KOO APPROACH FOR SCALABLE VARIABLE SELECTION PROBLEM IN LARGE-DIMENSIONAL REGRESSION

Zhidong Bai[1], Kwok Pui Choi[2], Yasunori Fujikoshi[3] and Jiang Hu[1]

[1]*KLASMOE and School of Mathematics & Statistics, Northeast Normal University, China,*

[2]*Department of Statistics and Data Science, National University of Singapore, Singapore.*

[3]*Department of Mathematics, Graduate School of Science, Hiroshima University, Japan*

Supplementary Material

This supplementary material includes additional simulation studies, additional real data analysis and proofs of the main theorems for general error distributions using random matrix theory.

S1 Additional simulation results

The simulation results for Settings I and II, and six cases of distribution of $\mathbf{E}$ are tabulated in Tables 1–6.

$\alpha = 0.2,\ c = 0.2$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	79	0	15	0	0	589	0	0	0
T-S	198	921	228	983	966	570	411	655	999	954
O-S	802	0	772	2	34	430	0	345	1	46
A-S	**2.05**	–	1.97	1	1	1.37	–	1.27	1	1.04
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	993	0	0	0	0	1000	0	0	0
T-S	956	7	972	1000	958	1000	0	1000	999	951
O-S	44	0	28	0	42	0	0	0	1	49
A-S	1.02	–	1	–	1.02	–	–	–	1	1.04
$\alpha = 0.2,\ c = 0.4$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	938	0	640	19	0	1000	0	0	0
T-S	35	62	0	360	940	2	0	0	1000	953
O-S	965	0	1000	0	41	998	0	1000	0	47
A-S	3.69	–	7.06	–	1.05	6.86	–	46.30	–	1.04
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	23	0	0	998	957	505	0	0	1000	961
O-S	977	0	1000	2	43	495	0	1000	0	39
A-S	3.92	–	95.28	1	1.05	1.47	–	194.82	–	1
$\alpha = 0.4,\ c = 0.2$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	42	0	828	41	0	129	0	0	0
T-S	0	923	3	172	919	0	871	0	998	965
O-S	1000	35	997	0	40	1000	0	1000	2	35
A-S	16.50	1.09	6.67	–	1.12	100.87	–	8	1	1
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	729	0	0	0	0	999	0	0	0
T-S	0	271	41	1000	954	0	1	748	995	940
O-S	1000	0	959	0	46	1000	0	252	5	60
A-S	213.52	–	3.28	–	1.02	450.55	–	1.16	1	1
$\alpha = 0.4,\ c = 0.4$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	623	0	999	889	0	1000	0	10	0
T-S	0	294	0	1	103	0	0	0	990	963
O-S	1000	83	1000	0	8	1000	0	1000	0	37
A-S	31.05	1.51	29.35	–	1.25	194.59	–	193.17	–	1.05
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	0	0	0	998	952	0	0	0	996	930
O-S	1000	0	1000	2	48	1000	0	1000	4	70
A-S	394.99	–	394.85	1	1.06	795	–	795	1	1.01

Table 1: Selection times of the KOO methods with AIC, BIC, C_p thresholds and bootstrap methods under Settings (I) and (i) based on 1,000 replications.

$\alpha = 0.2$, $c = 0.2$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	1	109	1	45	10	0	307	0	0	0
T-S	198	891	219	955	958	563	693	646	1000	961
O-S	801	0	780	0	32	437	0	354	0	39
A-S	2.09	–	1.96	–	1.03	1.35	–	1.25	–	1.05
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	780	0	0	0	0	1000	0	0	0
T-S	963	220	978	999	946	1000	0	1000	999	953
O-S	37	0	22	1	54	0	0	0	1	47
A-S	1.03	–	1	1	1.02	–	–	–	1	1.04

$\alpha = 0.2$, $c = 0.4$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	2	734	0	487	41	0	1000	0	0	0
T-S	40	266	0	513	937	1	0	0	1000	945
O-S	958	0	1000	0	22	999	0	1000	0	55
A-S	3.63	–	6.99	–	1.05	6.59	–	45.85	–	1.02
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	26	0	0	1000	962	490	0	0	999	943
O-S	974	0	1000	0	38	510	0	1000	1	57
A-S	3.95	–	95.28	–	1	1.46	–	194.80	1	1.02

$\alpha = 0.4$, $c = 0.2$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	56	2	245	53	0	113	0	0	0
T-S	0	917	2	755	913	0	887	0	999	958
O-S	1000	27	996	0	34	1000	0	1000	1	42
A-S	16.56	1.04	6.79	–	1.03	102.01	–	8.19	1	1
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	465	0	0	0	0	902	0	0	0
T-S	0	535	54	1000	940	0	98	780	999	959
O-S	1000	0	946	0	60	1000	0	220	1	41
A-S	213.34	–	3.27	–	1.02	449.91	–	1.11	1	1.05

$\alpha = 0.4$, $c = 0.4$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	465	1	984	744	0	996	0	17	3
T-S	0	420	0	16	240	0	4	0	983	937
O-S	1000	115	999	0	16	1000	0	1000	0	60
A-S	30.98	1.39	29.18	–	1.06	194.59	–	193.12	–	1
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	3	1	0	1000	0	0	0
T-S	0	0	0	996	949	0	0	0	999	950
O-S	1000	0	1000	1	50	1000	0	1000	1	50
A-S	394.99	–	394.86	1	1	795	–	795	1	1

Table 2: Selection times of the KOO methods with AIC, BIC, C_p thresholds and bootstrap methods under Settings (I) and (ii) based on 1,000 replications.

$\alpha = 0.2,\ c = 0.2$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	120	0	25	0	0	543	0	0	0
T-S	182	880	204	974	969	572	457	663	999	955
O-S	818	0	796	1	31	428	0	337	1	45
A-S	2.10	–	2.02	1	1.10	1.35	–	1.23	1	1.04
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	975	0	0	0	0	1000	0	0	0
T-S	963	25	981	1000	963	999	0	1000	999	953
O-S	37	0	19	0	37	1	0	0	1	47
A-S	1	–	1	–	1	1	–	–	1	1.02

$\alpha = 0.2,\ c = 0.4$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	895	0	219	24	0	1000	0	0	0
T-S	37	105	1	778	932	4	0	0	999	941
O-S	963	0	999	3	44	996	0	1000	1	59
A-S	3.68	–	7.11	1	1.07	6.60	–	46.14	1	1.02
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	35	0	0	1000	937	454	0	0	1000	941
O-S	965	0	1000	0	63	546	0	1000	0	59
A-S	3.90	–	95.25	–	1.05	1.46	–	194.82	–	1

$\alpha = 0.4,\ c = 0.2$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	61	0	331	43	0	153	0	0	0
T-S	0	894	5	666	897	0	847	0	999	969
O-S	1000	45	995	3	60	1000	0	1000	1	31
A-S	16.59	1.02	6.91	1	1.08	101.70	–	8.23	1	1
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	715	0	0	0	0	1000	0	0	0
T-S	0	285	44	1000	947	0	0	788	999	963
O-S	1000	0	956	0	53	1000	0	212	1	37
A-S	213.18	–	3.22	–	1.04	449.63	–	1.13	1	1

$\alpha = 0.4,\ c = 0.4$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	589	0	994	837	0	1000	0	2	0
T-S	0	319	0	6	150	0	0	0	995	950
O-S	1000	92	1000	0	13	1000	0	1000	3	50
A-S	30.97	1.34	29.24	–	1.38	194.60	–	193.15	1	1.02
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	0	0	0	1000	950	0	0	0	994	945
O-S	1000	0	1000	0	50	1000	0	1000	6	55
A-S	394.99	–	394.83	–	1.06	795	–	795	1	1.05

Table 3: Selection times of the KOO methods with AIC, BIC, C_p thresholds and bootstrap methods under Settings (I) and (iii) based on 1,000 replications.

$\alpha = 0.2$, $c = 0.2$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	77	0	971	215	0	648	0	0	0
T-S	41	847	46	29	753	0	352	1	1000	967
O-S	959	76	954	0	32	1000	0	999	0	33
A-S	3.11	1.04	3.01	–	1	7.40	–	6.68	–	1.03
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	993	0	0	0	0	1000	0	0	0
T-S	4	7	6	1000	944	177	0	267	997	942
O-S	996	0	994	0	56	823	0	733	3	58
A-S	5.51	–	4.65	–	1.04	2.11	–	1.77	1	1.02

$\alpha = 0.2$, $c = 0.4$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	925	0	991	622	0	1000	0	2	0
T-S	2	74	0	9	361	0	0	0	997	934
O-S	998	1	1000	0	17	1000	0	1000	1	66
A-S	4.74	1	7	–	1.06	16.40	–	45.88	1	1
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	0	0	0	1000	938	0	0	0	999	924
O-S	1000	0	1000	0	62	1000	0	1000	1	76
A-S	18.74	–	95.36	–	1.03	13.77	–	194.65	1	1.01

$\alpha = 0.4$, $c = 0.2$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	4	0	999	597	0	7	0	7	0
T-S	0	348	0	1	386	0	993	0	993	961
O-S	1000	648	1000	0	17	1000	0	1000	0	39
A-S	15.31	1.67	9.23	–	1	94.64	–	28.61	–	1
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	27	0	0	0	0	354	0	0	0
T-S	0	973	0	1000	939	0	646	0	1000	956
O-S	1000	0	1000	0	61	1000	0	1000	0	44
A-S	198.92	–	31.12	–	1.03	417.94	–	21.32	–	1.02

$\alpha = 0.4$, $c = 0.4$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	243	0	1000	898	0	988	0	188	2
T-S	0	233	0	0	97	0	12	0	812	965
O-S	1000	524	1000	0	5	1000	0	1000	0	33
A-S	28.58	1.99	26.89	–	1.20	191.27	–	186.32	–	1.03
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	0	0	0	1000	942	0	0	0	1000	928
O-S	1000	0	1000	0	58	1000	0	1000	0	72
A-S	394.33	–	391.74	–	1.05	794.99	–	794.74	–	1.04

Table 4: Selection times of the KOO methods with AIC, BIC, C_p thresholds and bootstrap methods under Settings (II) and (iv) based on 1,000 replications.

