Supplemental Material for "Multivariate-sign-based high-dimensional tests for the two-sample location problem"

This supplemental file contains some more discussions on the proposed SS test, all the technical proofs and some additional simulation results. The following materials are included:

- Appendix B: Additional Simulation
- Appendix C: Discussion about non-negligible biases
- Appendix D: Proofs of Lemmas and Proposition 2
- Appendix E: Some properties of r^{-1}
- Appendix F: Discussion on the test using sparse covariance matrix estimates

Appendix B: Additional Simulation

First, we add some simulation results when $p < n$. Table A1 reports the empirical sizes of SS and TS tests with different sample sizes and dimensions. The sample size n_i is chosen as $20, 50$ and 100 and the dimension $p = rn_i$, where $r = 0.2, 0.4, 0.8$. Table A3 reports the empirical size-corrected power of these two methods with the similar settings to those in Table 1 in the paper. The results are similar to those in Table 1; Our SS test still performs better than the TS test in most cases.

Table A1: Empirical size comparison at 5% significance under Scenarios (I)-(V) when $p < n$

Scenario	r	n_i = 25		50		75		100	
		SS	TS	SS	TS	SS	TS	SS	TS
(I)	0.2	6.5	5.7	6.5	3.9	5.3	4.5	7.2	5.1
	0.4	6.5	2.6	4.6	3.9	5.7	2.9	5.4	2.3
	0.8	6.0	2.5	6.5	0.8	4.8	1.7	6.8	1.3
(II)	0.2	7.0	5.3	7.0	4.0	5.8	4.5	6.9	4.5
	0.4	6.6	2.6	5.2	2.7	5.9	3.8	5.8	2.6
	0.8	6.0	0.9	6.4	1.5	4.6	1.8	6.5	1.1
(III)	0.2	6.0	5.0	6.5	4.7	5.7	4.3	7.4	5.0
	0.4	5.8	4.7	4.5	3.6	5.3	2.7	5.4	3.4
	0.8	5.7	1.7	6.3	1.1	4.8	1.1	6.4	1.0
(IV)	0.2	5.8	5.6	4.9	4.8	6.0	5.4	4.9	3.8
	0.4	5.2	4.6	5.8	4.3	5.6	4.2	5.8	4.3
	0.8	4.7	1.2	4.5	2.6	5.5	1.4	6.0	1.6
(V)	0.2	6.2	5.6	6.4	4.0	5.7	3.9	6.9	4.9
	0.4	5.9	3.8	4.4	3.3	5.6	4.0	5.5	4.1
	0.8	5.8	0.8	6.4	1.7	4.9	1.6	6.5	1.9

Next, we consider the case $p > n$. Here we also include the test proposed by Feng et al.

Table A2: Empirical size-corrected power comparison at 5% significance under Scenarios (I)-(V) when $p < n$

		Equal Allocation				Linear Allocation			
		$p = 40$		$p = 60$		$p = 40$		$p = 60$	
Scenario	%	SS	TS	SS	TS	SS	TS	SS	TS
(I)	50%	43	19	67	25	41	18	67	23
	95%	67	40	84	46	44	35	72	41
(II)	50%	81	34	98	49	82	36	98	49
	95%	99	74	100	85	91	78	100	88
(III)	50%	68	26	92	44	68	28	91	35
	95%	73	62	96	76	77	63	97	81
(IV)	50%	92	56	94	76	99	49	98	75
	95%	100	77	100	89	100	75	100	86
(V)	50%	75	35	95	48	75	37	94	48
	95%	79	70	98	82	83	72	98	83

(2014) (BF). Table A3 reports the empirical sizes at a 5% nominal significance level with $(n_i, p) = (25, 120), (50, 480), (50, 1440)$. The empirical sizes of SKK, BF and GCT are a little conservative under the Scenario (III)-(V). It is not strange because all these methods are based on the diverging factor model or a strict moment condition. Tables A4, A5 and A6 show the empirical power of all these six methods with the same settings as those in Table 3 in the main paper. Now, the BF test performs similar to SKK in most cases and is less powerful than our SS test under non-normal cases.

Next, we show some results of the empirical sizes and power of these six methods under the moving average models (VI)-(IX). Because the empirical sizes of GCT largely deviate from the nominal level, we also tabulate the size-corrected power of GCT (GCT*; the last column). GCT* is not as powerful as the other scalar-invariant tests under these scenarios.

The comparison conclusion drawn in the paper still holds here. The SS test can well control the empirical sizes in most cases, though it is a little conservative when p/n_i is large. Our SS test outperforms all the other tests under the non-normal cases. Under the non-normal distributions, the advantage of our SS test is obvious.

Table A3: Empirical size at 5% significance under Scenarios (I)-(V)

(n_i, p)	Scenario	Size				
		SS	SKK	CQ	BF	GCT
(25,120)	(I)	4.0	9.1	5.6	6.5	3.6
	(II)	4.2	9.1	5.5	6.5	2.6
	(III)	3.8	5.0	5.7	4.3	2.1
	(IV)	3.4	4.9	4.2	3.3	1.6
	(V)	3.2	3.7	6.0	2.9	1.3
(50,480)	(I)	4.9	7.4	5.6	5.9	4.2
	(II)	3.9	7.3	5.4	6.0	5.1
	(III)	4.7	3.9	6.0	3.3	2.6
	(IV)	4.8	5.0	6.6	3.6	3.5
	(V)	4.9	4.4	6.2	4.7	2.6
(50,1440)	(I)	4.4	5.0	7.0	5.3	3.4
	(II)	4.4	5.0	7.2	5.4	4.6
	(III)	4.4	1.6	6.0	2.6	2.5
	(IV)	3.2	1.4	6.0	2.0	4.0
	(V)	4.2	1.0	5.9	2.5	2.5

Finally, we compare our SS test with another method. Here, we simply applied scaling by the sample variance instead of our estimates $\hat{\mathbf{D}}$. The corresponding test is denoted as SS*. We consider the multivariate mixture normal distribution $MN_{p,\gamma,100}$. $\mathbf{X}_{ij}$'s are generated from $\gamma f_p(\boldsymbol{\theta}_i, \mathbf{R}_i) + (1-\gamma) f_p(\boldsymbol{\theta}_i, 100\mathbf{R}_i)$ and $\gamma = 0.8$. The other settings are the same as (V), except that $\eta =: ||\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2||^2/\sqrt{\mathrm{tr}({\boldsymbol{\Sigma}_1}^2) + \mathrm{tr}(\boldsymbol{\Sigma}_2^2)} = 0.01$. Table A8 reports the empirical sizes and power comparison between SS and SS*. We can observe that the proposed SS test performs slightly better than SS* test. This demonstrates the usefulness of our robust estimators.

Table A4: Empirical power comparison at 5% significance under Scenarios (I)-(V) with $p = 120$, $n_1 = n_2 = 25$

Scenario	%	Equal Allocation					Linear Allocation				
		SS	SKK	CQ	BF	GCT	SS	SKK	CQ	BF	GCT
(I)	50%	31	43	38	38	26	32	43	39	38	28
	95%	38	51	45	45	16	33	45	38	39	12
(II)	50%	74	85	38	82	66	75	84	39	79	64
	95%	83	91	44	88	29	76	85	39	82	18
(III)	50%	59	41	41	36	26	57	39	40	36	23
	95%	67	48	49	43	13	60	42	44	37	9.1
(IV)	50%	76	69	43	61	49	82	71	45	65	41
	95%	83	89	50	85	15	93	62	44	54	10
(V)	50%	65	39	40	35	23	65	39	41	34	21
	95%	74	47	47	42	12	66	39	41	35	8.5

Table A5: Empirical power comparison at 5% significance under Scenarios (I)-(V) with $p = 480$, $n_1 = n_2 = 50$

