} &= \left(\textstyle\sum_{i=1}^{n}\sum_{j=1}^{n}\epsilon_{ir}\epsilon_{jr}p_{ij}\right)-k\sigma_{rr}. \end{aligned} \tag{S6.6}$$

It is easy to note that $\mathrm{E}(\sigma_{rr}^{-1}\epsilon_{ir}^2)=1$. Let $\gamma_{j\ell}, j,\ell=1,\dots,p$ denote the (j,ℓ)-th entry of $\boldsymbol{\Gamma}$. Then under Condition C1, we have $\epsilon_{ir}=\sum_{j=1}^{p}\gamma_{jr}v_{ij}$ and $\sigma_{rr}=\sum_{j=1}^{p}\gamma_{jr}^2$. Furthermore, under Conditions C1–C3, we have

$$\begin{aligned} \mathrm{E}\left(\frac{\epsilon_{ir}^4}{\sigma_{rr}^2}\right) &= \sigma_{rr}^{-2}\mathrm{E}\left(\sum_{j=1}^{p}\gamma_{jr}^4 v_{ij}^4\right)+3\sigma_{rr}^{-2}\mathrm{E}\left(\sum_{j\neq s}\gamma_{jr}^2\gamma_{\ell r}^2 v_{ij}^2 v_{i\ell}^2\right)\\ &= (3+\Delta)\sigma_{rr}^{-2}\sum_{j=1}^{p}\gamma_{jr}^4+3\sigma_{rr}^{-2}\sum_{j\neq\ell}\gamma_{jr}^2\gamma_{\ell r}^2=\Delta\sigma_{rr}^{-2}\sum_{j=1}^{p}\gamma_{rj}^4+3\\ &\leq 3+\Delta, \end{aligned}$$

where we use the fact that $\sigma_{rr}^{-2}\sum_{j=1}^{p}\gamma_{rj}^4\leq\sigma_{rr}^{-2}(\sum_{j=1}^{p}\gamma_{rj}^2)^2=\sigma_{rr}^{-2}\sigma_{rr}^2=1$. Therefore, together with Condition C4, for all $r=1,\dots,p$, we have

$$\zeta_{rr}=\mathrm{Var}(\sigma_{rr}^{-1}\epsilon_{ir}^2)\leq 2+\Delta<\infty, \tag{S6.7}$$

uniformly for all r. By the central limit theorem, for any $r \in \{1, \dots, p\}$, as $n \to \infty$, we have

$$I_{r1} = \sqrt{n}\left(n^{-1}\sum_{i=1}^{n}\frac{\epsilon_{ir}^2}{\sigma_{rr}} - 1\right) \xrightarrow{L} \mathrm{N}(0, \zeta_{rr}).$$

We now consider I_{r2}. We first have $\mathrm{E}(I_{r2}) = 0$. Since $\epsilon_{ir}, i = 1, \dots, n$ are i.i.d., $\mathrm{E}(\epsilon_{ir}) = 0$, and $\mathrm{Var}(\epsilon_{ir}) = \sigma_{rr}$, under Conditions C1–C3, we have

$$\begin{aligned}
\mathrm{Var}(I_{r2}) &= \mathrm{Var}\left(\sum_{i=1}^{n} p_{ii}(\epsilon_{ir}^2 - \sigma_{rr}) + 2\sum_{j<\ell} p_{j\ell}\epsilon_{jr}\epsilon_{\ell r}\right) \\
&= \sum_{i=1}^{n} p_{ii}^2 \mathrm{Var}(\epsilon_{ir}^2 - \sigma_{rr}) + 4\sum_{j<\ell} p_{j\ell}^2\sigma_{rr}^2 \\
&\le \sum_{i=1}^{n} p_{ii}^2\sigma_{rr}^2(2+\Delta) + 4\sum_{j<\ell} p_{j\ell}^2\sigma_{rr}^2,
\end{aligned}$$

where we use the fact that by (S6.7), we have $\mathrm{Var}(\epsilon_{ir}^2 - \sigma_{rr}) = \sigma_{rr}^2\mathrm{Var}(\sigma_{rr}^{-1}\epsilon_{ir}^2) \le \sigma_{rr}^2(2+\Delta)$. It follows that