$\alpha = 0.2$, $c = 0.2$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	63	0	66	0	0	732	0	0	0
T-S	127	936	144	933	963	169	268	245	998	969
O-S	873	1	856	1	37	831	0	755	2	31
A-S	2.38	1	2.27	1	1	2.16	–	1.92	1	1
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	688	0	779	994	949	994	0	998	1000	937
O-S	312	0	221	6	51	6	0	2	0	63
A-S	1.16	–	1.11	1	1.02	1	–	1	–	1.05

$\alpha = 0.2$, $c = 0.4$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	973	0	615	100	0	1000	0	0	0
T-S	21	27	2	385	863	3	0	0	999	948
O-S	979	0	998	0	37	997	0	1000	1	52
A-S	3.86	–	6.98	–	1.05	8.99	–	46.12	1	1.02
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	2	0	0	999	952	132	0	0	998	947
O-S	998	0	1000	1	48	868	0	1000	2	53
A-S	6.64	–	95.24	1	1.02	2.30	–	193.88	1	1.06

$\alpha = 0.4$, $c = 0.2$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1	0	384	17	0	1	0	0	0
T-S	0	841	1	615	939	0	999	0	1000	944
O-S	1000	158	999	1	44	1000	0	1000	0	56
A-S	15.86	1.09	7.40	1	1.07	98.76	–	13.10	–	1.02
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	6	0	0	0	0	374	0	0	0
T-S	0	994	0	1000	954	0	626	197	999	954
O-S	1000	0	1000	0	46	1000	0	803	1	46
A-S	208.66	–	7.73	–	1.02	438.75	–	2.01	1	1.02

$\alpha = 0.4$, $c = 0.4$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	263	0	969	684	0	994	0	1	0
T-S	0	540	0	31	302	0	6	0	999	948
O-S	1000	197	1000	0	14	1000	0	1000	0	52
A-S	30.52	1.38	28.73	–	1.07	194.24	–	192.21	–	1.04
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	0	0	0	999	953	0	0	0	999	950
O-S	1000	0	1000	1	47	1000	0	1000	1	50
A-S	394.98	–	394.64	1	1.02	795	–	795	1	1.02

Table 5: Selection times of the KOO methods with AIC, BIC, C_p thresholds and bootstrap methods under Settings (II) and (v) based on 1,000 replications.

$\alpha = 0.2,\ c = 0.2$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	61	0	0	0	0	788	0	0	0
T-S	483	939	534	999	974	960	212	977	1000	956
O-S	517	0	466	1	26	40	0	23	0	44
A-S	1.49	–	1.44	1	1.08	1	–	1	–	1
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	1000	0	1000	1000	937	1000	0	1000	999	951
O-S	0	0	0	0	63	0	0	0	1	49
A-S	–	–	–	–	1.10	–	–	–	1	1.02

$\alpha = 0.2,\ c = 0.4$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	987	0	161	13	0	1000	0	0	0
T-S	60	13	0	838	951	28	0	0	1000	958
O-S	940	0	1000	1	36	972	0	1000	0	42
A-S	3.01	–	6.83	1	1.06	3.94	–	45.52	–	1.05
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	237	0	0	999	949	870	0	0	999	954
O-S	763	0	1000	1	51	130	0	1000	1	46
A-S	1.92	–	94.99	1	1	1.05	–	193.56	1	1

$\alpha = 0.4,\ c = 0.2$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	0	0	3	0	0	0	0	0	0
T-S	0	995	17	996	941	0	1000	82	998	943
O-S	1000	5	983	1	59	1000	0	918	2	57
A-S	16.54	1	4.84	1	1.07	102.68	–	2.96	1	1.05
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	3	0	0	0	0	345	0	0	0
T-S	0	997	655	1000	946	0	655	992	1000	951
O-S	1000	0	345	0	54	1000	0	8	0	49
A-S	217.91	–	1.17	–	1.07	465.89	–	1	–	1.06

$\alpha = 0.4,\ c = 0.4$										
	$n = 100$					$n = 500$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	237	0	961	464	0	997	0	0	0
T-S	0	658	0	39	505	0	3	0	1000	952
O-S	1000	105	1000	0	31	1000	0	1000	0	48
A-S	32.02	1.32	30.24	–	1.13	194.90	–	194.26	–	1.02
	$n = 1000$					$n = 2000$				
	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$	$\hat{\mathbf{j}}_*^A$	$\hat{\mathbf{j}}_*^B$	$\hat{\mathbf{j}}_*^C$	$\hat{\mathbf{j}}_*^{(0)}$	$\hat{\mathbf{j}}_*^{(5)}$
U-S	0	1000	0	0	0	0	1000	0	0	0
T-S	0	0	0	999	937	0	0	0	999	939
O-S	1000	0	1000	1	63	1000	0	1000	1	61
A-S	395	–	394.98	1	1.02	795	–	795	1	1.05

Table 6: Selection times of the KOO methods with AIC, BIC, C_p thresholds and bootstrap methods under Settings (II) and (vi) based on 1,000 replications.

S2 Additional real data analysis

The second example is chemometrics data taken from Skagerberg et al. (1992) (we replaced the value 19203 with 1.9203 in the 37th observation). The data are taken from a simulation of a low-density tubular polyethylene reactor studying the relationship between polymer properties and the process. The predictor variables consist of 20 temperatures measured at equal distances along the low-density polyethylene reactor section, together with the wall temperature of the reactor and the solvent feed rate. The responses are the output characteristics of the polymers, including two molecular weights, two branching frequencies and the contents of two groups. This data set has been studied by Breiman and Friedman (1997) and Similä and Tikka (2007). Similar to Breiman and Friedman (1997), we log-transformed the response values because they are highly skewed to the right. In total, there are $n = 56$ observations with $k = 22$ predictor variables and $p = 6$ responses.

We present the scatterplot of $\{\mathcal{K}_j\}$ in descending order in Figure 1. We also indicate the critical values of KAIC, KBIC and KCp, and $\hat{K}_0$, $\hat{K}_{0.05}$ estimated by Algorithm 1 with standard normal distribution and $N = 1,000$. Since the dimension is relatively small, we recommend using a larger significance level ν to prevent under-specifying. It seems that the variables

$\{22, 3, 4\}$ are significant and variables $\{21, 11\}$ are potentially significant too. KAIC and KCp, however, select many more variables, which are likely to be spurious.

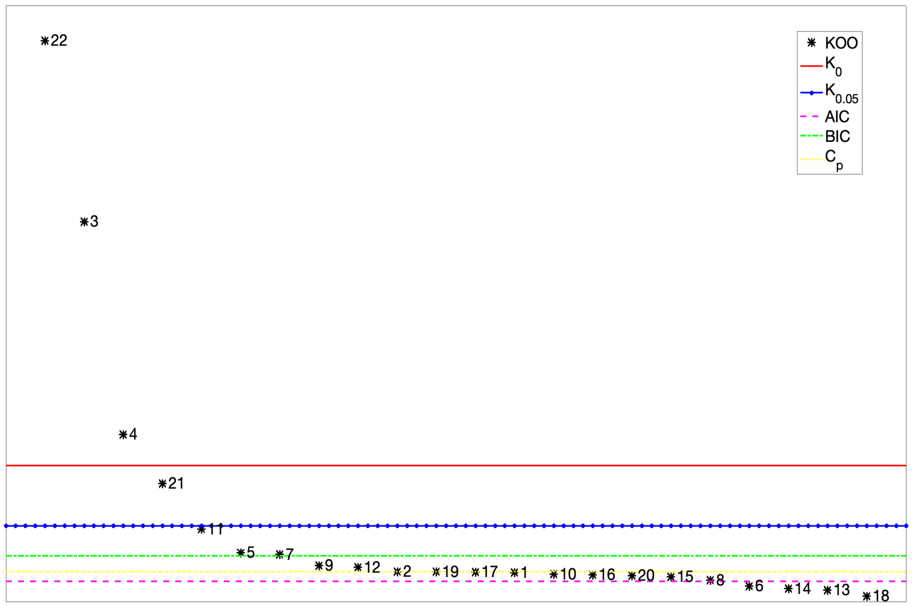

Figure 1: Scatterplots for the chemometrics dataset.

S3 Proofs of Theorems 1-4

In this appendix, we present the proofs of Theorems 1–4 under general distributions by random matrix theory. Before that, we first give some notation and preliminary results which will be used in the sequel frequently. For simplicity, we denote $\mathbf{M} = p^{-1}\mathbf{E}'\mathbf{Q}\mathbf{E}$ and $\mathbf{M}_l = \frac{1}{p}\mathbf{E}_l'\mathbf{Q}\mathbf{E}_l$, where $\mathbf{E}_l$ is the $n \times (p-1)$ submatrix of $\mathbf{E}$ with the l-th column removed. Denote by $\mathbb{E}_l$ the conditional expectation given $\{\mathbf{e}_1, \ldots, \mathbf{e}_l\}$ and by $\mathbb{E}_0 = \mathbb{E}$ the unconditional expectation, where $\mathbf{e}_i$ is the n-vector of the i-th column of $\mathbf{E}$. Let $\mathbf{b} = \mathbf{\Sigma}^{-1/2}\mathbf{\Theta}_*\mathbf{X}_*'\mathbf{a}_1$ and $\mathbf{b}_l$ be the $p-1$ sub-vector of $\mathbf{b}$ with the l-th entry b_l removed. Then we have

$$\mathbf{a}_1'\mathbf{Y}\mathbf{\Sigma}^{-1/2}(\mathbf{E}'\mathbf{Q}\mathbf{E})^{-1}\mathbf{\Sigma}^{-1/2}\mathbf{Y}'\mathbf{a}_1 = p^{-1}(\mathbf{b}' + \mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1 + \mathbf{b}).$$

Modifying the truncation argument of Bai et al. (2018), we can assume that the variables $\{e_{ij}, i = 1 \ldots n, j = 1 \ldots p\}$ satisfy the following additional condition:

$$|e_{ij}| < \eta_n \sqrt{n}, \quad \text{for all } i, j, \tag{S3.1}$$

where $\eta_n \to 0$ slowly enough. By the theorem in the appendix of Bai and Silverstein (2004), we know for any positive constant $d < (1-\sqrt{c})^2$ and any given $t > 0$, $\lambda_{\min}^{\frac{1}{n}\mathbf{E}'\mathbf{E}} \xrightarrow{a.s.} (1-\sqrt{c})^2$ and

$$\mathbb{P}(\lambda_{\min}^{\frac{1}{n}\mathbf{E}'\mathbf{E}} < d) = o(n^{-t}).$$

Moreover, by Theorem 1.2 in Bai and Silverstein (1999), we conclude that for any positive constant $d < (1-\sqrt{c/(1-\alpha)})^2$ and any given $t > 0$, $\lambda_{\min}^{\frac{1}{n}\mathbf{E}'\mathbf{Q}\mathbf{E}} \xrightarrow{a.s.} (1-\sqrt{c/(1-\alpha)})^2$ and

$$\mathbb{P}(\lambda_{\min}^{\frac{1}{n}\mathbf{E}'\mathbf{Q}\mathbf{E}} < d) = o(n^{-t}).$$

Denote

$$\beta_l = \frac{1}{p}\mathbf{e}_l'\mathbf{Q}\mathbf{e}_l - \frac{1}{p^2}\mathbf{e}_l'\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l$$

and

$$\beta_1^{tr} = \mathrm{tr}[\frac{1}{p}\mathbf{Q} - \frac{1}{p^2}\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}] = \frac{n-k-p+1}{p}.$$

It follows that

$$\frac{1}{\beta_l} = \frac{1}{\beta_1^{tr}} - \frac{\xi_l}{\beta_l\beta_1^{tr}}, \tag{S3.2}$$

where $\xi_l = \beta_l - \beta_1^{tr}$. By Lemma 7.2 in Bai and Yao (2005) (see Lemma 4), we have that for any $2 \leq \ell \leq \log(n)$,

$$\mathbb{E}|\xi_l|^\ell = O(p^{-1}\eta_n^{2\ell-4}), \tag{S3.3}$$

which indicate that ξ_l tends to 0 in probability with order of $o(n^{-t})$ for any $t > 0$. Analogously, for application later, together with the condition that $p^{-1/2}\mathbf{b}$ is bounded in Euclidean norm, we conclude that for $2 \leq \ell \leq \log(n)$,

$$\max\{\mathbb{E}|p^{-1}\mathbf{b}_l'\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l|^\ell, \mathbb{E}|p^{-1}\mathbf{a}_l\mathbf{E}_l'\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l|^\ell\} = O(p^{\ell/2-1}\eta_n^{\ell-2}) \tag{S3.4}$$

and

$$\max\{\mathbb{E}|p^{-2}\mathbf{b}_l'\mathbf{\Psi}_l\mathbf{b}_l|^\ell, \mathbb{E}|p^{-2}\mathbf{a}_1\mathbf{E}_l'\mathbf{\Psi}_l\mathbf{b}_l|^\ell, \mathbb{E}|p^{-2}\mathbf{a}_1\mathbf{E}_l'\mathbf{\Psi}_l\mathbf{E}_l\mathbf{a}_1|^\ell\} = O(p^{\ell-2}\eta_n^{2\ell-4}) \tag{S3.5}$$

where

$$\mathbf{\Psi}_l = \mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l\mathbf{e}_l'\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1} - \mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}.$$

As we only need to prove the weak convergence conclusion and $\beta_1^{tr} \to (1-\alpha-c)/c > 0$, thus throughout the proofs, we can safely assume $\|\mathbf{M}^{-1}\|$, $\|\mathbf{M}_l^{-1}\|$ and $|1/\beta_l|$ are all bounded for large n.