Scenario	%	Equal Allocation					Linear Allocation				
		SS	SKK	CQ	BF	GCT	SS	SKK	CQ	BF	GCT
(I)	50%	73	81	79	78	73	72	81	77	78	72
	95%	76	83	79	80	62	73	80	77	77	52
(II)	50%	100	100	80	100	100	100	100	79	100	100
	95%	100	100	84	100	98	100	100	81	100	93
(III)	50%	97	75	78	73	67	97	75	79	73	69
	95%	100	78	82	77	56	99	75	80	73	49
(IV)	50%	100	96	82	93	90	100	96	81	95	89
	95%	100	97	84	95	78	100	95	83	92	63
(V)	50%	99	76	76	76	73	99	77	76	77	67
	95%	100	80	81	80	61	99	77	78	77	50

Table A6: Empirical power comparison at 5% significance under Scenarios (I)-(V) with $p = 1440$, $n_1 = n_2 = 50$

Scenario	%	Equal Allocation					Linear Allocation				
		SS	SKK	CQ	BF	GCT	SS	SKK	CQ	BF	GCT
(I)	50%	77	82	84	82	80	77	82	84	82	79
	95%	78	83	84	82	75	76	82	84	83	71
(II)	50%	99	100	85	100	100	99	100	86	100	100
	95%	100	100	86	100	100	100	100	85	100	100
(III)	50%	97	60	85	69	69	98	61	86	69	67
	95%	98	63	85	72	61	98	62	85	70	62
(IV)	50%	100	92	87	91	88	98	92	87	92	91
	95%	100	94	87	94	85	100	95	84	96	82
(V)	50%	99	63	84	76	75	98	64	85	76	71
	95%	100	67	85	77	68	98	65	85	76	67

Table A7: Empirical size and power comparison at 5% significance under Scenarios (VI)-(IX) with $p = 480$, $n_1 = n_2 = 50$

Scenario	%			Size			
		SS	SKK	CQ	BF	GCT	GCT*
(VI)		6.2	4.7	5.0	4.3	24.8	–
(VII)		6.1	8.0	7.8	4.7	23.4	–
(VIII)		5.3	6.0	6.5	5.3	23.3	–
(IX)		5.1	5.3	8.2	5.0	25.9	–
				Equal Allocation			
(VI)	50%	42	45	43	42	74	29
	95%	61	48	47	43	83	2.9
(VII)	50%	64	56	45	36	87	30
	95%	99	90	50	66	98	2.4
(VIII)	50%	42	45	45	44	77	30
	95%	58	52	49	49	83	2.6
(IX)	50%	44	44	45	41	77	28
	95%	61	51	49	44	82	2.6
				Linear Allocation			
(VI)	50%	44	46	44	43	79	21
	95%	59	45	43	41	66	0.7
(VII)	50%	79	67	46	43	93	21
	95%	100	96	47	76	85	0.6
(VIII)	50%	44	47	45	43	80	19
	95%	55	49	46	46	66	1.0
(IX)	50%	45	44	44	40	78	17
	95%	58	48	46	42	67	1.4

Table A8: Empirical size and power comparison at 5% significance

(n_i, p)	%	Size		Equal Allocation		Linear Allocation	
		SS*	SS	SS*	SS	SS*	SS
(25,120)	50%	3.6	3.5	34	39	34	39
	95%	–	–	40	48	34	41
(50,480)	50%	3.8	4.2	88	91	89	91
	95%	–	–	91	93	89	91
(50,1440)	50%	2.3	2.1	89	93	91	93
	95%	–	–	90	93	91	94

Appendix C: Discussion about non-negligible biases

A natural idea is mimicking Chen and Qin (2010) (or Bai and Saranadasa 1996) and considering the following test statistic

$$G_n = \frac{\sum_{i\neq j}^{n_1} \hat{\mathbf{V}}_{1i}^T \hat{\mathbf{V}}_{1j}}{n_1(n_1-1)} + \frac{\sum_{i\neq j}^{n_2} \hat{\mathbf{V}}_{2i}^T \hat{\mathbf{V}}_{2j}}{n_2(n_2-1)} - 2\frac{\sum_{i=1}^{n_1}\sum_{j=1}^{n_2} \hat{\mathbf{V}}_{1i}^T \hat{\mathbf{V}}_{2j}}{n_1 n_2}.$$

where $\hat{\mathbf{V}}_{ij} = U(\mathbf{X}_{ij} - \hat{\boldsymbol{\theta}})$ and $\hat{\boldsymbol{\theta}}$ satisfy $\sum_{i=1}^{2}\sum_{j=1}^{n_i} U(\mathbf{X}_{ij} - \boldsymbol{\theta}) = 0$. If the sample spatial median $\hat{\boldsymbol{\theta}}$ in G_n is replaced by the true value $\boldsymbol{\theta}$, it can be shown that under null hypothesis the resulting test statistic G_n' satisfies

$$\frac{G_n'}{\sqrt{\mathrm{var}(G_n')}} \xrightarrow{d} N(0,1), \text{ as } n \to \infty,\ p \to \infty, \tag{A.1}$$

which is parallel to the null distribution of Chen and Qin's (2010) test. Unfortunately, this result is not valid for G_n since the sample spatial median $\hat{\boldsymbol{\theta}}$ is only root-n consistent. When p is fixed, this is acceptable because it does not affect the asymptotic properties of G_n. However, this substitution would yield a bias-term which is not negligible when $n/p = O(1)$. Even worse, when $n/p = o(1)$, the test based on (A.1) would have an asymptotic size 1 under H_0. In addition, it seems difficult to calculate the bias numerically in a high-dimensional setting.

Here we provide the detail calculation of the bias term of G_n. Taking the same procedure as Lemma 3 and the assumption that the spatial median of $\mathbf{X}_{ij}-\boldsymbol{\theta}_i$ is unique and zero, we can show that, for $k=1,2$

$$\widehat{\boldsymbol{\theta}}=\boldsymbol{\theta}+\frac{1}{n_k}\mathbf{A}_4^{-1}\sum_{k=1}^{2}\sum_{i=1}^{n_k}\mathbf{V}_{ki}(1+o_p(1)),$$

where $\mathbf{A}_4=\tilde{c}_1\kappa(\mathbf{I}_p-\boldsymbol{\Omega}_1)+\tilde{c}_2(1-\kappa)(\mathbf{I}_p-\boldsymbol{\Omega}_2)$, $\boldsymbol{\Omega}_k=E(\mathbf{V}_{ki}\mathbf{V}_{ki}^T)$, $\tilde{c}_k=E||\mathbf{X}_{ki}-\boldsymbol{\theta}_k||^{-1}$, $\tilde{r}_{ki}=||\mathbf{X}_{ki}-\boldsymbol{\theta}_k||$ and $\mathbf{V}_{ki}=U(\mathbf{X}_{ki}-\boldsymbol{\theta}_k)$. Then,

$$\hat{\mathbf{V}}_{1i}=\mathbf{V}_{1i}-\frac{1}{\tilde{r}_{1i}}[\mathbf{I}_p-\mathbf{V}_{1i}\mathbf{V}_{1i}^T]\widehat{\boldsymbol{\theta}}-\frac{||\hat{\boldsymbol{\theta}}||^2}{2\tilde{r}_{1i}^2}\mathbf{V}_{1i}+o_p(n^{-1}).$$