$$\begin{aligned}
\mathrm{Var}(I_{r2}) &\le \Delta\sigma_{rr}^2(\sum_{i=1}^{n} p_{ii}^2) + 2\sigma_{rr}^2(\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}^2) \\
&\le (\Delta+2)\sigma_{rr}^2(\sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}^2) \\
&= (\Delta+2)k\sigma_{rr}^2,
\end{aligned}$$

where we use the fact that $\sum_{i=1}^{n} p_{ii}^2 \le \sum_{i=1}^{n}\sum_{j=1}^{n} p_{ij}^2 = \mathrm{tr}(\mathbf{P}_X^2) = \mathrm{tr}(\mathbf{P}_X) = k$. It follows that as $n \to \infty$, we have

$$\mathrm{Var}\left[I_{r2}/(\sqrt{n}\sigma_{rr})\right] = \mathrm{Var}(I_{r2})/(n\sigma_{rr}^2) \le (\Delta+2)k/n \to 0,$$

uniformly for all $r = 1, \ldots, p$. Therefore, as $n \to \infty$, we have $I_{r2}/(\sqrt{n}\sigma_{rr})$ converges to 0 in probability uniformly. Thus, for any $r \in \{1, \ldots, p\}$, as $n \to \infty$, under Conditions C1–C4, we have

$$\sqrt{n}\left(\frac{\hat{\sigma}_{rr}}{\sigma_{rr}} - 1\right) \xrightarrow{L} \mathrm{N}(0, \zeta_{rr}).$$

Set $g(x) = 1/x, x > 0$. We have $g'(x) = -1/x^2$. Set $\lambda_{rr} = \mathrm{E}\left(\frac{\hat{\sigma}_{rr}}{\sigma_{rr}}\right) = 1$. It follows that for any $r \in \{1, \ldots, p\}$, as $n \to \infty$, we have

$$\sqrt{n}\left\{g\left(\frac{\hat{\sigma}_{rr}}{\sigma_{rr}}\right) - g\left(\lambda_{rr}\right)\right\} \xrightarrow{L} \mathrm{N}\left\{0, \left[g'(\lambda_{rr})\right]^2 \zeta_{rr}\right\}.$$

Since $g(\lambda_{rr}) = 1$ and $[g'(\lambda_{rr})]^2 = \left(-1/\lambda_{rr}^2\right)^2 = 1$, for any $r \in \{1, \ldots, p\}$, we have

$$\sqrt{n}\left(\frac{\sigma_{rr}}{\hat{\sigma}_{rr}} - 1\right) \xrightarrow{L} \mathrm{N}(0, \zeta_{rr}).$$

Then we have

$$\begin{aligned}
&\Pr\left[\max_{1\le r\le p}\left|\frac{\hat{\sigma}_{rr}}{\sigma_{rr}} - 1\right| > \sqrt{2(2+\Delta)\log(p)/n}\right] \\
&\le 2\sum_{r=1}^{p} \Pr\left[\sqrt{n}\left(\frac{\hat{\sigma}_{rr}}{\sigma_{rr}} - 1\right) > \sqrt{2(2+\Delta)\log(p)}\right] \\
&= 2\sum_{r=1}^{p}\left\{1 - \Phi\left[\sqrt{2(2+\Delta)\log(p)/\zeta_{rr}}\right]\right\}[1+o(1)] \\
&\le 2p\left\{1 - \Phi\left[\sqrt{2\log(p)}\right]\right\}[1+o(1)] \\
&\le \pi^{-1/2}\log(p)^{-1/2}[1+o(1)] \to 0,
\end{aligned}$$

as $p \to \infty$. As $n, p \to \infty$, we then have $\max_{1\le r\le p}\left|\frac{\sigma_{rr}}{\hat{\sigma}_{rr}} - 1\right| = O_p\left[n^{-1}\log(p)\right]$, resulting in (2.1). The expression (2.2) follows from (S6.5) immediately.