S3.1 Proof of Theorem 1

Theorem 1 can be obtained from Proposition 3.1 in Bai et al. (2022) with letting $z \downarrow 0$ directly. That is, for any non-random vectors $\mathbf{r}_1$, $\mathbf{r}_2$ $\mathbf{r}_3$ and $\mathbf{r}_4$ with suitable dimensions and bounded in Euclidean norm, under conditions in Theorem 1, we have that for any $t > 0$ and $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\mathbf{r}_1'\mathbf{M}^{-1}\mathbf{r}_2 - \frac{c_n\mathbf{r}_1'\mathbf{r}_2}{1-c_n-\alpha_n}\right| \geq \varepsilon\right) = o(n^{-t}), \tag{S3.6}$$

$$\mathbb{P}\left(\left|\frac{1}{\sqrt{p}}\mathbf{r}_1'\mathbf{M}^{-1}\mathbf{E}'\mathbf{r}_3\right| \geq \varepsilon\right) = o(n^{-t}), \tag{S3.7}$$

and

$$\mathbb{P}\left(\left|\frac{1}{p}\mathbf{r}_3'\mathbf{E}\mathbf{M}^{-1}\mathbf{E}'\mathbf{r}_4 - \frac{c_n\mathbf{r}_3'\mathbf{r}_4}{1-c_n-\alpha_n} + \frac{c_n^2\mathbf{r}_3'\mathbf{Q}\mathbf{r}_4}{(1-c_n-\alpha_n)(1-\alpha_n)}\right| \geq \varepsilon\right) = o(n^{-t}). \tag{S3.8}$$

Then the proof of Theorem 1 is complete.

S3.2 Proof of Theorem 2

For simple presentation, in the following we assume $\{1,\ldots,q\}\subset[k]\backslash\mathbf{j}_*$ and $j_i=i$. To prove Theorem 2, it is sufficient to show that for any non-null vector $\mathbf{h}=(h_1,\ldots,h_q)'$, $\sqrt{p}[(\mathcal{K}_1,\ldots,\mathcal{K}_q)\mathbf{h}-\frac{c_n}{1-c_n-\alpha_n}\mathbf{1}_q'\mathbf{h}]$ converges weakly to a normal distribution with mean zero and variance $\frac{c^2}{(1-\alpha_n-c_n)^2}[\frac{2(1-\alpha_n)}{(1-\alpha_n-c_n)}\mathbf{h}'(\boldsymbol{\mathcal{A}}_q'\boldsymbol{\mathcal{A}}_q)^2\mathbf{h}+\tau\mathbf{h}'(\boldsymbol{\mathcal{A}}_q\circ\boldsymbol{\mathcal{A}}_q)'(\boldsymbol{\mathcal{A}}_q\circ\boldsymbol{\mathcal{A}}_q)\mathbf{h}]$, where $\boldsymbol{\mathcal{A}}_q=(\mathbf{a}_1,\ldots,\mathbf{a}_q)$.

We split the proof of this theorem into two parts. First, we show the asymptotic normality of the sequence of random variables

$$\mathcal{M}_1^{(n)}:=\sqrt{p}[(\mathcal{K}_1,\ldots,\mathcal{K}_q)\mathbf{h}-\mathbb{E}(\mathcal{K}_1,\ldots,\mathcal{K}_q)\mathbf{h}].$$

Second, we prove the non-random sequence

$$\mathcal{M}_2^{(n)}=\sqrt{p}[\mathbb{E}(\mathcal{K}_1,\ldots,\mathcal{K}_q)\mathbf{h}-\frac{c_n\mathbf{1}_q'\mathbf{h}}{1-c_n-\alpha_n}]$$

tends to zero. Note that for notational simplicity the superscript $^{(n)}$ in $\mathcal{M}_1^{(n)}$ and $\mathcal{M}_2^{(n)}$ are suppressed in the sequel.

We start to consider $\mathcal{M}_1$. Let $\boldsymbol{\mathcal{H}}=Diag(h_1,\ldots,h_q)$. It follows that

$$\begin{aligned}\mathcal{M}_1=&p^{-1/2}\sum_{l=1}^{p}(\mathbb{E}_l-\mathbb{E}_{l-1})\mathrm{tr}(\mathbf{E}\mathbf{M}^{-1}\mathbf{E}'\boldsymbol{\mathcal{A}}_q\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_q')\\=&p^{-1/2}\sum_{l=1}^{p}(\mathbb{E}_l-\mathbb{E}_{l-1})\mathrm{tr}[(\mathbf{E}\mathbf{M}^{-1}\mathbf{E}'-\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l')\boldsymbol{\mathcal{A}}_q\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_q'].\end{aligned}$$

By the inversion formula of block matrix, we obtain

$$\mathbf{E}\mathbf{M}^{-1}\mathbf{E}' - \mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l' = \frac{1}{\beta_l p^2}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l\mathbf{e}_l'\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'$$
$$- \frac{1}{\beta_l p}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l\mathbf{e}_l' - \frac{1}{\beta_l p}\mathbf{e}_l\mathbf{e}_l'\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l' + \frac{\mathbf{e}_l\mathbf{e}_l'}{\beta_l}. \tag{S3.9}$$

Then, by the equation (S3.2), we can rewrite $\mathcal{M}_1$ as

$$\mathcal{M}_1 = \frac{1}{p^{1/2}\beta_1^{tr}}\sum_{l=1}^{p}\mathbb{E}_l(\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \mathrm{tr}\mathbf{\Gamma}_l) - \frac{1}{p^{1/2}(\beta_1^{tr})^2}\sum_{l=1}^{p}\mathbb{E}_l(\xi_l\mathrm{tr}\mathbf{\Gamma}_l) + \mathcal{M}_{10},$$

where

$$\mathbf{\Gamma}_l = p^{-2}\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\boldsymbol{\mathcal{A}}_q\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_q'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q} - p^{-1}\boldsymbol{\mathcal{A}}_q\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_q'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}$$
$$- p^{-1}\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l\boldsymbol{\mathcal{A}}_q\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_q' + \boldsymbol{\mathcal{A}}_q\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_q'$$

and

$$\mathcal{M}_{10} = -\sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})\frac{\xi_l(\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \mathrm{tr}\mathbf{\Gamma}_l)}{p^{1/2}(\beta_1^{tr})^2} + \sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})\frac{\xi_l^2\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l}{p^{1/2}\beta_l(\beta_1^{tr})^2}.$$

It follows from (S3.3) that

$$\mathbb{E}|\frac{1}{p^{1/2}}\sum_{l=1}^{p}\mathbb{E}_l(\xi_l\mathrm{tr}\mathbf{\Gamma}_l)|^2 = \frac{1}{p}\sum_{l=1}^{p}\mathbb{E}|\mathbb{E}_l(\xi_l\mathrm{tr}\mathbf{\Gamma}_l)|^2 = O(p^{-1}).$$

By (S3.4), (S3.5) and the BurkHölder's inequality (see Lemma 2) we have that $\mathcal{M}_{10} = o_p(1)$. Applying Lemma 2.7 in Bai and Silverstein (1998) (see Lemma 3), we have that

$$\mathbb{E}|\mathbb{E}_l(\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \mathrm{tr}\mathbf{\Gamma}_l)|^4 \le \mathbb{E}|\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \mathrm{tr}\mathbf{\Gamma}_l|^4 = O(p\eta_n^4)$$

which verifies the condition (ii) in Lemma 1. Thus, what we need is to obtain the limit of

$$\frac{1}{p(\beta_1^{tr})^2}\sum_{l=1}^{p}\mathbb{E}_{l-1}\{\mathbb{E}_l[\mathbf{e}_l'\boldsymbol{\Gamma}_l\mathbf{e}_l-\mathrm{tr}\boldsymbol{\Gamma}_l]\}^2.$$

By Lemma 5, we have that

$$\mathbb{E}_{l-1}\{\mathbb{E}_l[\mathbf{e}_l'\boldsymbol{\Gamma}_l\mathbf{e}_l-\mathrm{tr}\boldsymbol{\Gamma}_l]\}^2=2\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l\boldsymbol{\Gamma}_l\mathbb{E}_l\boldsymbol{\Gamma}_l)+\tau\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l\boldsymbol{\Gamma}_l\circ\mathbb{E}_l\boldsymbol{\Gamma}_l),$$

where $\circ$ stands for the Hadamard product. Notice that

$$\mathrm{tr}(\mathbb{E}_l\boldsymbol{\Gamma}_l\mathbb{E}_l\boldsymbol{\Gamma}_l)=p^{-4}\mathrm{tr}(\mathbb{E}_l\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\boldsymbol{\mathcal{A}}_q\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_q'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})^2 \tag{S3.10}$$

$$+p^{-2}2\mathrm{tr}[\mathbb{E}_l(\boldsymbol{\mathcal{A}}_q\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_q'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})\mathbb{E}_l(\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l\boldsymbol{\mathcal{A}}_q\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_q')] \tag{S3.11}$$

$$+\mathrm{tr}(\boldsymbol{\mathcal{A}}_q\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_q')^2.$$