Note that

$$\begin{aligned}
&E(\hat{\mathbf{V}}_{1i}^T\hat{\mathbf{V}}_{1j})\\
=&E\left\{\left[\mathbf{V}_{1i}-\frac{1}{\tilde{r}_{1i}}[\mathbf{I}_p-\mathbf{V}_{1i}\mathbf{V}_{1i}^T]\widehat{\boldsymbol{\theta}}-\frac{||\hat{\boldsymbol{\theta}}||^2}{2\tilde{r}_{1i}^2}\mathbf{V}_{1i}\right]^T\times\left[\mathbf{V}_{1j}-\frac{1}{\tilde{r}_{1j}}[\mathbf{I}_p-\mathbf{V}_{1j}\mathbf{V}_{1j}^T]\widehat{\boldsymbol{\theta}}-\frac{||\hat{\boldsymbol{\theta}}||^2}{2\tilde{r}_{1j}^2}\mathbf{V}_{1j}\right]\right\}+o(n^{-2})\\
=&E(\mathbf{V}_{1i}^T\mathbf{V}_{1j})-2E\left(\frac{1}{\tilde{r}_{1i}}\widehat{\boldsymbol{\theta}}^T[\mathbf{I}_p-\mathbf{V}_{1i}\mathbf{V}_{1i}^T]\mathbf{V}_{1j}\right)-2E\left(\frac{||\hat{\boldsymbol{\theta}}||^2}{2\tilde{r}_{1i}^2}\mathbf{V}_{1i}^T\mathbf{V}_{1j}\right)\\
&+E\left(\frac{1}{\tilde{r}_1^2}\widehat{\boldsymbol{\theta}}^T[\mathbf{I}_p-\mathbf{V}_{1i}\mathbf{V}_{1i}^T][\mathbf{I}_p-\mathbf{V}_{1j}\mathbf{V}_{1j}^T]\widehat{\boldsymbol{\theta}}\right)+2E\left(\frac{||\hat{\boldsymbol{\theta}}||^2}{2\tilde{r}_1^3}\mathbf{V}_{1i}^T[\mathbf{I}_p-\mathbf{V}_{1j}\mathbf{V}_{1j}^T]\widehat{\boldsymbol{\theta}}\right)\\
&+E\left(\frac{||\hat{\boldsymbol{\theta}}||^4}{2\tilde{r}_1^4}\mathbf{V}_{1i}^T\mathbf{V}_{1j})\right)+o(n^{-2})\doteq E(\mathbf{V}_{1i}^T\mathbf{V}_{1j})-2\Delta_1-\Delta_2+\Delta_3+2\Delta_4+\Delta_5+o(n^{-2}).
\end{aligned}$$

It can be verified that

$$\begin{aligned}
\Delta_1&=\frac{\tilde{c}_1}{n}\mathrm{tr}\left[\mathbf{A}_4^{-1}(\mathbf{I}_p-\boldsymbol{\Omega}_1)\boldsymbol{\Omega}_1\right](1+o(1));\\
\Delta_2&=\frac{1}{n^2}E(\tilde{r}_1^{-2})\mathrm{tr}\left[\mathbf{A}_4^{-2}\boldsymbol{\Omega}_1^2\right](1+o(1))=o(n^{-2});\\
\Delta_3&=\frac{2}{n^2}E(\tilde{r}_1^{-2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p-\boldsymbol{\Omega}_1]\mathbf{A}_4^{-1}\boldsymbol{\Omega}_1\right)+\frac{n_2}{n^2}E(\tilde{r}_1^{-2})\mathrm{tr}\left(A^{-1}[\mathbf{I}_p-\boldsymbol{\Omega}_1]^2\mathbf{A}_4^{-1}\boldsymbol{\Omega}_2\right)\\
&+\frac{n_1-2}{n^2}E(\tilde{r}_1^{-2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p-\boldsymbol{\Omega}_1]^2\mathbf{A}_4^{-1}\boldsymbol{\Omega}_1\right)+o(n^{-2});\\
\Delta_4&=E(\tilde{r}_1^{-3})O_p(n^{-1})=o(n^{-2});\\
\Delta_5&=E(\tilde{r}_1^{-4})O_p(n^{-2})=o(n^{-2}),
\end{aligned}$$

where $\tilde{r}_k = ||\mathbf{X}_{ki} - \boldsymbol{\theta}_k||$, $k = 1, 2$. Combining all these, we know that

$$
\begin{aligned}
E(\hat{\mathbf{V}}_{1i}^T\hat{\mathbf{V}}_{1j}) =& -\frac{2\tilde{c}_1}{n}\mathrm{tr}\left[\mathbf{A}_4^{-1}(\mathbf{I}_p - \boldsymbol{\Omega}_1)\boldsymbol{\Omega}_1\right] + \frac{n_2}{n^2}E(\frac{1}{\tilde{r}_1^2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_1]^2\mathbf{A}_4^{-1}\boldsymbol{\Omega}_2\right) \\
&+ \frac{n_1 - 2}{n^2}E(\frac{1}{\tilde{r}_1^2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_1]^2\mathbf{A}_4^{-1}\boldsymbol{\Omega}_1\right) + \frac{2}{n^2}E(\frac{1}{\tilde{r}_1^2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_1]\mathbf{A}_4^{-1}\boldsymbol{\Omega}_1\right) + o(n^{-2}).
\end{aligned}
$$

Taking the same procedure, we can obtain that

$$
\begin{aligned}
E(\hat{\mathbf{V}}_{2i}^T\hat{\mathbf{V}}_{2j}) =& -\frac{2\tilde{c}_2}{n}\mathrm{tr}\left[\mathbf{A}_4^{-1}(\mathbf{I}_p - \boldsymbol{\Omega}_2)\boldsymbol{\Omega}_2\right] + \frac{n_1}{n^2}E(\frac{1}{\tilde{r}_2^2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_2]^2\mathbf{A}_4^{-1}\boldsymbol{\Omega}_1\right) \\
&+ \frac{n_2 - 2}{n^2}E(\frac{1}{\tilde{r}_2^2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_2]^2\mathbf{A}_4^{-1}\boldsymbol{\Omega}_2\right) + \frac{2}{n^2}E(\frac{1}{\tilde{r}_2^2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_2]\mathbf{A}_4^{-1}\boldsymbol{\Omega}_2\right) + o(n^{-2}); \\
E(\hat{\mathbf{V}}_{1i}^T\hat{\mathbf{V}}_{2j}) =& -\frac{\tilde{c}_1}{n}\mathrm{tr}\left[\mathbf{A}_4^{-1}(\mathbf{I}_p - \boldsymbol{\Omega}_1)\boldsymbol{\Omega}_2\right] + \frac{n_2 - 1}{n^2}\tilde{c}_1\tilde{c}_2\mathrm{tr}(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_1][\mathbf{I}_p - \boldsymbol{\Omega}_2]\mathbf{A}_4^{-1}\boldsymbol{\Omega}_2) \\
&- \frac{\tilde{c}_2}{n}\mathrm{tr}\left[\mathbf{A}_4^{-1}(\mathbf{I}_p - \boldsymbol{\Omega}_2)\boldsymbol{\Omega}_1\right] + \frac{n_1 - 1}{n^2}\tilde{c}_1\tilde{c}_2\mathrm{tr}(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_1][\mathbf{I}_p - \boldsymbol{\Omega}_2]\mathbf{A}_4^{-1}\boldsymbol{\Omega}_1) \\
&+ \frac{1}{n^2}E(\frac{1}{\tilde{r}_1\tilde{r}_2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_1]\mathbf{A}_4^{-1}\boldsymbol{\Omega}_2\right) + \frac{1}{n^2}E(\frac{1}{\tilde{r}_1\tilde{r}_2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_2]\mathbf{A}_4^{-1}\boldsymbol{\Omega}_1\right) + o(n^{-2}).
\end{aligned}
$$