The proof is complete.

S6.2 Proof of Theorem 1

We first prove the first expression of Part (a). Write $T^*_{n,p,0} = (pq)^{-1}\mathrm{tr}(\mathbf{Z}^\top\mathbf{H}\mathbf{Z})$, where $\mathbf{Z} = \boldsymbol{\epsilon}\mathbf{D}^{-1/2} = (\mathbf{z}_1, \ldots, \mathbf{z}_n)^\top$ with $\mathbf{z}_1, \ldots, \mathbf{z}_n$ being i.i.d. with $\mathrm{E}(\mathbf{z}_1) = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{z}_1) = \mathbf{R}$. Let $\mathbf{u}_{p,1}, \ldots, \mathbf{u}_{p,p}$ denote the orthonormal eigenvectors associated with the eigenvalues $\lambda_{p,1}, \ldots, \lambda_{p,p}$ of $\mathbf{R}$ in descending order. We can write $\mathbf{z}_i = \sum_{r=1}^p \xi_{ir}\mathbf{u}_{p,r}, i = 1, \ldots, n$. It is known that $\xi_{ir}, r = 1, \ldots, p$ are uncorrelated with $\mathrm{E}(\xi_{ir}) = 0$ and $\mathrm{Var}(\xi_{ir}) = \lambda_{p,r}$. It follows that $\mathbf{z}_i^\top\mathbf{z}_i = \sum_{r=1}^p \xi_{ir}^2$ and $\mathbf{z}_i^\top\mathbf{z}_j = \sum_{r=1}^p \xi_{ir}\xi_{jr}$. Thus, $T^*_{n,p,0} = (pq)^{-1}\sum_{r=1}^p B_{n,r}$, where $B_{n,r} = \boldsymbol{\xi}_r^\top\mathbf{H}\,\boldsymbol{\xi}_r$ with $\boldsymbol{\xi}_r = (\xi_{1r}, \ldots, \xi_{nr})^\top,\ r = 1, \ldots, p$. Set

$$\mathbf{W}_n = [\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{C}^\top]^{-1/2}\mathbf{C}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top, \qquad \text{(S6.8)}$$

a $q \times n$ matrix of rank q. We have $\mathbf{H} = \mathbf{W}_n^\top\mathbf{W}_n$ and $\mathbf{W}_n\mathbf{W}_n^\top = \mathbf{I}_q$. In addition, we have $\mathrm{tr}(\mathbf{H}) = q$ and $\mathrm{tr}(\mathbf{H}^2) = q$.

Let h_{ij} denote the (i,j)-th entry of $\mathbf{H}$. We have $B_{n,r} = \sum_{i=1}^n h_{ii}\xi_{ir}^2 + 2\sum_{1\le i<j\le n} h_{ij}\xi_{ir}\xi_{jr}, r = 1, \ldots, p$. It follows that $\mathrm{E}(B_{n,r}) = \sum_{i=1}^n h_{ii}\lambda_{p,r} = q\lambda_{p,r}$ and $\mathrm{Var}(B_{n,r}) = \sum_{i=1}^n h_{ii}^2\mathrm{Var}(\xi_{ir}^2) + 4\sum_{1\le i<j\le n} h_{ij}^2\lambda_{p,r}^2$. Under Condition C1, we have $\xi_{ir}^2 = (\mathbf{z}_i^\top\mathbf{u}_{p,r})^2 = \mathbf{v}_i^\top\mathbf{S}\mathbf{v}_i$ where $\mathrm{E}(\mathbf{v}_i) = \mathbf{0}$ and $\mathrm{Cov}(\mathbf{v}_i) = \mathbf{I}_p$ and $\mathbf{S} = \mathbf{R}^{1/2}\mathbf{u}_{p,r}\mathbf{u}_{p,r}^\top\mathbf{R}^{1/2}$. Then we have $\mathrm{tr}(\mathbf{S}^2) = (\mathbf{u}_{p,r}^\top\mathbf{R}\mathbf{u}_{p,r})^2 = \lambda_{p,r}^2$. Under Conditions C1, C2, and C4, applying Lemma 6.1 (b) of Srivastava and