Let $\widetilde{\mathbf{E}}_l$ be $\mathbf{E}_l$ by replacing $\{\mathbf{e}_{l+1},\ldots,\mathbf{e}_p\}$ with $\{\tilde{\mathbf{e}}_{l+1},\ldots,\tilde{\mathbf{e}}_p\}$, where $\{\tilde{\mathbf{e}}_i\}$ are i.i.d. copies of $\mathbf{e}_1$. We define $\widetilde{\mathbf{M}}_l=\frac{1}{p}\widetilde{\mathbf{E}}_l'\mathbf{Q}\widetilde{\mathbf{E}}_l$, correspondingly. As $\boldsymbol{\mathcal{H}}$ is a diagonal matrix, thus we have that

$$\begin{aligned}&\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\boldsymbol{\mathcal{A}}_q\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_q'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})^2\\=&\mathbb{E}_l\mathrm{tr}[\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_q'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\widetilde{\boldsymbol{\Xi}}_l\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\boldsymbol{\mathcal{A}}_q]\\=&\mathbb{E}_l\sum_{i=1}^{q}h_i\mathbf{a}_i'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\widetilde{\boldsymbol{\Xi}}_l\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{a}_i,\end{aligned}$$

where $\widetilde{\widetilde{\boldsymbol{\Xi}}}_l = \mathbf{Q}\widetilde{\mathbf{E}}_l\widetilde{\mathbf{M}}_l^{-1}\widetilde{\mathbf{E}}_l'\boldsymbol{\mathcal{A}}_q\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_q'\widetilde{\mathbf{E}}_l\widetilde{\mathbf{M}}_l^{-1}\widetilde{\mathbf{E}}_l'\mathbf{Q}$. By applying the inversion formula of block matrix to $\mathbf{M}_l$, similar to (S3.9), we have that

$$\begin{aligned}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l' - \mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}' &= \frac{1}{\beta_{lp}p^2}\mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}'\mathbf{Q}\mathbf{e}_p\mathbf{e}_p'\mathbf{Q}\mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}' \\ - \frac{1}{\beta_{lp}p}\mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}'\mathbf{Q}\mathbf{e}_p\mathbf{e}_p' &- \frac{1}{\beta_{lp}p}\mathbf{e}_p\mathbf{e}_p'\mathbf{Q}\mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}' + \frac{\mathbf{e}_p\mathbf{e}_p'}{\beta_{lp}}, \end{aligned} \tag{S3.12}$$

where $\mathbf{E}_{li}$ is the $n \times (i-2)$ submatrix of $\mathbf{E}$ with the columns $\{\mathbf{e}_l, \mathbf{e}_i, \ldots, \mathbf{e}_p\}$ removed, $\mathbf{M}_{li} = \frac{1}{p}\mathbf{E}_{li}'\mathbf{Q}\mathbf{E}_{li}$ and

$$\beta_{li} = \frac{1}{p}\mathbf{e}_i'\mathbf{Q}\mathbf{e}_i - \frac{1}{p^2}\mathbf{e}_i'\mathbf{Q}\mathbf{E}_{li}\mathbf{M}_{li}^{-1}\mathbf{E}_{li}'\mathbf{Q}\mathbf{e}_i. \tag{S3.13}$$

Denote

$$\beta_i^{tr} = \mathrm{tr}[\frac{1}{p}\mathbf{Q} - \frac{1}{p^2}\mathbf{Q}\mathbf{E}_{li}\mathbf{M}_{li}^{-1}\mathbf{E}_{li}'\mathbf{Q}] = \frac{n-k-p+i}{p}$$

and

$$\xi_{li} = \beta_{li} - \beta_i^{tr}.$$

We can easily check that the orders of (S3.3)–(S3.5) hold for replacing the subscripts l by li. Thus, analogous to the above discussion, we have that

$$\begin{aligned} &\mathbb{E}_l\mathbf{a}_i'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\widetilde{\widetilde{\boldsymbol{\Xi}}}_l\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{a}_i \\ =&\mathbb{E}_l[\mathbf{a}_i'\mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}'\widetilde{\widetilde{\boldsymbol{\Xi}}}_l\mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}'\mathbf{a}_i] + o_p(p^3) \\ =&\mathbb{E}_l[\mathbf{a}_i'\mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}'\widetilde{\widetilde{\boldsymbol{\Xi}}}_{lp}\mathbf{E}_{lp}\mathbf{M}_{lp}^{-1}\mathbf{E}_{lp}'\mathbf{a}_i] + o_p(p^3), \end{aligned}$$

where $\widetilde{\widetilde{\boldsymbol{\Xi}}}_{lp}$ is defined by removing b_p and $\tilde{\mathbf{e}}_p$ from $\widetilde{\widetilde{\boldsymbol{\Xi}}}_l$. We then repeat the procedure that remove b_i, $\mathbf{e}_i$ and $\tilde{\mathbf{e}}_i$, $i = l+1, \ldots, p-1$ from $\boldsymbol{\Xi}_{lp}$ and $\widetilde{\widetilde{\boldsymbol{\Xi}}}_{lp}$, respectively. Then applying Proposition 3.1 in (Bai et al., 2022), we finally obtain that

$$
\begin{aligned}
\text{(S3.18)} =& \sum_{i,j}^{q} h_i h_j p^{-2} [\mathbf{a}_i' \mathbf{E}_{l(l-1)} \mathbf{M}_{l(l-1)}^{-1} \mathbf{E}_{l(l-1)}' \mathbf{a}_j)]^2 + o_p(1) \\
=& \frac{(l-1)^2}{(n-k-l+1)^2} \mathbf{h}' (\boldsymbol{\mathcal{A}}_q' \boldsymbol{\mathcal{A}}_q)^2 \mathbf{h} + o_p(1). \qquad \text{(S3.14)}
\end{aligned}
$$

Analogously, we have that

$$
\begin{aligned}
\text{(S3.11)} =& p^{-2} \sum_{i=1}^{q} h_i h_j \mathbb{E}_l (\mathbf{a}_i' \mathbf{E}_l \mathbf{M}_l^{-1} \mathbf{E}_l' \mathbf{Q}) \mathbb{E}_l (\mathbf{Q} \mathbf{E}_l \mathbf{M}_l^{-1} \mathbf{E}_l \mathbf{a}_i) \\
=& \frac{(l-1) \mathbf{h}' (\boldsymbol{\mathcal{A}}_q' \boldsymbol{\mathcal{A}}_q)^2 \mathbf{h}}{n-k-l+1} + o_p(1),
\end{aligned}
$$

which together with (S3.14) and the fact that $\mathrm{tr}(\boldsymbol{\mathcal{A}}_q \boldsymbol{\mathcal{H}} \boldsymbol{\mathcal{A}}_q')^2 = \mathbf{h}' (\boldsymbol{\mathcal{A}}_q' \boldsymbol{\mathcal{A}}_q)^2 \mathbf{h}$ implies

$$
\begin{aligned}
\frac{1}{p} \sum_{l=1}^{p} \mathbb{E}_{l-1} \mathrm{tr}(\mathbb{E}_l \boldsymbol{\Gamma}_l \mathbb{E}_l \boldsymbol{\Gamma}_l) =& \frac{\mathbf{h}' (\boldsymbol{\mathcal{A}}_q' \boldsymbol{\mathcal{A}}_q)^2 \mathbf{h}}{p} \sum_{l=1}^{p} \left(\frac{l-1}{n-k-l+1} + 1 \right)^2 + o_p(1) \\
=& \mathbf{h}' (\boldsymbol{\mathcal{A}}_q' \boldsymbol{\mathcal{A}}_q)^2 \mathbf{h} \frac{1-\alpha_n}{1-\alpha_n-c_n} + o_p(1).
\end{aligned}
$$

We now turn to prove the term $\mathbb{E}_{l-1} \mathrm{tr}(\mathbb{E}_l \boldsymbol{\Gamma}_l \circ \mathbb{E}_l \boldsymbol{\Gamma}_l) = o_p(1)$. Let $\mathbf{u}_j$ be an n-dimensional column vector with the j-th element being 1 and 0 otherwise. Then we have that

$$
\begin{aligned}
\mathbb{E}(\mathbb{E}_{l-1} \mathrm{tr}(\mathbb{E}_l \boldsymbol{\Gamma}_l \circ \mathbb{E}_l \boldsymbol{\Gamma}_l)) =& \sum_{j=1}^{n} \mathbb{E}(\mathbf{u}_j' \mathbb{E}_l \boldsymbol{\Gamma}_l \mathbf{u}_j)^2 \\
=& \sum_{j=1}^{n} (\mathbb{E} \mathbf{u}_j' \boldsymbol{\Gamma}_l \mathbf{u}_j)^2 + \sum_{j=1}^{n} \mathbb{E}(\mathbb{E}_l \mathbf{u}_j' \boldsymbol{\Gamma}_l \mathbf{u}_j - \mathbb{E} \mathbf{u}_j' \boldsymbol{\Gamma}_l \mathbf{u}_j)^2.
\end{aligned}
$$

By BurkHölder's inequality, we have that

$$\mathbb{E}(\mathbb{E}_l\mathbf{u}_j'\mathbf{\Gamma}_l\mathbf{u}_j-\mathbb{E}\mathbf{u}_j'\mathbf{\Gamma}_l\mathbf{u}_j)^2\le\sum_{s\neq l}^{p}\mathbb{E}(\mathbf{u}_j'\mathbf{\Gamma}_l\mathbf{u}_j-\mathbf{u}_j'\mathbf{\Gamma}_{l\cdot s}\mathbf{u}_j)^2,$$

where $\mathbf{\Gamma}_{l\cdot s}$ is the submatrix of $\mathbf{\Gamma}_l$ with $\mathbf{e}_s$ removed. Applying the inversion formula of block matrix (S3.12) again, we have that

$$\begin{aligned}\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'=&\mathbf{Q}\mathbf{E}_{l\cdot s}\mathbf{M}_{l\cdot s}^{-1}\mathbf{E}_{l\cdot s}'+\frac{\mathbf{Q}\mathbf{e}_s\mathbf{e}_s'}{\beta_{l\cdot s}}\\&+\frac{1}{\beta_{l\cdot s}p^2}\mathbf{Q}\mathbf{E}_{l\cdot s}\mathbf{M}_{l\cdot s}^{-1}\mathbf{E}_{l\cdot s}'\mathbf{Q}\mathbf{e}_s\mathbf{e}_s'\mathbf{Q}\mathbf{E}_{l\cdot s}\mathbf{M}_{l\cdot s}^{-1}\mathbf{E}_{l\cdot s}'\\&-\frac{1}{\beta_{l\cdot s}p}\mathbf{Q}\mathbf{E}_{l\cdot s}\mathbf{M}_{l\cdot s}^{-1}\mathbf{E}_{l\cdot s}'\mathbf{Q}\mathbf{e}_s\mathbf{e}_s'-\frac{1}{\beta_{l\cdot s}p}\mathbf{Q}\mathbf{e}_s\mathbf{e}_s'\mathbf{Q}\mathbf{E}_{l\cdot s}\mathbf{M}_{l\cdot s}^{-1}\mathbf{E}_{l\cdot s}'\\:=&\boldsymbol{\mathcal{U}}_{ls0}+\boldsymbol{\mathcal{U}}_{ls1}+\boldsymbol{\mathcal{U}}_{ls2}-\boldsymbol{\mathcal{U}}_{ls3}-\boldsymbol{\mathcal{U}}_{ls4}\end{aligned}\tag{S3.15}$$

and

$$\begin{aligned}&\mathbb{E}(\mathbf{u}_j'\mathbf{\Gamma}_l\mathbf{u}_j-\mathbf{u}_j'\mathbf{\Gamma}_{ls}\mathbf{u}_j)^2\\=&\mathbb{E}\{p^{-4}\sum_{i=1}^{q}h_i[\mathbf{u}_j'(\boldsymbol{\mathcal{U}}_{ls1}+\boldsymbol{\mathcal{U}}_{ls2}-\boldsymbol{\mathcal{U}}_{ls3}-\boldsymbol{\mathcal{U}}_{ls4})\mathbf{a}_i]^2\\&+2p^{-2}\sum_{i=1}^{q}h_i\mathbf{u}_j'\boldsymbol{\mathcal{U}}_{ls0}\mathbf{a}_i\mathbf{u}_j'(\boldsymbol{\mathcal{U}}_{ls1}+\boldsymbol{\mathcal{U}}_{ls2}-\boldsymbol{\mathcal{U}}_{ls3}-\boldsymbol{\mathcal{U}}_{ls4})\mathbf{a}_i\\&-2p^{-1}\sum_{i=1}^{q}h_i\mathbf{u}_j'\mathbf{a}_i\mathbf{u}_j'(\boldsymbol{\mathcal{U}}_{ls1}+\boldsymbol{\mathcal{U}}_{ls2}-\boldsymbol{\mathcal{U}}_{ls3}-\boldsymbol{\mathcal{U}}_{ls4})\mathbf{a}_i\}^2.\end{aligned}$$