So the expectation of G_n is

$$
\begin{aligned}
E(G_n) =& -\frac{2\tilde{c}_1}{n}\mathrm{tr}\left[\mathbf{A}_4^{-1}(\mathbf{I}_p - \boldsymbol{\Omega}_1)\boldsymbol{\Omega}_1\right] + \frac{n_2}{n^2}E(\frac{1}{\tilde{r}_1^2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_1]^2\mathbf{A}_4^{-1}\boldsymbol{\Omega}_2\right) \\
&+ \frac{n_1 - 2}{n^2}E(\frac{1}{\tilde{r}_1^2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_1]^2\mathbf{A}_4^{-1}\boldsymbol{\Omega}_1\right) \\
&- \frac{2\tilde{c}_2}{n}\mathrm{tr}\left[\mathbf{A}_4^{-1}(\mathbf{I}_p - \boldsymbol{\Omega}_2)\boldsymbol{\Omega}_2\right] + \frac{n_1}{n^2}E(\frac{1}{\tilde{r}_2^2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_2]^2\mathbf{A}_4^{-1}\boldsymbol{\Omega}_1\right) \\
&+ \frac{n_2 - 2}{n^2}E(\frac{1}{\tilde{r}_2^2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_2]^2\mathbf{A}_4^{-1}\boldsymbol{\Omega}_2\right) \\
&+ \frac{2\tilde{c}_1}{n}\mathrm{tr}\left[\mathbf{A}_4^{-1}(\mathbf{I}_p - \boldsymbol{\Omega}_1)\boldsymbol{\Omega}_2\right] - \frac{2(n_2 - 1)}{n^2}\tilde{c}_1\tilde{c}_2\mathrm{tr}(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_1][\mathbf{I}_p - \boldsymbol{\Omega}_2]\mathbf{A}_4^{-1}\boldsymbol{\Omega}_2) \\
&+ \frac{2\tilde{c}_2}{n}\mathrm{tr}\left[\mathbf{A}_4^{-1}(\mathbf{I}_p - \boldsymbol{\Omega}_2)\boldsymbol{\Omega}_1\right] - \frac{2(n_1 - 1)}{n^2}\tilde{c}_1\tilde{c}_2\mathrm{tr}(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_1][\mathbf{I}_p - \boldsymbol{\Omega}_2]\mathbf{A}_4^{-1}\boldsymbol{\Omega}_1) \\
&+ \frac{2}{n^2}E(\frac{1}{\tilde{r}_1^2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_1]\mathbf{A}_4^{-1}\boldsymbol{\Omega}_1\right) + \frac{2}{n^2}E(\frac{1}{\tilde{r}_2^2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_2]\mathbf{A}_4^{-1}\boldsymbol{\Omega}_2\right) \\
&- \frac{2}{n^2}E(\frac{1}{\tilde{r}_1\tilde{r}_2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_1]\mathbf{A}_4^{-1}\boldsymbol{\Omega}_2\right) - \frac{2}{n^2}E(\frac{1}{\tilde{r}_1\tilde{r}_2})\mathrm{tr}\left(\mathbf{A}_4^{-1}[\mathbf{I}_p - \boldsymbol{\Omega}_2]\mathbf{A}_4^{-1}\boldsymbol{\Omega}_1\right) + o(n^{-2}).
\end{aligned}
$$

Similarly, we can also show that

$$
\mathrm{var}(G_n) = \frac{2}{n_1(n_1 - 1)}\mathrm{tr}(\boldsymbol{\Omega}_1^2) + \frac{2}{n_2(n_2 - 1)}\mathrm{tr}(\boldsymbol{\Omega}_2^2) + \frac{4}{n_1 n_2}\mathrm{tr}(\boldsymbol{\Omega}_1\boldsymbol{\Omega}_2) + o(n^{-2}).
$$

If $n/p = O(1)$, $E(G_n) = O(n^{-1})$ and $\mathrm{var}(G_n) = O(n^{-2})$. Then, $E(G_n)$ is a non-negligible bias term to the asymptotic variance $\mathrm{var}(G_n)$.

Appendix D: Proofs of Lemmas and Proposition 2

Proof of Lemma 2:

$$U(\mathbf{D}_i^{-1/2}(\mathbf{X}_{ij}-\boldsymbol{\theta}_i)) = \frac{\mathbf{D}_i^{-1/2}\boldsymbol{\Sigma}_i^{-1/2}\boldsymbol{\varepsilon}_{ij}}{(\boldsymbol{\varepsilon}_{ij}^T\mathbf{R}_i\boldsymbol{\varepsilon}_{ij})^{1/2}} = \frac{\mathbf{D}_i^{-1/2}\boldsymbol{\Sigma}_i^{-1/2}\mathbf{u}_{ij}}{(1+\mathbf{u}_{ij}^T(\mathbf{R}_i-\mathbf{I}_p)\mathbf{u}_{ij})^{1/2}}$$
$$= \mathbf{D}_i^{-1/2}\boldsymbol{\Sigma}_i^{-1/2}\mathbf{u}_{ij} + C_1\mathbf{u}_{ij}^T(\mathbf{R}_i-\mathbf{I}_p)\mathbf{u}_{ij}\mathbf{D}_i^{-1/2}\boldsymbol{\Sigma}_i^{-1/2}\mathbf{u}_{ij},$$

where C_1 is a bounded random variable between -0.5 and $-0.5(1+\mathbf{u}_{ij}^T(\mathbf{R}_i-\mathbf{I}_p)\mathbf{u}_{ij})^{-3/2}$. Thus, according to the Cauchy inequality,

$$E\left(U(\mathbf{D}_i^{-1/2}(\mathbf{X}_{ij}-\boldsymbol{\theta}_i))\right) \le C_2\{E(\mathbf{u}_{ij}^T(\mathbf{R}_i-\mathbf{I}_p)\mathbf{u}_{ij})^2 E(\mathbf{D}_i^{-1/2}(\mathbf{X}_{ij}-\boldsymbol{\theta}_i))^2\}^{1/2}$$
$$= O\left(p^{-1}\sqrt{\mathrm{tr}(\mathbf{R}_i^2)-p}\right) = o(n^{-1/2})$$

by Condition (C4). Similarly, we can show that

$$U(\mathbf{D}_i^{-1/2}(\mathbf{X}_{ij}-\boldsymbol{\theta}_i))U(\mathbf{D}_i^{-1/2}(\mathbf{X}_{ij}-\boldsymbol{\theta}_i))^T$$
$$= \frac{1}{1+\mathbf{u}_{ij}^T(\mathbf{R}_i-\mathbf{I}_p)\mathbf{u}_{ij}}\mathbf{D}_i^{-1/2}\boldsymbol{\Sigma}_i^{-1/2}\mathbf{u}_{ij}\mathbf{u}_{ij}^T\boldsymbol{\Sigma}_i^{-1/2}\mathbf{D}_i^{-1/2}$$
$$= \mathbf{D}_i^{-1/2}\boldsymbol{\Sigma}_i^{-1/2}\mathbf{u}_{ij}\mathbf{u}_{ij}^T\boldsymbol{\Sigma}_i^{-1/2}\mathbf{D}_i^{-1/2} + C_3(\mathbf{u}_{ij}^T(\mathbf{R}_i-\mathbf{I}_p)\mathbf{u}_{ij})\mathbf{D}_i^{-1/2}\boldsymbol{\Sigma}_i^{-1/2}\mathbf{u}_{ij}\mathbf{u}_{ij}^T\boldsymbol{\Sigma}_i^{-1/2}\mathbf{D}_i^{-1/2},$$

where C_3 is a bounded random variable between -1 and $-(1+\mathbf{u}_{ij}^T(\mathbf{R}_i-\mathbf{I}_p)\mathbf{u}_{ij})$. Thus, by the Cauchy inequality,

$$E\left(\mathrm{diag}\left\{E\left(U(\mathbf{D}_i^{-1/2}(\mathbf{X}_{ij}-\boldsymbol{\theta}_i))U(\mathbf{D}_i^{-1/2}(\mathbf{X}_{ij}-\boldsymbol{\theta}_i))^T\right)\right\} - p^{-1}\mathbf{I}_p\right)$$
$$\le C_4\{E(\mathbf{u}_{ij}^T(\mathbf{R}_i-\mathbf{I}_p)\mathbf{u}_{ij})^2 E(\mathrm{diag}\{D_i^{-1/2}\boldsymbol{\Sigma}_i^{-1/2}\mathbf{u}_{ij}\mathbf{u}_{ij}^T\boldsymbol{\Sigma}_i^{-1/2}\mathbf{D}_i^{-1/2}\} - p^{-1}\mathbf{I}_p)^2\}^{1/2}$$
$$= O\left(p^{-1}\sqrt{\mathrm{tr}(\mathbf{R}_i^2)-p}\right) = o(n^{-1/2})$$

by Condition (C4). The above two equations define the functional equation for each component of $\boldsymbol{\eta}_i$,

$$T_{ij}(F_i, \boldsymbol{\eta}_{ij}) = o_p(n^{-1/2}), \tag{A.2}$$

where F_i is the distribution function of $\mathbf{X}_{ij}$, $i = 1, 2$, $\boldsymbol{\eta}_i = (\eta_{i1}, \cdots, \eta_{i2p})$. Similar to Hettmansperger and Randles (2002), the linearisation of this equation produces