Kubokawa (2013), we have $\mathrm{Var}(\xi_{ir}^2) = \mathrm{Var}(\mathbf{v}_i^\top \mathbf{S}\mathbf{v}_i) = \Delta \sum_{r=1}^p s_{rr}^2 + 2\mathrm{tr}(\mathbf{S}^2) \leq (\Delta+2)\mathrm{tr}(\mathbf{S}^2) = (\Delta+2)\lambda_{p,r}^2$, where s_{rr}'s denote the diagonal entries of $\mathbf{S}$. It follows that

$$\mathrm{Var}(B_{n,r}) = \sum_{i=1}^n h_{ii}^2 \mathrm{Var}(\xi_{ir}^2) + 4 \sum_{1\leq i<j\leq n} h_{ij}^2 \lambda_{p,r}^2 \leq (2q + \Delta \sum_{i=1}^n h_{ii}^2)\lambda_{p,r}^2, \quad \text{(S6.9)}$$

where we use the fact that $\sum_{i=1}^n \sum_{j=1}^n h_{ij}^2 = \mathrm{tr}(\mathbf{H}^2) = q$.

Write $\tilde{T}_{n,p,0}^* = [pqT_{n,p,0}^* - q\mathrm{tr}(\mathbf{R})]/[2q\mathrm{tr}(\mathbf{R}^2)]^{1/2} = [\sum_{r=1}^p (B_{n,r} - q\lambda_{p,r})]/[2q\mathrm{tr}(\mathbf{R}^2)]^{1/2}$. Set $\tilde{T}_{n,p,0}^{*(m)} = [\sum_{r=1}^m (B_{n,r} - q\lambda_{p,r})]/[2q\mathrm{tr}(\mathbf{R}^2)]^{1/2}$. By (S6.9), we have

$$\begin{aligned}
& \mathrm{E}\left(\tilde{T}_{n,p,0}^* - \tilde{T}_{n,p,0}^{*(m)}\right)^2 = \mathrm{E}\left\{\textstyle\sum_{r=m+1}^p (B_{n,r} - q\lambda_{p,r})/[2q\mathrm{tr}(\mathbf{R}^2)]^{1/2}\right\}^2 \\
= \;& \mathrm{Var}\left(\textstyle\sum_{r=m+1}^q B_{n,r}\right)/[2q\mathrm{tr}(\mathbf{R}^2)] \leq \left[\textstyle\sum_{r=m+1}^q \sqrt{\mathrm{Var}(B_{n,r})}\right]^2/[2q\mathrm{tr}(\mathbf{R}^2)] \\
\leq \;& (2q + \Delta \textstyle\sum_{i=1}^n h_{ii}^2)\left(\textstyle\sum_{r=m+1}^q \lambda_{p,r}\right)^2/[2q\mathrm{tr}(\mathbf{R}^2)] \\
= \;& \left(1 + \tfrac{\Delta}{2q}\textstyle\sum_{i=1}^n h_{ii}^2\right)\left(\textstyle\sum_{r=m+1}^q \rho_{p,r}\right)^2.
\end{aligned}$$

It follows that

$$\begin{aligned}
|\psi_{\tilde{T}_{n,p,0}^*}(t) - \psi_{\tilde{T}_{n,p,0}^{*(m)}}(t)| &\leq |t|\left[\mathrm{E}(\tilde{T}_{n,p,0}^* - \tilde{T}_{n,p,0}^{*(m)})^2\right]^{1/2} \\
&\leq |t|\left(1 + \tfrac{\Delta}{2q}\textstyle\sum_{i=1}^n h_{ii}^2\right)^{1/2} \textstyle\sum_{r=m+1}^p \rho_{p,r},
\end{aligned}$$

where $\psi_{\tilde{T}_{n,p,0}^*}(t)$ and $\psi_{\tilde{T}_{n,p,0}^{*(m)}}(t)$ are the characteristic functions of $\tilde{T}_{n,p,0}^*$ and $\tilde{T}_{n,p,0}^{*(m)}$, respectively.