We first consider $\mathbb{E}(\mathbf{u}_j'\boldsymbol{\mathcal{U}}_{ls1}\mathbf{a}_i\mathbf{a}_i'\boldsymbol{\mathcal{U}}_{ls1}'\mathbf{u}_j)^2$. Notice that

$$
\begin{aligned}
&\mathbb{E}(\mathbf{e}_s'\mathbf{Q}\mathbf{u}_j\mathbf{u}_j'\mathbf{Q}\mathbf{e}_s\mathbf{e}_s'\mathbf{a}_i\mathbf{a}_i'\mathbf{e}_s)^2\\
=&\mathbb{E}[(\mathbf{e}_s'\mathbf{Q}\mathbf{u}_j\mathbf{u}_j'\mathbf{Q}\mathbf{e}_s-\mathbf{u}_j'\mathbf{Q}\mathbf{u}_j+\mathbf{u}_j'\mathbf{Q}\mathbf{u}_j)(\mathbf{e}_s'\mathbf{a}_i\mathbf{a}_i'\mathbf{e}_s-1+1)]^2\\
=&\mathbb{E}[(\mathbf{e}_s'\mathbf{Q}\mathbf{u}_j\mathbf{u}_j'\mathbf{Q}\mathbf{e}_s-\mathbf{u}_j'\mathbf{Q}\mathbf{u}_j)(\mathbf{e}_s'\mathbf{a}_i\mathbf{a}_i'\mathbf{e}_s-1)+\mathbf{u}_j'\mathbf{Q}\mathbf{u}_j(\mathbf{e}_s'\mathbf{a}_i\mathbf{a}_i'\mathbf{e}_s-1)\\
&+(\mathbf{e}_s'\mathbf{Q}\mathbf{u}_j\mathbf{u}_j'\mathbf{Q}\mathbf{e}_s-\mathbf{u}_j'\mathbf{Q}\mathbf{u}_j)+\mathbf{u}_j'\mathbf{Q}\mathbf{u}_j]^2.
\end{aligned}
$$

From Lemma 4 we have that for $\ell\geq 2$,

$$\mathbb{E}(\mathbf{e}_s'\mathbf{Q}\mathbf{u}_j\mathbf{u}_j'\mathbf{Q}\mathbf{e}_s-\mathbf{u}_j'\mathbf{Q}\mathbf{u}_j)^\ell=O(n^{\ell-1}\eta_n^{2\ell-4})$$

and

$$\mathbb{E}(\mathbf{e}_s'\mathbf{a}_i\mathbf{a}_i'\mathbf{e}_s-1)^\ell=O(n^{\ell-1}\eta_n^{2\ell-4}).$$

Then, together with the fact that $\frac{1}{\beta_{l\cdot s}}$ and $\mathbf{u}_j'\mathbf{Q}\mathbf{u}_j$ are both bounded, and the c_r-inequality, we obtain

$$\mathbb{E}(\mathbf{u}_j'\boldsymbol{\mathcal{U}}_{ls1}\mathbf{a}_i\mathbf{a}_i'\boldsymbol{\mathcal{U}}_{ls1}'\mathbf{u}_j)^2=O(n^3\eta_n^{-4}).$$

Next, we consider the term $\mathbb{E}(\mathbf{u}_j'\boldsymbol{\mathcal{U}}_{ls0}\mathbf{a}_i\mathbf{a}_i'\boldsymbol{\mathcal{U}}_{ls1}'\mathbf{u}_j)^2$. It follows $\mathbf{a}_i'\mathbf{Q}=\mathbf{0}$ that

$$
\begin{aligned}
&\mathbb{E}(\mathbf{u}_j'\mathbf{Q}\mathbf{E}_{l\cdot s}\mathbf{M}_{l\cdot s}^{-1}\mathbf{E}_{l\cdot s}'\mathbf{a}_i\mathbf{a}_i'\mathbf{e}_s\mathbf{e}_s'\mathbf{Q}\mathbf{u}_j)^2\\
=&\mathbb{E}(\mathbf{e}_s'\mathbf{Q}\mathbf{u}_j\mathbf{u}_j'\mathbf{Q}\mathbf{E}_{l\cdot s}\mathbf{M}_{l\cdot s}^{-1}\mathbf{E}_{l\cdot s}'\mathbf{a}_i\mathbf{a}_i'\mathbf{e}_s-\mathbf{a}_i'\mathbf{Q}\mathbf{u}_j\mathbf{u}_j'\mathbf{Q}\mathbf{E}_{l\cdot s}\mathbf{M}_{l\cdot s}^{-1}\mathbf{E}_{l\cdot s}'\mathbf{a}_i)^2=O(1).
\end{aligned}
$$

As other terms are analogous, thus by combining the above argument, we conclude that $\sum_{j=1}^n \mathbb{E}(\mathbb{E}_l \mathbf{u}_j'\mathbf{\Gamma}_l\mathbf{u}_j - \mathbb{E}\mathbf{u}_j'\mathbf{\Gamma}_l\mathbf{u}_j)^2 = o(1)$.

For $\sum_{j=1}^n(\mathbb{E}\mathbf{u}_j'\mathbf{\Gamma}_l\mathbf{u}_j)^2$, it follows from the assumption that $\{e_{ij}\}$ are i.i.d.,

$$\begin{aligned}\sum_{j=1}^n(\mathbb{E}\mathbf{u}_j'\mathbf{\Gamma}_l\mathbf{u}_j)^2 =& \sum_{j=1}^n(n^{-1}p^{-1}\mathbb{E}\mathrm{tr}\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_q'\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\boldsymbol{\mathcal{A}}_q + \mathbf{u}_j'\boldsymbol{\mathcal{A}}_q\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_q'\mathbf{u}_j)^2 \\ =& \sum_{j=1}^n(\mathbf{u}_j'\boldsymbol{\mathcal{A}}_q\boldsymbol{\mathcal{H}}\boldsymbol{\mathcal{A}}_q'\mathbf{u}_j)^2 + O(n^{-1}) \\ =& \mathbf{h}'(\boldsymbol{\mathcal{A}}_q \circ \boldsymbol{\mathcal{A}}_q)'(\boldsymbol{\mathcal{A}}_q \circ \boldsymbol{\mathcal{A}}_q)\mathbf{h} + O(n^{-1}).\end{aligned}$$

Here we use a result similar to (S3.8), that is

$$\mathbb{E}\frac{1}{p}\mathbf{r}_3'\mathbf{E}\mathbf{M}^{-1}\mathbf{E}'\mathbf{r}_4 - \frac{c_n\mathbf{r}_3'\mathbf{r}_4}{1-c_n-\alpha_n} + \frac{c_n^2\mathbf{r}_3'\mathbf{Q}\mathbf{r}_4}{(1-c_n-\alpha_n)(1-\alpha_n)} \to 0,$$

and the proof can be found in the proof of Proposition 3.1 in Bai et al. (2022). Then we conclude that

$$\sum_{l=1}^p \mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l\mathbf{\Gamma}_l \circ \mathbb{E}_l\mathbf{\Gamma}_l) = \mathbf{h}'(\boldsymbol{\mathcal{A}}_q \circ \boldsymbol{\mathcal{A}}_q)'(\boldsymbol{\mathcal{A}}_q \circ \boldsymbol{\mathcal{A}}_q)\mathbf{h} + o_p(1).$$

Next, we will prove that the non-random sequence

$$\mathcal{M}_2 = \mathcal{M}_2^{(n)} = o(1).$$

Write $\mathbf{M}^{-1} = (M^{ij})$. Without loss of generality, we only need to prove $p^{-1}\mathbb{E}\mathbf{a}_1'\mathbf{E}\mathbf{M}^{-1}\mathbf{E}'\mathbf{a}_1 - \frac{c_n}{1-c_n-\alpha_n} = o(p^{-1/2})$. Because the entries of $\mathbf{E}$ are i.i.d., we have

$$p^{-1}\mathbb{E}\mathbf{a}_1'\mathbf{E}\mathbf{M}^{-1}\mathbf{E}'\mathbf{a}_1 = p^{-1}\sum_{i,j=1}^{p}\mathbb{E}\mathbf{a}_1'\mathbf{e}_i M^{ij}\mathbf{e}_j'\mathbf{a}_1$$

$$=\mathbb{E}\mathbf{e}_1'\mathbf{a}_1\mathbf{a}_1'\mathbf{e}_1 M^{11} + (p-1)\mathbb{E}\mathbf{a}_1'\mathbf{e}_1 M^{12}\mathbf{e}_2'\mathbf{a}_1. \tag{S3.16}$$

From the inverse matrix formula, we know that

$$M^{11} = \frac{1}{\beta_1} = \frac{1}{\beta_1^{tr}} - \frac{\xi_1}{(\beta_1^{tr})^2} + \frac{\xi_1^2}{\beta_1(\beta_1^{tr})^2}.$$

and

$$M^{12} = \frac{\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1}{p\beta_1^{tr}} - \frac{\xi_1\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1}{p(\beta_1^{tr})^2} + \frac{\xi_1^2\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1}{p\beta_1(\beta_1^{tr})^2}. \tag{S3.17}$$

Then it follows from (S3.2), (S3.3) and the Hölder's inequality that

$$\mathbb{E}\mathbf{e}_1'\mathbf{a}_1\mathbf{a}_1'\mathbf{e}_1 M^{11} - \frac{c_n}{1-c_n-\alpha_n} = \mathbb{E}\frac{\mathbf{e}_1'\mathbf{a}_1\mathbf{a}_1'\mathbf{e}_1\xi_1^2}{\beta_1(\beta_1^{tr})^2} = o(p^{-1/2}).$$