$$\sqrt{n_i}(\hat{\boldsymbol{\eta}}_{ij} - \boldsymbol{\eta}_{ij}) = -\mathbf{H}_{ij}^{-1}\sqrt{n_i}(T_{ij}(F_{ni}, \boldsymbol{\eta}_{ij}) - T_{ij}(F_i, \boldsymbol{\eta}_{ij})) + o_p(1),$$

where F_{ni} is the empirical distribution function of $\mathbf{X}_{ij}, j = 1, \cdots, n_i$, $\mathbf{H}_{ij}$ is the corresponding Hessian matrix of the functional defined in (A.2), and

$$T_i(F_{ni}, \boldsymbol{\eta}_i) = \left(n_i^{-1} \sum_{j=1}^{n_i} \mathbf{U}_{ij}^T, vec\left(\text{diag}\left(n_i^{-1} \sum_{j=1}^{n_i} \mathbf{U}_{ij}\mathbf{U}_{ij}^T - p^{-1}\mathbf{I}_p \right)\right)\right)^T,$$

where $T_i(F_{ni}, \boldsymbol{\eta}_i) = (T_{i1}(F_{ni}, \eta_{i1}), \cdots, T_{i2p}(F_{ni}, \eta_{i2p}))$ and $vec(A)$ means the vector of the diagonal matrix of A. For each variance estimator $\hat{d}_{ij}$, we have

$$\sqrt{n}(\hat{d}_{ij} - d_{ij}) \xrightarrow{p} N(0, \zeta_{ij}^2),$$

where ζ_{ij}^2 is the corresponding asymptotic variance. Define $\zeta_{max} = \max_{1 \le i \le 2, 1 \le j \le p} \zeta_{ij}$. As $p \to \infty$,

$$\begin{aligned}
& P\left(\max_{1 \le j \le p} (\hat{d}_{ij} - d_{ij}) > \sqrt{2}\zeta_{max} n_i^{-1/2} (\log p)^{1/2} \right) \\
& \le \sum_{j=1}^{p} P\left(\sqrt{n_i}(\hat{d}_{ij} - d_{ij}) > \sqrt{2}\zeta_{max} (\log p)^{1/2} \right) \\
& = \sum_{i=1}^{p} \left(1 - \Phi(\sqrt{2}\zeta_{max}\sigma_{ij}^{-1} (\log p)^{1/2}) \right) \le p\left(1 - \Phi((2\log p)^{1/2})\right) \\
& \le \frac{p}{\sqrt{4\pi \log p}} e^{-\log p} = (4\pi)^{-1/2} (\log p)^{-1/2} \to 0.
\end{aligned}$$

Finally, $\max_{1 \le j \le p}(\hat{d}_{ij} - d_{ij}) = O_p(n_i^{-1/2} (\log p)^{1/2})$. □

Proof of Lemma 4: By the Taylor expansion,

$$
\begin{aligned}
&U(\hat{\mathbf{D}}_{1,i}^{-1/2}(\mathbf{X}_{1i}-\hat{\boldsymbol{\theta}}_{2,j}))\\
&\quad=U(\mathbf{D}_1^{-1/2}(\mathbf{X}_{1i}-\boldsymbol{\theta})-\mathbf{D}_1^{-1/2}\hat{\boldsymbol{\mu}}_{2,j}+(\hat{\mathbf{D}}_{1,i}^{-1/2}-\mathbf{D}_1^{-1/2})(\mathbf{X}_{1i}-\boldsymbol{\theta})-(\hat{\mathbf{D}}_{1,i}^{-1/2}-\mathbf{D}_1^{-1/2})\hat{\boldsymbol{\mu}}_{2,j})\\
&\quad=\mathbf{U}_{1i}-\frac{1}{r_{1i}}[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T]\mathbf{D}_1^{-1/2}\hat{\boldsymbol{\mu}}_{2,j}+\frac{1}{r_{1i}}[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T](\hat{\mathbf{D}}_{1,i}^{-1/2}-\mathbf{D}_1^{-1/2})(\mathbf{X}_{1i}-\boldsymbol{\theta})\\
&\qquad-\frac{1}{r_{1i}}[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T](\hat{\mathbf{D}}_{1,i}^{-1/2}-\mathbf{D}_1^{-1/2})\hat{\boldsymbol{\mu}}_{2,j}\\
&\qquad+\frac{1}{2r_{1i}^2}||(\hat{\mathbf{D}}_{1,i}^{-1/2}-\mathbf{D}_1^{-1/2})(\mathbf{X}_{1i}-\boldsymbol{\theta})-\hat{\mathbf{D}}_{1,i}^{-1/2}\hat{\boldsymbol{\mu}}_{2,j}||^2\mathbf{U}_{1i}+o_p(n^{-1})\\
&\quad=\mathbf{U}_{1i}-\frac{1}{r_{1i}}[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T]\mathbf{D}_1^{-1/2}\hat{\boldsymbol{\mu}}_{2,j}+[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T](\hat{\mathbf{D}}_{1,i}^{-1/2}\mathbf{D}_1^{1/2}-\mathbf{I}_p)\mathbf{U}_{1i}\\
&\qquad-\frac{1}{r_{1i}}[\mathbf{I}_p-\mathbf{U}_{1i}\mathbf{U}_{1i}^T](\hat{\mathbf{D}}_{1,i}^{-1/2}-\mathbf{D}_1^{-1/2})\hat{\boldsymbol{\mu}}_{2,j}\\
&\qquad+\frac{1}{2r_{1i}^2}||(\hat{\mathbf{D}}_{1,i}^{-1/2}-\mathbf{D}_1^{-1/2})(\mathbf{X}_{1i}-\boldsymbol{\theta})-\hat{\mathbf{D}}_{1,i}^{-1/2}\hat{\boldsymbol{\mu}}_{2,j}||^2\mathbf{U}_{1i}+o_p(n^{-1}).
\end{aligned}
$$

□

Proof of Lemma 6: Note that

$$
\begin{aligned}
E[Z_{nj}^2|\mathcal{F}_{n,j-1}]&=\frac{1}{\tilde{n}_j^2(\tilde{n}_j-1)^2}E\left\{\left(\sum_{i=1}^{j-1}\mathbf{Y}_i^T\boldsymbol{\Xi}_{i,j}\mathbf{Y}_j\right)^2|\mathcal{F}_{n,j-1}\right\}\\
&=\frac{1}{\tilde{n}_j^2(\tilde{n}_j-1)^2}E\left\{\sum_{i_1,i_2=1}^{j-1}\mathbf{Y}_{i_1}^T\boldsymbol{\Xi}_{i_1,j}\mathbf{Y}_j\mathbf{Y}_j^T\boldsymbol{\Xi}_{i_2,j}^T\mathbf{Y}_{i_2}|\mathcal{F}_{n,j-1}\right\}\\
&=\frac{1}{\tilde{n}_j^2(\tilde{n}_j-1)^2}\sum_{i_1,i_2=1}^{j-1}\mathbf{Y}_{i_1}^T\boldsymbol{\Xi}_{i_1,j}E\left(\mathbf{Y}_j\mathbf{Y}_j^T|\mathcal{F}_{n,j-1}\right)\boldsymbol{\Xi}_{i_2,j}^T\mathbf{Y}_{i_2}\\
&=\frac{1}{\tilde{n}_j^2(\tilde{n}_j-1)^2}\sum_{i_1,i_2=1}^{j-1}\mathbf{Y}_{i_1}^T\tilde{\boldsymbol{\Lambda}}_{i_1,i_2,j}\mathbf{Y}_{i_2},
\end{aligned}
$$

where $\tilde{\boldsymbol{\Lambda}}_{i_1,i_2,j}=p^{-1}\boldsymbol{\Xi}_{i_1,j}\boldsymbol{\Xi}_{i_2,j}^T$, $\tilde{n}_j=n_1$, for $j\in[1,n_1]$ and $\tilde{\boldsymbol{\Lambda}}_{i_1,i_2,j}=p^{-1}\boldsymbol{\Xi}_{i_1,j}\boldsymbol{\Xi}_{i_2,j}^T$, $\tilde{n}_j=n_2$, for $j\in[n_1,n]$, and

$$
\boldsymbol{\Xi}_{i,j}=\begin{cases}\mathbf{A}_1, & i,j\in\{1,2,\ldots,n_1\},\\ \mathbf{A}_3, & i\in\{1,2,\ldots,n_1\}, j\in\{n_1+1,\ldots,n\},\\ \mathbf{A}_2, & i,j\in\{n_1+1,n_1+2,\ldots,n\}.\end{cases}
$$

Define $\eta_n = \sum_{j=2}^n E[Z_{nj}^2|\mathcal{F}_{n,j-1}]$. By some tedious algebra, we can obtain that $E(\eta_n) = \frac{1}{4}\sigma_n^2(1+o(1))$.