Let t be fixed. By Condition C5, for any fixed m, as $p\to\infty$, we have $\sum_{r=1}^\infty \rho_r < \infty$ and

$$\sum_{r=m+1}^p \rho_{p,r} = \sum_{r=1}^p \rho_{p,r} - \sum_{r=1}^m \rho_{p,r} \to \sum_{r=1}^\infty \rho_r - \sum_{r=1}^m \rho_r \to \sum_{r=m+1}^\infty \rho_r.$$

By letting $m \to \infty$, we further have $\sum_{r=m+1}^{\infty} \rho_r \to 0$. Notice that under Condition C3, we have $\sum_{i=1}^{n} h_{ii}^2 = O(n^{-1})$. Thus, for any given $\epsilon > 0$, there exist P_1, M_1 and N_1, depending on t and ϵ, such that for any $p \geq P_1$, $m \geq M_1$ and $n \geq N_1$, we have

$$\left|\psi_{\tilde{T}^*_{n,p,0}}(t) - \psi_{\tilde{T}^{*(m)}_{n,p,0}}(t)\right| \leq \epsilon. \tag{S6.10}$$

Set $\boldsymbol{\zeta}_{n,r} = \mathbf{W}_n\, \boldsymbol{\xi}_r, r = 1, \ldots, p$ where $\mathbf{W}_n$ is defined in (S6.8). Then $B_{n,r} = \boldsymbol{\xi}_r^\top \mathbf{H}\, \boldsymbol{\xi}_r = \|\, \boldsymbol{\zeta}_{n,r}\|^2, r = 1, \ldots, p$. For any fixed finite r, by the central limit theorem, as $n \to \infty$, we have $\boldsymbol{\zeta}_{n,r} \xrightarrow{L} \boldsymbol{\zeta}_r \sim \mathrm{N}_q(\mathbf{0}, \lambda_{p,r}\mathbf{I}_q)$ and hence $B_{n,r} \stackrel{d}{=} \lambda_{p,r} A_r,\ A_r \sim \chi_q^2$. Therefore, for any fixed $p \geq P_1, m \geq M_1$, as $n \to \infty$, we have $\tilde{T}^{*(m)}_{n,p,0} \xrightarrow{L} \tilde{T}^{(m)}_{p,0}$ where $\tilde{T}^{(m)}_{p,0} = \sum_{r=1}^{m} \rho_{p,r}(A_r - q)/\sqrt{2q}$. That is, under Condition C3, there exists N_2, depending on p, m, t and ϵ, such that for any $n \geq N_2$ we have

$$\left|\psi_{\tilde{T}^{*(m)}_{n,p,0}}(t) - \psi_{\tilde{T}^{(m)}_{p,0}}(t)\right| \leq \epsilon. \tag{S6.11}$$

Recall that $\zeta = \sum_{r=1}^{\infty} \rho_r (A_r - q)/\sqrt{2q}$. Set $\zeta^{(m)} = \sum_{r=1}^{m} \rho_r (A_r - q)/\sqrt{2q}$. Then, under Condition C5, for any fixed m, as $p \to \infty$, we have $\tilde{T}^{(m)}_{p,0} \xrightarrow{L} \zeta^{(m)}$. That is, there exists a P_2, depending on m, t and ϵ, such that for any $p \geq P_2$ we have

$$\left|\psi_{\tilde{T}^{(m)}_{p,0}}(t) - \psi_{\zeta^{(m)}}(t)\right| \leq \epsilon. \tag{S6.12}$$