Moreover, substituting (S3.17) into the second term of (S3.16), we have three terms. The first one is

$$\mathbb{E}\frac{\mathbf{e}_2'\mathbf{a}_1\mathbf{a}_1'\mathbf{e}_1\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1}{p\beta_1^{tr}} = \mathbb{E}\frac{\mathbf{e}_2'\mathbf{a}_1\mathbf{a}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1}{p\beta_1^{tr}} = 0,$$

because of $\mathbf{a}_1'\mathbf{Q} = \mathbf{0}$. Applying the inversion formula to $\mathbf{M}_1^{-1}$ again, we obtain that

$$\begin{aligned}
&\mathbb{E}\frac{\xi_1\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1\mathbf{a}_1'\mathbf{e}_1\mathbf{e}_2'\mathbf{a}_1}{p}\\
=&\mathbb{E}\frac{\xi_1\mathbf{e}_1'\mathbf{Q}\mathbf{e}_2\mathbf{a}_1'\mathbf{e}_1\mathbf{e}_2'\mathbf{a}_1}{\beta_{1\cdot 2}p} - \mathbb{E}\frac{\xi_1\mathbf{e}_1'\mathbf{Q}\mathbf{E}_{1\cdot 2}\mathbf{M}_{1\cdot 2}^{-1}\mathbf{E}_{1\cdot 2}'\mathbf{Q}\mathbf{e}_2\mathbf{a}_1'\mathbf{e}_1\mathbf{e}_2'\mathbf{a}_1}{\beta_{1\cdot 2}p^2}\\
=&\mathbb{E}\frac{\xi_1\mathbf{e}_1'\mathbf{Q}\mathbf{e}_2\mathbf{a}_1'\mathbf{e}_1\mathbf{e}_2'\mathbf{a}_1}{\beta_{1\cdot 2}p}.
\end{aligned}$$

Rewrite $1/\beta_{1\cdot 2}$ as

$$\frac{1}{\beta_{1\cdot 2}} = \frac{1}{\beta_2^{tr}} - \frac{\xi_{1\cdot 2}}{\beta_{1\cdot 2}\beta_2^{tr}}$$

and by the the fact that $\mathbf{Q}\mathbf{a}_1 = \mathbf{0}$, we have that

$$\mathbb{E}\frac{\xi_1 \mathbf{e}_1' \mathbf{Q}\mathbf{e}_2 \mathbf{a}_1' \mathbf{e}_1 \mathbf{e}_2' \mathbf{a}_1}{\beta_{1\cdot 2}} = -\mathbb{E}\frac{\xi_{1\cdot 2}\xi_1 \mathbf{e}_1' \mathbf{Q}\mathbf{e}_2 \mathbf{a}_1' \mathbf{e}_1 \mathbf{e}_2' \mathbf{a}_1}{\beta_2^{tr}\beta_{1\cdot 2}} = o(p^{-1/2}).$$

Therefore, by combining the above results, we conclude that

$$\mathcal{M}_2 = o(1),$$

and we complete the proof of the theorem.

S3.3 Proof of Theorem 3

Note that

$$\mathrm{tr}[(\mathbf{Y}'\mathbf{Q}\mathbf{Y} - (n-k)\mathbf{I}) \circ (\mathbf{Y}'\mathbf{Q}\mathbf{Y} - (n-k)\mathbf{I})] = \sum_{i=1}^{p} (\mathbf{e}_i' \mathbf{Q}\mathbf{e}_i - (n-k))^2$$

and

$$\mathbb{E}(\mathbf{e}_i' \mathbf{Q}\mathbf{e}_i - (n-k))^2 = 2(n-k) + \tau \mathrm{tr}(\mathbf{Q} \circ \mathbf{Q}).$$

Thus by the definition of $\hat{\tau}$ and $\{\mathbf{e}_i\}$ are i.i.d., we have $\mathbb{E}\hat{\tau} = \tau$. Next we will show that

$$\mathbb{E}(\hat{\tau} - \tau)^2 \to 0.$$

It follows from (S3.1) and Lemma 3 that

$$\begin{aligned}
&\mathbb{E}(\hat{\tau}-\tau)^2 = \mathbb{E}(\hat{\tau}-\mathbb{E}\hat{\tau})^2\\
=&p^{-1}\mathbb{E}[(\mathbf{e}_1'\mathbf{Q}\mathbf{e}_1-(n-k))^2-\mathbb{E}(\mathbf{e}_1'\mathbf{Q}\mathbf{e}_1-(n-k))]^2/\mathrm{tr}^2(\mathbf{Q}\circ\mathbf{Q})\\
=&\{p^{-1}\mathbb{E}(\mathbf{e}_1'\mathbf{Q}\mathbf{e}_1-(n-k))^4-p^{-1}[\mathbb{E}(\mathbf{e}_1'\mathbf{Q}\mathbf{e}_1-(n-k))]^2\}/\mathrm{tr}^2(\mathbf{Q}\circ\mathbf{Q})\\
\leq&Kp^{-1}[((n-k)^2+n^2(n-k)\eta_n^4+(n-k)^2+\tau^2\mathrm{tr}^2(\mathbf{Q}\circ\mathbf{Q})]/\mathrm{tr}^2(\mathbf{Q}\circ\mathbf{Q}),
\end{aligned}$$

where K is a positive constant. By c_r-inequality, we have that $\mathrm{tr}(\mathbf{Q}\circ\mathbf{Q})\geq n^{-1}(n-k)^2$, which together with condition (C1) implies $\mathbb{E}(\hat{\tau}-\tau)^2\to 0$. Then we complete the proof of this theorem.

S3.4 Proof of Theorem 4

For simple presentation, in the following we assume $\{1\}\subset\mathbf{j}_*$ and let $j=1$. Then by the notation $\mathbf{b}=\mathbf{\Sigma}^{-1/2}\mathbf{\Theta}_*\mathbf{X}_*'\mathbf{a}_1$, $\mathcal{K}_1=p^{-1}(\mathbf{b}'+\mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1+\mathbf{b})$. Note that the proof procedure of Theorem 4 is the same as that of Theorem 2. And the difference is that Theorem 4 requires the consideration of linear combinations of three different forms of random variables, namely $\mathbf{a}_1'\mathbf{E}\mathbf{M}^{-1}\mathbf{E}'\mathbf{a}_1$, $\mathbf{a}_1'\mathbf{E}\mathbf{M}^{-1}\mathbf{b}$ and $\mathbf{b}'\mathbf{M}^{-1}\mathbf{b}$. As the asymptotic normality of $\mathbf{a}_1'\mathbf{E}\mathbf{M}^{-1}\mathbf{E}'\mathbf{a}_1$ is proved in last subsection, in the sequel we only focus on the other two terms and their correlations.

Analogously, we split the proof of this theorem into two parts. It is worthy noting that next we may use the same notation as in the proof of Theorem 2, but they represent a little different content. First, we show the asymptotic normality of the sequence of random variables

$$\mathcal{M}_3 := \sqrt{p}[p^{-1}(\mathbf{b}' + \mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1 + \mathbf{b}) - \mathbb{E}p^{-1}(\mathbf{b}' + \mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1 + \mathbf{b})].$$

Second, we prove the non-random sequence

$$\mathcal{M}_4 = \sqrt{p}[\mathbb{E}p^{-1}(\mathbf{b}' + \mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1 + \mathbf{b}) - \frac{c_n(1+\delta_1)}{1 - c_n - \alpha_n}]$$

tends to zero. It follows that

$$\begin{aligned}\mathcal{M}_1 =& p^{-1/2}\sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})(\mathbf{b}' + \mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1 + \mathbf{b})\\ =& p^{-1/2}\sum_{l=1}^{p}(\mathbb{E}_l - \mathbb{E}_{l-1})[(\mathbf{b}' + \mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1 + \mathbf{b}) - (\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}(\mathbf{E}_l'\mathbf{a}_1 + \mathbf{b}_l)].\end{aligned}$$

By the inversion formula of block matrix, we obtain

$$\begin{aligned}&(\mathbf{b}' + \mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1 + \mathbf{b}) - (\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}(\mathbf{E}_l'\mathbf{a}_1 + \mathbf{b}_l)\\ =&\frac{1}{\beta_l p^2}(\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l\mathbf{e}_l'\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}(\mathbf{E}_l'\mathbf{a}_1 + \mathbf{b}_l)\\ &- \frac{2}{\beta_l p}(\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l(\mathbf{e}_l'\mathbf{a}_1 + b_l) + \frac{(\mathbf{e}_l'\mathbf{a}_1 + b_l)^2}{\beta_l}.\end{aligned}$$

Then, by the equation (S3.2), we can rewrite $\mathcal{M}_3$ as

$$\mathcal{M}_3 = \frac{1}{p^{1/2}\beta_1^{tr}}\sum_{l=1}^{p}\mathbb{E}_l[\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \mathrm{tr}\mathbf{\Gamma}_l + 2\mathbf{e}_l'\boldsymbol{\gamma}_l] + \mathcal{M}_{30},$$

where

$$\begin{aligned}\mathbf{\Gamma}_l =& p^{-2}\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}(\mathbf{b}_l+\mathbf{E}_l'\mathbf{a}_1)(\mathbf{b}_l+\mathbf{E}_l'\mathbf{a}_1)'\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\\ &-p^{-1}\mathbf{a}_1(\mathbf{b}_l+\mathbf{E}_l'\mathbf{a}_1)'\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}-p^{-1}\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}(\mathbf{b}_l+\mathbf{E}_l'\mathbf{a}_1)\mathbf{a}_1'+\mathbf{a}_1\mathbf{a}_1',\end{aligned}$$

$$\boldsymbol{\gamma}_l=-p^{-1}b_l\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}(\mathbf{b}_l+\mathbf{E}_l'\mathbf{a}_1)+\mathbf{a}_1 b_l$$

and

$$\begin{aligned}\mathcal{M}_{30}=&-\sum_{l=1}^{p}(\mathbb{E}_l-\mathbb{E}_{l-1})\frac{\xi_l(\mathbf{b}_l'+\mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l\mathbf{e}_l'\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}(\mathbf{E}_l'\mathbf{a}_1+\mathbf{b}_l)}{p^{5/2}\beta_l\beta_1^{tr}}\\ &+2\sum_{l=1}^{p}(\mathbb{E}_l-\mathbb{E}_{l-1})\frac{\xi_l(\mathbf{b}_l'+\mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l(\mathbf{e}_l'\mathbf{a}_1+b_l)}{p^{3/2}\beta_l\beta_1^{tr}}\\ &-\sum_{l=1}^{p}(\mathbb{E}_l-\mathbb{E}_{l-1})\frac{\xi_l(\mathbf{e}_l'\mathbf{a}_1+b_l)^2}{p^{1/2}\beta_l\beta_1^{tr}}\\ :=&-\mathcal{M}_{301}+2\mathcal{M}_{302}-\mathcal{M}_{303}.\end{aligned}$$

Next we will prove $\mathcal{M}_{10}=o_p(1)$. Substitute (S3.2) into $\mathcal{M}_{101}$, $\mathcal{M}_{102}$ and