Now write $E(\eta_n^2)$ as

$$\begin{aligned}
E(\eta_n^2) =& E\left\{\sum_{j=2}^{n}\frac{1}{\tilde{n}_j^2(\tilde{n}_j-1)^2}\sum_{i_1,i_2=1}^{j-1}\mathbf{Y}_{i_1}^T\tilde{\mathbf{\Lambda}}_{i_1,i_2j}\mathbf{Y}_{i_2}\right\}^2\\
=&2E\left\{\sum_{2\le j_1<j_2}^{n}\frac{1}{\tilde{n}_{j_1}^2(\tilde{n}_{j_1}-1)^2}\frac{1}{\tilde{n}_{j_2}^2(\tilde{n}_{j_2}-1)^2}\sum_{i_1,i_2=1}^{j_1-1}\sum_{i_3,i_4=1}^{j_2-1}\mathbf{Y}_{i_1}^T\tilde{\mathbf{\Lambda}}_{i_1,i_2,j_1}\mathbf{Y}_{i_2}\mathbf{Y}_{i_3}^T\tilde{\mathbf{\Lambda}}_{i_3,i_4,j_2}\mathbf{Y}_{i_4}\right\}\\
&+E\left\{\sum_{j=2}^{n}\frac{1}{\tilde{n}_j^4(\tilde{n}_j-1)^4}\sum_{i_1,i_2=1}^{j-1}\sum_{i_3,i_4=1}^{j-1}\mathbf{Y}_{i_1}^T\tilde{\mathbf{\Lambda}}_{i_1,i_2,j}\mathbf{Y}_{i_2}\mathbf{Y}_{i_3}^T\tilde{\mathbf{\Lambda}}_{i_3,i_4,j}\mathbf{Y}_{i_4}\right\}\\
\doteq& L_1+L_2
\end{aligned}$$

Consider the first part L_1.

$$\begin{aligned}
&E\left\{\sum_{2\le j_1<j_2}^{n}\frac{1}{\tilde{n}_{j_1}^2(\tilde{n}_{j_1}-1)^2}\frac{1}{\tilde{n}_{j_2}^2(\tilde{n}_{j_2}-1)^2}\sum_{i_1,i_2=1}^{j_1-1}\sum_{i_3,i_4=1}^{j_2-1}\mathbf{Y}_{i_1}^T\tilde{\mathbf{\Lambda}}_{i_1,i_2,j_1}\mathbf{Y}_{i_2}\mathbf{Y}_{i_3}^T\tilde{\mathbf{\Lambda}}_{i_3,i_4,j_2}\mathbf{Y}_{i_4}\right\}\\
=&E\left\{\sum_{2\le j_1<j_2}^{n}\frac{1}{\tilde{n}_{j_1}^2(\tilde{n}_{j_1}-1)^2}\frac{1}{\tilde{n}_{j_2}^2(\tilde{n}_{j_2}-1)^2}\sum_{i=1}^{j_1-1}\sum_{i=1}^{j_2-1}\mathbf{Y}_{i}^T\tilde{\mathbf{\Lambda}}_{i,i,j_1}\mathbf{Y}_{i}\mathbf{Y}_{i}^T\tilde{\mathbf{\Lambda}}_{i,i,j_2}\mathbf{Y}_{i}\right\}\\
&+E\left\{\sum_{2\le j_1<j_2}^{n}\frac{1}{\tilde{n}_{j_1}^2(\tilde{n}_{j_1}-1)^2}\frac{1}{\tilde{n}_{j_2}^2(\tilde{n}_{j_2}-1)^2}\sum_{i_1=1}^{j_1-1}\sum_{i_2=1}^{j_2-1}\mathbf{Y}_{i_1}^T\tilde{\mathbf{\Lambda}}_{i_1,i_1,j_1}\mathbf{Y}_{i_1}\mathbf{Y}_{i_2}^T\tilde{\mathbf{\Lambda}}_{i_2,i_2,j_2}\mathbf{Y}_{i_2}\right\}\\
&+E\left\{\sum_{2\le j_1<j_2}^{n}\frac{1}{\tilde{n}_{j_1}^2(\tilde{n}_{j_1}-1)^2}\frac{1}{\tilde{n}_{j_2}^2(\tilde{n}_{j_2}-1)^2}\sum_{i_1=1}^{j_1-1}\sum_{i_2=1}^{j_2-1}\mathbf{Y}_{i_2}^T\tilde{\mathbf{\Lambda}}_{i_2,i_1,j_1}\mathbf{Y}_{i_1}\mathbf{Y}_{i_1}^T\tilde{\mathbf{\Lambda}}_{i_1,i_2,j_2}\mathbf{Y}_{i_2}\right\}\\
\doteq& L_{11}+L_{12}+L_{13}
\end{aligned}$$

Taking the same procedure as Lemma 5 and some tedious calculations, we can verify that $L_{11} = o(\sigma_n^4)$, $L_{12}+L_{13} = E^2(\eta_n)$ and $E(L_2^2) = o(\sigma_n^4)$. So, $\text{var}(\eta_n) = E(\eta_n^2) - E^2(\eta_n) = o(\sigma_n^4)$. This completes the proof of Lemma 6. $\square$

Proof of Lemma 7: First of all, we note that

$$\sigma_n^{-2}\sum_{j=2}^{n}E[Z_{nj}^2 I(|Z_{nj}|>\epsilon\sigma_n|)|\mathcal{F}_{n,j-1}] \le \sigma_n^{-4}\epsilon^{-2}\sum_{j=2}^{n}E[Z_{nj}^4|\mathcal{F}_{n,j-1}].$$

Accordingly, the assertion of this lemma is true if we can show

$$E\left\{\sum_{j=2}^{n} E[Z_{nj}^4|\mathcal{F}_{n,j-1}]\right\} = o(\sigma_n^4).$$

Notice that

$$E\left\{\sum_{j=2}^{n} E[Z_{nj}^4|\mathcal{F}_{n,j-1}]\right\} = \sum_{j=2}^{n} E(Z_{nj}^4) = O(n^{-8})\sum_{j=2}^{n} E\left(\sum_{i=1}^{j-1}\phi_{ij}\right)^4.$$

Similar to Chen and Qin (2010), the last term can be decomposed as $3Q+P$, where

$$Q = O(n^{-8})\sum_{j=2}^{n}\sum_{s\neq t}^{j-1} E(\mathbf{Y}_j^T\mathbf{\Xi}_{j,s}\mathbf{Y}_s\mathbf{Y}_s^T\mathbf{\Xi}_{s,j}\mathbf{Y}_j\mathbf{Y}_j^T\mathbf{\Xi}_{j,t}\mathbf{Y}_t\mathbf{Y}_t^T\mathbf{\Xi}_{t,j}\mathbf{Y}_j),$$
$$P = O(n^{-8})\sum_{j=2}^{n}\sum_{s=1}^{j-1} E(\mathbf{Y}_s^T\mathbf{\Xi}_{s,j}\mathbf{Y}_j)^4.$$