Furthermore, we have

$$\begin{aligned}\left|\psi_{\zeta^{(m)}}(t)-\psi_\zeta(t)\right| &\le |t|\left\{\mathrm{E}\left[\sum_{r=m+1}^{\infty}\rho_r(A_r-q)/\sqrt{2q}\right]^2\right\}^{1/2}\\ &\le |t|\left\{\mathrm{Var}\left[\sum_{r=m+1}^{\infty}\rho_r(A_r-q)/\sqrt{2q}\right]\right\}^{1/2}\\ &= |t|\left(\sum_{r=m+1}^{\infty}\rho_r^2\right)^{1/2}\le |t|\left(\sum_{r=m+1}^{\infty}\rho_r\right),\end{aligned}$$

which, under Condition C5, tends to 0 as $m\to\infty$. Thus, there exists M_2, depending on t and ϵ, such that for any $m\ge M_2$ we have

$$\left|\psi_{\zeta^{(m)}}(t)-\psi_\zeta(t)\right|\le\epsilon. \tag{S6.13}$$

It follows from (S6.10)–(S6.13) that for any $n\ge\max(N_1,N_2)$, $p\ge\max(P_1,P_2)$ and $m\ge\max(M_1,M_2)$ we have

$$\begin{aligned}\left|\psi_{\tilde{T}^*_{n,p,0}}(t)-\psi_\zeta(t)\right| &\le \left|\psi_{\tilde{T}^*_{n,p,0}}(t)-\psi_{\tilde{T}^{*(m)}_{n,p,0}}(t)\right|+\left|\psi_{\tilde{T}^{*(m)}_{n,p,0}}(t)-\psi_{\tilde{T}^{(m)}_{p,0}}(t)\right|\\ &+ \left|\psi_{\tilde{T}^{(m)}_{p,0}}(t)-\psi_{\zeta^{(m)}}(t)\right|+\left|\psi_{\zeta^{(m)}}(t)-\psi_\zeta(t)\right|\le 4\epsilon.\end{aligned}$$

The convergence in distribution of $\tilde{T}^*_{n,p,0}$ to ζ given in (2.11) follows as we can let $\epsilon\to 0$. The first expression of Theorem 1(a) is then proved.

When the measurement error matrix $\boldsymbol{\epsilon}$ are normally distributed, Conditions C1 and C2 are automatically satisfied and we have $T^*_{n,p,0}\overset{d}{=}T^*_{p,0}$ so that under Conditions C3, C4, and C5, the second expression of (2.11) follows immediately.

We now prove (b). Under Conditions C1–C4, and C6, the first expression of (2.12) follows from Theorem 2.1 of Srivastava and Kubokawa (2013). Again, when the measurement error matrix $\boldsymbol{\epsilon}$ are normally distributed, Conditions C1 and C2 are automatically satisfied and we have $T^*_{n,p,0} \stackrel{d}{=} T^*_{p,0}$ so that the second expression of (2.12) follows immediately too.

Set $\tilde{x} = (x-1)/\left[2p^{-2}q^{-1}\mathrm{tr}(\mathbf{R}^2)\right]^{1/2}$ for any real number x. Since the limit ζ is a continuous random variable, by Lemma 2.11 of Vaart (1998), the uniform convergence result given in (2.13) follows directly from the convergence in distribution of both $\tilde{T}^*_{n,p,0}$ and $\tilde{T}^*_{p,0}$ to ζ and the triangular inequality:

$$\sup_x \left| \Pr(T^*_{n,p,0} \le x) - \Pr(T^*_{p,0} \le x) \right| = \sup_x \left| \Pr(\tilde{T}^*_{n,p,0} \le \tilde{x}) - \Pr(\tilde{T}^*_{p,0} \le \tilde{x}) \right|$$
$$\le \sup_x \left| \Pr(\tilde{T}^*_{n,p,0} \le \tilde{x}) - \Pr(\zeta \le \tilde{x}) \right| + \sup_x \left| \Pr(\tilde{T}^*_{p,0} \le \tilde{x}) - \Pr(\zeta \le \tilde{x}) \right| \to 0,$$

as $n, p \to \infty$. The theorem is then proved.