$\mathcal{M}_{103}$ respectively, we then have that

$$\mathcal{M}_{301} = \sum_{l=1}^{p} (\mathbb{E}_l - \mathbb{E}_{l-1}) \frac{\xi_l(\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{\Psi}_l(\mathbf{E}_l'\mathbf{a}_1 + \mathbf{b}_l)}{p^{5/2}(\beta_1^{tr})^2} + \sum_{l=1}^{p} (\mathbb{E}_l - \mathbb{E}_{l-1}) \frac{\xi_l(\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}(\mathbf{E}_l'\mathbf{a}_1 + \mathbf{b}_l)}{p^{3/2}(\beta_1^{tr})^2}$$
$$- \sum_{l=1}^{p} (\mathbb{E}_l - \mathbb{E}_{l-1}) \frac{\xi_l^2(\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l\mathbf{e}_l'\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}(\mathbf{E}_l'\mathbf{a}_1 + \mathbf{b}_l)}{p^{5/2}\beta_l(\beta_1^{tr})^2},$$

$$\mathcal{M}_{302} = \sum_{l=1}^{p} (\mathbb{E}_l - \mathbb{E}_{l-1}) \frac{\xi_l(\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l(\mathbf{e}_l'\mathbf{a}_1 + b_l)}{p^{3/2}(\beta_1^{tr})^2}$$
$$- \sum_{l=1}^{p} (\mathbb{E}_l - \mathbb{E}_{l-1}) \frac{\xi_l^2(\mathbf{b}_l' + \mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{e}_l(\mathbf{e}_l'\mathbf{a}_1 + b_l)}{p^{3/2}\beta_l(\beta_1^{tr})^2},$$

$$\mathcal{M}_{303} = \sum_{l=1}^{p} (\mathbb{E}_l - \mathbb{E}_{l-1}) \frac{\xi_l(\mathbf{e}_l'\mathbf{a}_1\mathbf{a}_1'\mathbf{e}_l - 1 + 2b_l\mathbf{a}_1'\mathbf{e}_l + b_l^2 + 1)}{p^{1/2}(\beta_1^{tr})^2} - \sum_{l=1}^{p} (\mathbb{E}_l - \mathbb{E}_{l-1}) \frac{\xi_l^2(\mathbf{e}_l'\mathbf{a}_1 + b_l)^2}{p^{1/2}\beta_l(\beta_1^{tr})^2}.$$

These together with (S3.4), (S3.5) and the BurkHölder's inequality (see Lemma 2) implies that $\mathcal{M}_{10} = o_p(1)$. Note that here we used the fact $\mathbf{Q}\mathbf{a}_1 = \mathbf{0}$.

Applying Lemma 3, we have that

$$\mathbb{E}|\mathbb{E}_l(\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \mathrm{tr}\mathbf{\Gamma}_l)|^4 \leq \mathbb{E}|\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \mathrm{tr}\mathbf{\Gamma}_l|^4 = O(p\eta_n^4)$$

and

$$\mathbb{E}|\mathbb{E}_l(\mathbf{e}_l'\boldsymbol{\gamma}_l)|^4 \leq \mathbb{E}|\mathbf{e}_l'\boldsymbol{\gamma}_l\boldsymbol{\gamma}_l^*\mathbf{e}_l|^2 = O(p^{-2}b_l^4),$$

which verify the condition (ii) in Lemma 1. Thus, what we need is to obtain the limit of

$$\frac{1}{p(\beta_1^{tr})^2} \sum_{l=1}^{p} \mathbb{E}_{l-1}\{\mathbb{E}_l[\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \mathrm{tr}\mathbf{\Gamma}_l + 2\mathbf{e}_l'\boldsymbol{\gamma}_l]\}^2.$$

By Lemma 5, we have that

$$\begin{aligned}
&\mathbb{E}_{l-1}\{\mathbb{E}_l[\mathbf{e}_l'\mathbf{\Gamma}_l\mathbf{e}_l - \mathrm{tr}\mathbf{\Gamma}_l + 2\mathbf{e}_l'\boldsymbol{\gamma}_l]\}^2 \\
=&2\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l\mathbf{\Gamma}_l\mathbb{E}_l\mathbf{\Gamma}_l) + 4\mathbb{E}_{l-1}(\mathbb{E}_l\boldsymbol{\gamma}_l\mathbb{E}_l\boldsymbol{\gamma}_l') \\
&+ \tau\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l\mathbf{\Gamma}_l \circ \mathbb{E}_l\mathbf{\Gamma}_l) + 4\mathbb{E}e_{11}^3\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l\mathbf{\Gamma}_l \circ \mathbb{E}_l\boldsymbol{\gamma}_l\mathbf{1}').
\end{aligned}$$

Notice that

$$\begin{aligned}
\mathrm{tr}(\mathbb{E}_l\mathbf{\Gamma}_l\mathbb{E}_l\mathbf{\Gamma}_l) =&p^{-4}\mathrm{tr}(\mathbb{E}_l\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}(\mathbf{b}_l + \mathbf{E}_l'\mathbf{a}_1)(\mathbf{b}_l + \mathbf{E}_l'\mathbf{a}_1)'\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})^2 \\
&+ p^{-2}2\mathbb{E}_l((\mathbf{b}_l + \mathbf{E}_l'\mathbf{a}_1)'\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})\mathbb{E}_l(\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}(\mathbf{b}_l + \mathbf{E}_l'\mathbf{a}_1)) + 1.
\end{aligned} \tag{S3.18}$$

In the proof of Theorem 2, we have shown that

$$\begin{aligned}
&p^{-4}\mathrm{tr}(\mathbb{E}_l\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{a}_1\mathbf{a}_1\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})^2 \\
&+ p^{-2}2\mathbb{E}_l(\mathbf{a}_1\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})\mathbb{E}_l(\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{a}_1) + 1 \\
=&\left(\frac{l-1}{n-k-l+1} + 1\right)^2 + o_p(1).
\end{aligned}$$

By the same procedure and the assumptions in Theorem 4, we can also have that

$$p^{-4}\mathrm{tr}(\mathbb{E}_l\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{b}_l\mathbf{b}_l'\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})^2 = \left(\frac{\sum_{i=1}^{l-1} b_i^2}{n-k-l+1}\right)^2 + o_p(1),$$

$$p^{-4}\mathrm{tr}(\mathbb{E}_l\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{b}_l\mathbf{a}_1\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})^2 = \frac{(l-1)\sum_{i=1}^{l-1} b_i^2}{n-k-l+1} + o_p(1),$$

$$p^{-2}\mathbb{E}_l(\mathbf{b}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})\mathbb{E}_l(\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{a}_1) = o_p(1),$$

and

$$p^{-2}\mathbb{E}_l(\mathbf{b}_l\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q})\mathbb{E}_l(\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}\mathbf{b}_l) = \frac{\sum_{i=1}^{l-1} b_i^2}{n-k-l+1} + o_p(1),$$

which together with (S3.18) and (S3.14) implies

$$\frac{1}{p}\sum_{l=1}^{p}\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l\mathbf{\Gamma}_l\mathbb{E}_l\mathbf{\Gamma}_l) = \frac{1}{p}\sum_{l=1}^{p}\left(\frac{(l-1+\sum_{i=1}^{l-1} b_i^2)}{n-k-l+1}+1\right)^2 + o_p(1).$$

For $\mathbb{E}_l\boldsymbol{\gamma}_l\mathbb{E}_l\boldsymbol{\gamma}_l'$, by the notation $\widetilde{\mathbf{M}}_l = \frac{1}{p}\widetilde{\mathbf{E}}_l'\mathbf{Q}\widetilde{\mathbf{E}}_l$, we have that

$$\begin{aligned}
&\mathbb{E}_l\boldsymbol{\gamma}_l'\mathbb{E}_l\boldsymbol{\gamma}_l \\
=&\mathbb{E}_l[(p^{-1}b_l\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}(\mathbf{b}_l+\mathbf{E}_l'\mathbf{a}_1)-\mathbf{a}_1b_l)'((p^{-1}b_l\mathbf{Q}\mathbf{E}_l\mathbf{M}_l^{-1}(\mathbf{b}_l+\mathbf{E}_l'\mathbf{a}_1)-\mathbf{a}_1b_l)] \\
=&\mathbb{E}_l[p^{-2}b_l^2(\mathbf{b}_l+\mathbf{a}_1'\mathbf{E}_l)\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\widetilde{\mathbf{E}}_l\widetilde{\mathbf{M}}_l^{-1}(\mathbf{b}_l+\widetilde{\mathbf{E}}_l'\mathbf{a}_1)+b_l^2].
\end{aligned}$$

Applying the inversion formula of block matrix (S3.12) again, we obtain that

$$p^{-1}\sum_{l=1}^{p}\mathbb{E}_{l-1}(\mathbb{E}_l\boldsymbol{\gamma}_l'\mathbb{E}_l\boldsymbol{\gamma}_l) = p^{-1}\sum_{l=1}^{p} b_l^2\left(\frac{(l-1+\sum_{i=1}^{l-1} b_i^2)}{n-k-l+1}+1\right) + o_p(1).$$

Then by applying Lemma 6, we have that as $n \to \infty$,

$$
\begin{aligned}
&\frac{1}{p}\sum_{l=1}^{p}\mathbb{E}_{l-1}(\mathrm{tr}(\mathbb{E}_l\boldsymbol{\Gamma}_l\mathbb{E}_l\boldsymbol{\Gamma}_l)+2\mathbb{E}_l\boldsymbol{\gamma}_l\mathbb{E}_l\boldsymbol{\gamma}_l')\\
=&\frac{1}{p}\sum_{l=1}^{p}\left(\left(\frac{l-1+\sum_{i=1}^{l-1}b_i^2}{n-k-l+1}+1\right)^2+2b_l^2\left(\frac{l-1+\sum_{i=1}^{l-1}b_i^2}{n-k-l+1}+1\right)\right)+o_p(1)\\
=&\frac{n}{p}\int_0^{c_n}\left[\left(\frac{t(1+\delta_1)}{1-\alpha_n-t}+1\right)^2+2\delta_1\left(\frac{t(1+\delta_1)}{1-\alpha_n-t}+1\right)\right]dt+o_p(1)\\
=&\frac{(1-\alpha_n)(1+2\delta_1)+c_n\delta_1^2}{1-\alpha_n-c_n}+o_p(1).
\end{aligned}
$$

We now turn to the term $\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l\boldsymbol{\Gamma}_l\circ\mathbb{E}_l\boldsymbol{\Gamma}_l)$. By the notation that $\mathbf{u}_j$ is an n-dimensional column vector with the j-th element being 1 and 0 otherwise and repeating the same argument in the proof of Theorem 2, we can obtain that

$$
\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l\boldsymbol{\Gamma}_l\circ\mathbb{E}_l\boldsymbol{\Gamma}_l)=\sum_{j=1}^{n}(\mathbb{E}\mathbf{u}_j'\boldsymbol{\Gamma}_l\mathbf{u}_j)^2+o_p(1).
$$

As $\{e_{ij}\}$ are i.i.d., thus from the assumptions of this theorem, we have that

$$
\begin{aligned}
&\sum_{j=1}^{n}(\mathbb{E}\mathbf{u}_j'\boldsymbol{\Gamma}_l\mathbf{u}_j)^2\\
=&\sum_{j=1}^{n}(n^{-1}p^{-1}\mathbb{E}(\mathbf{b}_l+\mathbf{E}_l'\mathbf{a}_1)'\mathbf{M}_l^{-1}(\mathbf{b}_l+\mathbf{E}_l'\mathbf{a}_1)\\
&-2p^{-1}\mathbf{u}_j'\mathbf{a}_1\mathbb{E}(\mathbf{b}_l+\mathbf{E}_l'\mathbf{a}_1)'\mathbf{M}_l^{-1}\mathbf{E}_l'\mathbf{Q}\mathbf{u}_j+(\mathbf{u}_j'\mathbf{a}_1)^2]^2\\
=&o(1).
\end{aligned}
$$