Note that

$$\begin{aligned}
Q =&O(n^{-8})\sum_{j=2}^{n}\sum_{s\neq t}^{j-1} E(\mathbf{Y}_j^T\mathbf{\Xi}_{j,s}\mathbf{Y}_s\mathbf{Y}_s^T\mathbf{\Xi}_{s,j}\mathbf{Y}_j\mathbf{Y}_j^T\mathbf{\Xi}_{j,t}\mathbf{Y}_t\mathbf{Y}_t^T\mathbf{\Xi}_{t,j}\mathbf{Y}_j)\}\\
=&O(n^{-8}p^{-2})\Bigg\{\sum_{j=2}^{n_1}\sum_{s\neq t}^{j-1} E(\mathbf{Y}_j^T\mathbf{A}_1\mathbf{A}_1\mathbf{Y}_j\mathbf{Y}_j^T\mathbf{A}_1\mathbf{A}_1\mathbf{Y}_j)\\
&+\sum_{j=n_1+1}^{n}\sum_{s\neq t}^{j-1} p^{-2}E(\mathbf{Y}_j^T\mathbf{\Xi}_{j,s}\mathbf{\Xi}_{s,j}\mathbf{Y}_j\mathbf{Y}_j^T\mathbf{\Xi}_{j,t}\mathbf{\Xi}_{t,j}\mathbf{Y}_j)\Bigg\}\\
=&O(n^{-1})\sigma_n^4,
\end{aligned}$$

where the last equation follows from the similar procedure in Lemma 5 with Condition (C2). Accordinglu, we can verify that $Q = o(\sigma_n^4)$. In addition,

$$\begin{aligned}
P =&O(n^{-8})\sum_{j=2}^{n}\sum_{s=1}^{j-1} E(\mathbf{Y}_s^T\mathbf{\Xi}_{s,j}\mathbf{Y}_j)^4\\
=&O(n^{-8})\left\{\sum_{j=2}^{n_1}\sum_{s=1}^{j-1} E(\mathbf{Y}_s^T\mathbf{A}_1\mathbf{Y}_j)^4 + \sum_{j=n_1+1}^{n}\sum_{s=1}^{n_1} E(\mathbf{Y}_s^T\mathbf{A}_3\mathbf{Y}_j)^4 + \sum_{j=n_1+1}^{n}\sum_{s=n_1+1}^{j-1} E(\mathbf{Y}_s^T\mathbf{A}_2\mathbf{Y}_j)^4\right\}\\
\doteq&O(n^{-8})(P_1+P_2+P_3).
\end{aligned}$$

As the procedures for handling P_1, P_2, P_3 are similar, let us only consider P_2. Define $\mathbf{A}_3 = (v_{ij})_{p\times p}$, and accordingly we can write

$$
\begin{aligned}
E(\mathbf{u}_{1s}^T\mathbf{A}_3\mathbf{u}_{2t})^4 &= E\left(\sum_{i=1}^{p}\sum_{j=1}^{p} v_{ij}u_{1si}u_{2tj}\right)^4 \\
&= \sum_{i_1,\ldots,i_4=1}^{p}\sum_{j_1,\ldots,j_4=1}^{p} v_{i_1j_1}v_{i_2j_2}v_{i_3j_3}v_{i_4j_4}E(u_{1si_1}u_{1si_2}u_{1si_3}u_{1si_4})E(u_{2tj_1}u_{2tj_2}u_{2tj_3}u_{2tj_4}) \\
&= O(p^{-4})\sum_{i_1,\ldots,i_4=1}^{p}\sum_{j_1,\ldots,j_4=1}^{p} v_{i_1j_1}v_{i_2j_2}v_{i_3j_3}v_{i_4j_4}.
\end{aligned}
$$

By the Cauchy inequality, we have

$$
\begin{aligned}
\sum_{i_1,i_2,i_3,i_4=1}^{p}\sum_{j_1,j_2,j_3,j_4=1}^{p} v_{i_1j_1}v_{i_2j_2}v_{i_3j_3}v_{i_4j_4} &\le \frac{1}{4}\sum_{i_1,i_2,i_3,i_4=1}^{p}\sum_{j_1,j_2,j_3,j_4=1}^{p}(v_{i_1j_1}^2+v_{i_2j_2}^2)(v_{i_3j_3}^2+v_{i_4j_4}^2) \\
&= \sum_{i_1,i_2,j_1,j_2=1}^{p} v_{i_1j_1}^2v_{i_2j_2}^2 = \left(\sum_{i_1,j_1} v_{i_1j_1}^2\right)^2 = \mathrm{tr}^2(\mathbf{A}_3^2).
\end{aligned}
$$

Similar to the proof of Lemma 5, we obtain that $n^{-2}P_2 = O(p^{-4}\mathrm{tr}(\mathbf{A}_3^2)^2)$ and thus $O(n^{-8}P_2) = o(\sigma_n^4)$. Similarly, $O(n^{-8}P_1) = o(\sigma_n^4)$ and $O(n^{-8}P_3) = o(\sigma_n^4)$. This completes the proof of the lemma. $\square$

Proof of Proposition 2 First, we will show the ratio consistency of $\hat{c}_i$.

$$
\begin{aligned}
||\hat{\mathbf{D}}_{i,j}^{-1/2}(\mathbf{X}_{ij}-\hat{\boldsymbol{\theta}}_{i,j})|| =&||\mathbf{D}_i^{-1/2}(\mathbf{X}_{ij}-\boldsymbol{\theta}_i)||(1+r_{ij}^{-2}||(\hat{\mathbf{D}}_{i,j}^{-1/2}-\mathbf{D}_i^{-1/2})(\mathbf{X}_{ij}-\boldsymbol{\theta}_i)||^2 \\
&+r_{ij}^{-2}||\hat{\mathbf{D}}_{ij}^{-1/2}\hat{\boldsymbol{\mu}}_{i,j}||^2+2r_{ij}^{-1}\mathbf{U}_{ij}^T(\hat{\mathbf{D}}_{i,j}^{-1/2}-\mathbf{D}_i^{-1/2})\mathbf{D}_i^{1/2}\mathbf{U}_{ij} \\
&-2r_{ij}^{-1}\mathbf{U}_{ij}^T\hat{\mathbf{D}}_{ij}^{-1/2}\hat{\boldsymbol{\mu}}_{i,j}-2r_{ij}^{-1}\mathbf{U}_{ij}\mathbf{D}_i^{1/2}(\hat{\mathbf{D}}_{i,j}^{-1/2}-\mathbf{D}_i^{-1/2})\hat{\mathbf{D}}_{ij}^{-1/2}\hat{\boldsymbol{\mu}}_{i,j})^{1/2}.
\end{aligned}
$$

According to Lemmas 2 and 3, we can show that $r_{ij}^{-2}||(\hat{\mathbf{D}}_{i,j}^{-1/2}-\mathbf{D}_i^{-1/2})(\mathbf{X}_{ij}-\boldsymbol{\theta}_i)||^2 = O_p((\log p/n)^{1/2}) = o_p(1)$ and $r_{ij}^{-2}||\hat{\mathbf{D}}_{ij}^{-1/2}\hat{\boldsymbol{\mu}}_{i,j}||^2 = O_p(n^{-1}) = o_p(1)$ and by the Cauchy inequality, the other parts are also $o_p(1)$. So,

$$
n_i^{-1}\sum_{j=1}^{n_i}||\hat{\mathbf{D}}_{i,j}^{-1/2}(\mathbf{X}_{ij}-\hat{\boldsymbol{\theta}}_{i,j})||^{-1} = \left(n_i^{-1}\sum_{j=1}^{n_i}||\mathbf{D}_i^{-1/2}(\mathbf{X}_{ij}-\boldsymbol{\theta}_i)||^{-1}\right)(1+o_p(1)).
$$

Obviously, we can show that $E(n_i^{-1}\sum_{j=1}^{n_i} r_{ij}^{-1}) = c_i$ and $\mathrm{var}(n_i^{-1}c_i^{-1}\sum_{j=1}^{n_i} r_{ij}^{-1}) = O(n^{-1})$. Then the ratio consistent of $\hat{c}_i$ follows. We only present the proof for the ratio consistency of $\widehat{\mathrm{tr}(\mathbf{A}_1^2)}$ as the proofs of the other two follow in the same way.