S6.3 Proof of Theorem 2

Notice that we can treat $\mathbf{R}$ as the covariance matrix of "the transformed data" $\boldsymbol{\epsilon}^*_i = \mathbf{D}^{-1/2}\boldsymbol{\epsilon}_i, i = 1, \ldots, n$. Under Conditions C1–C3, by Zhang et al. (2020), the ratio-consistent estimator of $\mathrm{tr}(\mathbf{R}^2)$ is given by

$$\frac{(n-k)^2}{(n-k-1)(n-k+2)}\left[\mathrm{tr}(\tilde{\mathbf{R}}^2) - \frac{1}{n-k}\mathrm{tr}^2(\tilde{\mathbf{R}})\right],$$

where $\tilde{\mathbf{R}} = \mathbf{D}^{-1/2}\hat{\mathbf{\Sigma}}\mathbf{D}^{-1/2}$. Notice that under Conditions C1–C4, by (2.1) and (2.3), we can write $\hat{\mathbf{D}}^{-1} = \mathbf{D}^{-1}[1 + o_p(1)]$. It follows that $\mathrm{tr}(\hat{\mathbf{R}}) = \mathrm{tr}(\tilde{\mathbf{R}})[1 + o_p(1)]$ and $\mathrm{tr}(\hat{\mathbf{R}}^2) = \mathrm{tr}(\tilde{\mathbf{R}}^2)[1 + o_p(1)]$. Therefore,

$$\widehat{\mathrm{tr}(\mathbf{R}^2)} = \frac{(n-k)^2}{(n-k-1)(n-k+2)}\left[\mathrm{tr}(\hat{\mathbf{R}}^2) - \frac{1}{n-k}\mathrm{tr}^2(\hat{\mathbf{R}})\right],$$

is ratio-consistent for $\mathrm{tr}(\mathbf{R}^2)$ as desired. That is, as $n, p \to \infty$, we have $\widehat{\mathrm{tr}(\mathbf{R}^2)}/\mathrm{tr}(\mathbf{R}^2) \to 1$ in probability. It follows that as $n, p \to \infty$, we have $\hat{d}/d \to 1$ in probability. The theorem is proved.

S6.4 Proof of Theorem 3

First of all, under Conditions C1–C4 and (2.3), by Theorem 2, $\hat{d}$ (2.17) is a ratio-consistent estimator of d (2.15). We now prove (a). Under (2.3), we have $T_{n,p} = T^*_{n,p}[1 + o_p(1)]$. This, together with (2.5) and (2.18), we have $T_{n,p} = [T^*_{n,p,0} + (pq)^{-1}\mathrm{tr}(\ \mathbf{\Omega}\mathbf{D}^{-1})][1+o_p(1)]$. Under Conditions C1–C5, Theorem 1(a) indicates that as $n, p \to \infty$, we have $(T^*_{n,p,0}-1)/[2p^{-2}q^{-1}\mathrm{tr}(\mathbf{R}^2)]^{1/2} \xrightarrow{L}$

ζ. We then have

$$\begin{aligned}
& \Pr\left[T_{n,p} \geq \chi^2_{\hat{d}}(\alpha)/\hat{d}\right] \\
= & \Pr\left[T^*_{n,p,0} \geq \chi^2_{\hat{d}}(\alpha)/\hat{d} - (pq)^{-1}\mathrm{tr}(\ \mathbf{\Omega}\mathbf{D}^{-1})\right][1+o(1)] \\
= & \Pr\left[\frac{T^*_{n,p,0}-1}{\sqrt{2p^{-2}q^{-1}\mathrm{tr}(\mathbf{R}^2)}} \geq \frac{\chi^2_{\hat{d}}(\alpha)/\hat{d}-1}{\sqrt{2p^{-2}q^{-1}\mathrm{tr}(\mathbf{R}^2)}} - \frac{\mathrm{tr}(\ \mathbf{\Omega}\mathbf{D}^{-1})}{\sqrt{2q\mathrm{tr}(\mathbf{R}^2)}}\right][1+o(1)] \\
= & \Pr\left[\zeta \geq \frac{\chi^2_{d}(\alpha)/d-1}{\sqrt{2p^{-2}q^{-1}\mathrm{tr}(\mathbf{R}^2)}} - \frac{\mathrm{tr}(\ \mathbf{\Omega}\mathbf{D}^{-1})}{\sqrt{2q\mathrm{tr}(\mathbf{R}^2)}}\right][1+o(1)] \\
= & \Pr\left[\zeta \geq \frac{\chi^2_{d}(\alpha)-d}{\sqrt{2d}} - \frac{\mathrm{tr}(\ \mathbf{\Omega}\mathbf{D}^{-1})}{\sqrt{2q\mathrm{tr}(\mathbf{R}^2)}}\right][1+o(1)].
\end{aligned}$$

Part (a) is proved.