Then we conclude that

$$\frac{1}{p}\sum_{l=1}^{p}\mathbb{E}_{l-1}\mathrm{tr}(\mathbb{E}_l\mathbf{\Gamma}_l\circ\mathbb{E}_l\mathbf{\Gamma}_l)=o_p(1).$$

Next, we will prove that the non-random sequence

$$\mathcal{M}_4=o(1).$$

By the notation $\mathbf{M}^{-1}=(M^{ij})$ and $\{e_{ij}\}$ are i.i.d., we have that

$$\mathbb{E}(\mathbf{b}'+\mathbf{a}_1'\mathbf{E})\mathbf{M}^{-1}(\mathbf{E}'\mathbf{a}_1+\mathbf{b})=\sum_{i,j=1}^{p}\mathbb{E}(b_i+\mathbf{a}_1'\mathbf{e}_i)M^{ij}(b_j+\mathbf{a}_1'\mathbf{e}_j)$$

$$=\mathbf{b}'\mathbf{b}\mathbb{E}M^{11}+p\mathbb{E}\mathbf{e}_1'\mathbf{a}_1\mathbf{a}_1'\mathbf{e}_1M^{11}+2\mathbb{E}\mathbf{a}_1'\mathbf{e}_1M^{11}\sum_{i=1}^{p}b_i$$

$$+\sum_{i\neq j}^{p}b_ib_j\mathbb{E}M^{12}+p(p-1)\mathbb{E}\mathbf{a}_1'\mathbf{e}_1M^{12}\mathbf{e}_2'\mathbf{a}_1+2(p-1)\mathbb{E}\mathbf{a}_1'\mathbf{e}_1M^{12}\sum_{i=1}^{p}b_i.$$

From the inverse matrix formula, we know that

$$M^{11}=\frac{1}{\beta_1}=\frac{1}{\beta_1^{tr}}-\frac{\xi_1}{(\beta_1^{tr})^2}+\frac{\xi_1^2}{\beta_1(\beta_1^{tr})^2}.$$

and

$$M^{12}=\frac{\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1}{p\beta_1^{tr}}-\frac{\xi_1\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1}{p(\beta_1^{tr})^2}+\frac{\xi_1^2\mathbf{e}_1'\mathbf{Q}\mathbf{E}_1\mathbf{M}_1^{-1}\mathbf{u}_1}{p\beta_1(\beta_1^{tr})^2}.$$

Then it follows from (S3.2), (S3.3) and the Hölder's inequality that

$$p^{-1}\mathbf{b}'\mathbf{b}\mathbb{E}M^{11}-\frac{c_n\delta_1}{1-c_n-\alpha_n}=o(p^{-1/2})$$

and

$$\mathbb{E}\mathbf{e}_1'\mathbf{a}_1\mathbf{a}_1'\mathbf{e}_1 M^{11} - \frac{c_n}{1 - c_n - \alpha_n} = o(p^{-1/2}).$$

By the facts that

$$|\mathbb{E}\mathbf{a}_1'\mathbf{e}_1\xi_1| \leq \sqrt{\mathbb{E}|\xi_1|^2} = O(p^{-1/2})$$

and

$$|\sum_{i=1}^{p} b_i| = O(p^{1/2}),$$

we can obtain that

$$\mathbb{E}\mathbf{a}_1'\mathbf{e}_1 M^{11} \sum_{i=1}^{p} b_i = O(1).$$

It follows from

$$|p^{-1} \sum_{i \neq j}^{p} b_i b_j| = O(1), \quad |\mathbb{E}\xi_1 \mathbf{e}_1' \mathbf{Q}\mathbf{E}_1 \mathbf{M}_1^{-1} \mathbf{u}_1| = O(1)$$

and the Hölder's inequality, we have that

$$\sum_{i \neq j}^{p} b_i b_j \mathbb{E} M^{12} = o(p^{-1/2}).$$

Therefore, similar to the proof of Theorem 2, we conclude that

$$\mathcal{M}_2 = o(1),$$

and we complete the proof of this theorem.

S3.5 Some useful lemmas

Lemma 1 (Theorem 35.12 of Billingsley (1995)). *Suppose that for each n, $Y_{n1}, Y_{n2}, \ldots, Y_{nr_n}$ is a real martingale difference sequence with respect to the increasing σ -field $\{\mathcal{F}_{nj}\}$ having second moments. If as $n \to \infty$, for each $\varepsilon > 0$,*

(i) $\sum_{j=1}^{r_n} \mathbb{E}\left(Y_{nj}^2 | \mathcal{F}_{n,j-1}\right) \xrightarrow{p} \sigma^2$, *where σ^2 is a positive constant;*

(ii) $\sum_{j=1}^{r_n} \mathbb{E}\left(Y_{nj}^2 I_{(|Y_{nj}| \geq \varepsilon)}\right) \to 0$,

then we have that

$$\sum_{j=1}^{r_n} Y_{nj} \xrightarrow{\mathcal{D}} N\left(0, \sigma^2\right).$$

Lemma 2 (Burkholder (1971)). *Let $\{Y_k\}$ be a martingale difference sequence with respect to the increasing σ-field $\{\mathcal{F}_k\}$. Then, for $\ell > 1$,*

$$\mathbb{E}\left|\sum X_k\right|^\ell \leq K_\ell \left(\mathbb{E}\left(\sum \mathbb{E}(|X_k|^2 \, |\mathcal{F}_{k-1})\right)^{\ell/2} + \sum \mathbb{E}\,|X_k|^\ell\right).$$

Lemma 3 (Lemma 2.7 of Bai and Silverstein (1998)). *For $\mathbf{e} = (e_1, \ldots, e_n)'$ i.i.d. standardized entries, $\mathbf{A}$ a $n \times n$ matrix, we have, for any $\ell \geq 2$*

$$\mathbb{E}\,|\mathbf{e}'\mathbf{A}\mathbf{e} - \mathrm{tr}\mathbf{A}|^\ell \leq K_\ell \left(\left(\mathbb{E}\,|e_1|^4 \,\mathrm{tr}\mathbf{A}\mathbf{A}'\right)^{\ell/2} + \mathbb{E}\,|e_1|^{2\ell}\,\mathrm{tr}(\mathbf{A}\mathbf{A}')^{\ell/2}\right).$$

Lemma 4 (Lemma 7.2 of Bai and Yao (2005)). *Let $\mathbf{e} = (e_1, \ldots, e_n)'$ be a random n-vector with i.i.d. standardized entries. Suppose $E\,|e_i|^4 < \infty$ and $|e_i| \leq \eta_n \sqrt{n}$ with $\eta_n \to 0$ slowly. Assume that $\mathbf{A}$ is a symmetric matrix of* *order n bounded in norm by M. Then, for any given $2 \leqslant \ell \leqslant b \log(n\eta_n^2)$ with some $b > 1$, there exists a constant K such that*

$$\mathbb{E}\left|\mathbf{e}'\mathbf{A}\mathbf{e} - \operatorname{tr}(\mathbf{A})\right|^{\ell} \leq n^{\ell}\left(n\eta_n^4\right)^{-1}\left(MK\eta_n^2\right)^{\ell}.$$

Lemma 5. *Let $\mathbf{B}$ and $\mathbf{C}$ be $n \times n$ matrices. Let $\mathbf{d}$ be a n-vector. Let $\mathbf{e} = (e_1, \ldots, e_n)'$ be a random n-vector with i.i.d. standardized entries. Let $\tau := \mathbb{E}e_i^4 - 3$Then, we have that*

$$\mathbb{E}\left\{(\mathbf{e}'\mathbf{B}\mathbf{e} - \operatorname{tr}\mathbf{B})(\mathbf{e}'\mathbf{C}\mathbf{e} - \operatorname{tr}\mathbf{C})\right\} = \operatorname{tr}(\mathbf{B}\mathbf{C}) + \operatorname{tr}(\mathbf{B}\mathbf{C}') + \tau\sum_{i=1}^{n} b_{ii}c_{ii}$$

and

$$\mathbb{E}(\mathbf{e}'\mathbf{B}\mathbf{e}\mathbf{e}'\mathbf{d}) = \mathbb{E}e_1^3\sum_{i=1}^{n} b_{ii}d_i.$$

Lemma 6 (Lemma 3.1 of Bai and Pan (2012))**.** *Let $\left\{\mathbf{d}_n = (d_1, \ldots, d_n)'\right\}$ be a sequence of unit vectors with $\max_{k\leq n}|d_k| \to 0$. There is a permutation m of $\{1, \ldots, n\}$ given by*

$$\begin{pmatrix} 1 & 2 & \cdots & n \\ m(1) & m(2) & \cdots & m(n) \end{pmatrix}$$

such that $\mathbf{d}_f = \left(d_{f(1)}, \ldots, d_{f(n)}\right)$ and F_{nm} tends to a uniform distribution over the interval $(0, 1)$, where F_{nm} is a distribution function defined by

$$F_{nm}(t) = \sum_{i\leq nt}\left|d_{f(i)}\right|^2.$$

Bibliography

Bai, Z., K. P. Choi, Y. Fujikoshi, and J. Hu (2022). Asymptotics of AIC, BIC and Cp model selection rules in high-dimensional regression. Bernoulli 28(4), 2375–2403.

Bai, Z. D., K. P. Choi, and Y. Fujikoshi (2018). Limiting behavior of eigenvalues in high-dimensional MANOVA via RMT. The Annals of Statistics 46(6A), 2985–3013.

Bai, Z. D. and G. M. Pan (2012). Limiting Behavior of Eigenvectors of Large Wigner Matrices. Journal of Statistical Physics 146(3), 519–549.

Bai, Z. D. and J. W. Silverstein (1998). No eigenvalues outside the support of the limiting spectral distribution of large-dimensional sample covariance matrices. The Annals of Probability 26(1), 316–345.

Bai, Z. D. and J. W. Silverstein (1999). Exact separation of eigenvalues of large dimensional sample covariance matrices. The Annals of Probability 27(3), 1536–1555.

Bai, Z. D. and J. W. Silverstein (2004). CLT for linear spectral statistics of large-dimensional sample covariance matrices. The Annals of Probability 32(1), 553–605.

Bai, Z. D. and J. F. Yao (2005). On the convergence of the spectral empirical process of Wigner matrices. Bernoulli 11(6), 1059–1092.

Billingsley, P. (1995). Probability and Measure. John Wiley&Sons, New York.

Breiman, L. and J. H. Friedman (1997). Predicting Multivariate Responses in Multiple Linear Regression. Journal of the Royal Statistical Society: Series B (Statistical Methodology) 59(1), 3–54.

Burkholder, D. L. (1971). Distribution function inequalities for martingales. The Annals of Probability 1(1), 19–42.

Similä, T. and J. Tikka (2007). Input selection and shrinkage in multiresponse linear regression. Computational Statistics & Data Analysis 52(1), 406–422.

Skagerberg, B., J. F. MacGregor, and C. Kiparissides (1992). Multivariate data analysis applied to low-density polyethylene reactors. Chemometrics and Intelligent Laboratory Systems 14(1), 341–356.