By Lemma 4 again,

$$\frac{p^2\hat{c}_2\hat{c}_1^{-1}}{n_1(n_1-1)}\sum_{k=1}^{n_1}\sum_{l\neq k}^{n_1}\left(\tilde{\mathbf{U}}_{1l}\hat{\mathbf{D}}_2^{-1/2}\hat{\mathbf{D}}_1^{1/2}\tilde{\mathbf{U}}_{1k}\right)^2 = \frac{p^2c_2c_1^{-1}}{n_1(n_1-1)}\sum_{k=1}^{p}\sum_{l\neq k}^{n_1}\left(\mathbf{U}_{1l}\mathbf{D}_2^{-1/2}\mathbf{D}_1^{1/2}\mathbf{U}_{1k}\right)^2 + M_{n1}.$$

Taking the same procedure as Lemma 5, we can obtain that $M_{n1} = o_p(\mathrm{tr}(\mathbf{A}_1^2))$. Also, by the same arguments as G_{n1},

$$E\left(\frac{p^2c_2c_1^{-1}}{n_1(n_1-1)}\sum_{k=1}^{n_1}\sum_{l\neq k}^{n_1}\left(\mathbf{U}_{1l}\mathbf{D}_2^{-1/2}\mathbf{D}_1^{1/2}\mathbf{U}_{1k}\right)^2\right) = \mathrm{tr}(\mathbf{A}_1^2)(1+o(1)).$$

Here, we only need to show that

$$\mathrm{var}\left(\frac{p^2c_2c_1^{-1}}{n_1(n_1-1)}\sum_{k=1}^{n_1}\sum_{l\neq k}^{n_1}\left(\mathbf{U}_{1l}\mathbf{D}_2^{-1/2}\mathbf{D}_1^{1/2}\mathbf{U}_{1k}\right)^2\right) = o(\mathrm{tr}^2(\mathbf{A}_1^2)).$$

By consider the possible combinations of the subscripts, it can be shown that

$$\begin{aligned}
&\mathrm{var}\left(\frac{p^2c_2c_1^{-1}}{n_1(n_1-1)}\sum_{k=1}^{n_1}\sum_{l\neq k}^{n_1}\left(\mathbf{U}_{1l}\mathbf{D}_2^{-1/2}\mathbf{D}_1^{1/2}\mathbf{U}_{1k}\right)^2\right)\\
&= \frac{2p^4c_2^2c_1^{-2}}{n_1^2(n_1-1)^2}\sum_{i=1}^{n_1}\sum_{j\neq i}^{n_1}E((\mathbf{U}_{1i}^T\mathbf{D}_2^{-1/2}\mathbf{D}_1^{1/2}\mathbf{U}_{1j})^4)\\
&\quad+\frac{4p^4c_2^2c_1^{-2}}{n_1(n_1-1)}\sum_{i=1}^{n_1}E((\mathbf{U}_{1i}^T\mathbf{D}_2^{-1/2}\mathbf{D}_1^{1/2}E(\mathbf{U}_{1l}\mathbf{U}_{1l}^T)\mathbf{D}_2^{-1/2}\mathbf{D}_1^{1/2}\mathbf{U}_{1i})^2) + o(\mathrm{tr}^2(\mathbf{A}_1^2)).
\end{aligned}$$

Taking the same procedure as in Lemma 1 and G_{n1}, we can show that

$$E(p^4c_2^2c_1^{-2}(\mathbf{U}_{1i}^T\mathbf{D}_2^{-1/2}\mathbf{D}_1^{1/2}\mathbf{U}_{1j})^4) = O(\mathrm{tr}^2(\mathbf{A}_1^2)).$$

Similarly, we have $E(p^4c_2^2c_1^{-2}(\mathbf{U}_{1i}^T\mathbf{D}_2^{-1/2}\mathbf{D}_1^{1/2}E(\mathbf{U}_{1l}\mathbf{U}_{1l}^T)\mathbf{D}_2^{-1/2}\mathbf{D}_1^{1/2}\mathbf{U}_{1i})^2) = O(\mathrm{tr}^2(\mathbf{A}_1^2))$. Thus we complete the proof. $\square$

Appendix E: Some properties of r^{-1}

When $\mathbf{X}_i \sim N_p(\mathbf{0}, \mathbf{I}_p)$, we have

$$E(r^{-1}) = \frac{1}{\sqrt{2}} \frac{\Gamma(\frac{p-1}{2})}{\Gamma(\frac{p}{2})},\ E(r^{-2}) = \frac{\Gamma(\frac{p}{2}-1)}{2\Gamma(\frac{1}{2})} = \frac{1}{p-2},\ E(r^{-3}) = \frac{\Gamma(\frac{p-3}{2})}{2\sqrt{2}\Gamma(\frac{p}{2})},$$

where $\Gamma(\cdot)$ are respectively the usual Beta and Gamma functions.

By Stirling formula

$$\lim_{x\to\infty} \frac{\Gamma(x+1)}{(x/e)^x\sqrt{2\pi x}} = 1,$$

it is straightforward to see

$$\lim_{p\to\infty} \frac{E(r^{-2})}{[E(r^{-1})]^2} = \lim_{p\to\infty} \frac{e(p-2)(p-4)}{(p-3)^2} \left[(1+\frac{1}{p-4})^{p-4}\right]^{-1} = 1,$$
$$\lim_{p\to\infty} \frac{E(r^{-3})}{[E(r^{-1})]^3} = \lim_{p\to\infty} \frac{(p-2)^2}{e(p-3)^2} \left(1+\frac{1}{p-3}\right)^{p-3} = 1.$$

Appendix F: Discussion on the test using sparse covariance matrix estimates

When $\mathbf{\Sigma}$ is a sparse covariance matrix, under regularity conditions, $\|\mathbf{\Sigma}^{-1} - \widehat{\mathbf{\Sigma}}^{-1}\| = O_p(\sqrt{\frac{\log p}{n}})$, where $\widehat{\mathbf{\Sigma}}$ is a "good" sparse covaraince estimator (e.g., Fan et al. 2011). By the lower bound derived by Cai et al. (2010), the convergence rate is minimax optimal for the sparse covariance estimation. Consider the Hotelling's T^2 test statistic $T_n^2 = \frac{n_1 n_2}{n}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)^T \widehat{\mathbf{\Sigma}}^{-1}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)$. If $\mathbf{\Sigma}$ is used in T_n^2 (denoted as $\tilde{T}_n^2$), it can be seen that $(\tilde{T}_n^2 - p)/\sqrt{2p} \xrightarrow{d} N(0,1)$ as $(n,p) \to \infty$ under H_0. In order to make this asymptotic applicable for T_n^2, one needs to show that

$$\frac{n}{\sqrt{p}}(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)^T(\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1})(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2) = o_p(1).$$

However, $\|\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2\|^2 = O_p(p/n)$ under H_0. By the Cauchy-Schwartz inequality, we have $\frac{n}{\sqrt{p}}|(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)^T(\widehat{\mathbf{\Sigma}}^{-1} - \mathbf{\Sigma}^{-1})(\bar{\mathbf{X}}_1 - \bar{\mathbf{X}}_2)| = O_p(\sqrt{\frac{p\log p}{n}})$. We see that it requires $p \log p = o(n)$, which is basically a low-dimensional scenario. The above simple derivation uses, however, a Cauchy-Schwartz bound, which is too crude for a large p. More elaborate analysis may be available but requires further regularity conditions and other more technical arguments.

References

Bai, Z. and Saranadasa, H. (1996), "Effect of High Dimension: by an Example of a Two Sample Problem," *Statistica Sinica,* 6, 311–329.

Cai, T., Zhang, C. and Zhou, H. (2010), "Optimal rates of convergence for covariance matrix estimation," *Annals of Statistics*, 38, 2118–2144.

Chen, S. X. and Qin, Y-L. (2010), "A Two-Sample Test for High-Dimensional Data with Applications to Gene-Set Testing," *The Annals of Statistics,* 38, 808–835.

Fan, J., Liao, Y. and Mincheva, M. (2011), "High dimensional covariance matrix estimation in approximate factor models," *Annals of Statistics*, 39, 3320–3356.

Hettmansperger, T. P. and Randles, R. H. (2002), "A Practical Affine Equivariant Multivariate Median," *Biometrika*, 89, 851–860.