Next we prove Part (b). Under Conditions C1–C4 and C6, by Theorem 1(b), we have $(T^*_{n,p,0} - 1)/[2p^{-2}q^{-1}\mathrm{tr}(\mathbf{R}^2)]^{1/2} \xrightarrow{L} \mathrm{N}(0,1)$ and $(T^*_{p,0} - 1)/[2p^{-2}q^{-1}\mathrm{tr}(\mathbf{R}^2)]^{1/2} \xrightarrow{L} \mathrm{N}(0,1)$ as $n, p \to \infty$. Since $T^*_{p,0}$ is a chi-square type mixture, by Theorems 4 and 5 of Zhang et al. (2020), we have $d \to \infty$ as $p \to \infty$. It follows that $(\chi^2_d - d)/\sqrt{2d} \xrightarrow{L} \mathrm{N}(0,1)$ and hence $[\chi^2_d(\alpha) - d]/\sqrt{2d} \to z_\alpha$ where z_α denotes the upper 100% percentile of $\mathrm{N}(0,1)$. It follows that under the given conditions, as $n, p \to \infty$, we have

$$\begin{aligned}
& \Pr\left[T_{n,p} \geq \chi^2_{\hat{d}}(\alpha)/\hat{d}\right] \\
= & \Pr\left[T^*_{n,p,0} \geq \chi^2_{\hat{d}}(\alpha)/\hat{d} - (pq)^{-1}\mathrm{tr}(\ \mathbf{\Omega}\mathbf{D}^{-1})\right][1+o(1)] \\
= & \Pr\left[\frac{T^*_{n,p,0}-1}{\sqrt{2p^{-2}q^{-1}\mathrm{tr}(\mathbf{R}^2)}} \geq \frac{\chi^2_{d}(\alpha)-d}{\sqrt{2d}} - \frac{\mathrm{tr}(\ \mathbf{\Omega}\mathbf{D}^{-1})}{\sqrt{2q\mathrm{tr}(\mathbf{R}^2)}}\right][1+o(1)] \\
= & \Phi\left[-z_\alpha + \frac{\mathrm{tr}(\ \mathbf{\Omega}\mathbf{D}^{-1})}{\sqrt{2q\mathrm{tr}(\mathbf{R}^2)}}\right][1+o(1)],
\end{aligned}$$

where $\Phi(\cdot)$ denotes the cumulative distribution of $\mathrm{N}(0,1)$. The proof is complete.

References

Fujikoshi, Y., Himeno, T., and Wakaki, H. (2004). Asymptotic results of a high dimensional MANOVA test and power comparison when the dimension is large compared to the sample size. *J. of the Japan Statistical Society*, 34(1):19–26.

Li, H., Aue, A., and Paul, D. (2020). High-dimensional general linear hypothesis tests via non-linear spectral shrinkage. *Bernoulli*, 26(4):2541–2571.

Srivastava, M. S. and Fujikoshi, Y. (2006). Multivariate analysis of variance with fewer observations than the dimension. *J. of Multivariate Analysis*, 97(9):1927–1940.

Srivastava, M. S. and Kubokawa, T. (2013). Tests for multivariate analysis of variance in high dimension under non-normality. *J. of Multivariate Analysis*, 115:204–216.

Vaart, A. W. v. d. (1998). *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press.

Yamada, T. and Srivastava, M. S. (2012). A test for multivariate analysis of variance in high dimension. *Communications in Statistics-Theory and Methods*, 41(13-14):2602–2615.

Zhang, J.-T., Guo, J., Zhou, B., and Cheng, M.-Y. (2020). A simple two-sample test in high dimensions based on L^2-norm. *J. of American Statistical Association*, 115(530):1011–1027.