Supplementary Material to "High-Dimensional Interaction Detection with False Sign Rate Control"

Daoji Li, Yinfei Kong, Yingying Fan and Jinchi Lv

This supplementary material consists of three parts. Appendix C reports additional numerical results. Appendix D provides Lemma 1 and its proof. This lemma establishes the sure screening property for both interaction and main effect screening and thus Condition 1 holds under some sufficient conditions. Appendix E contains some additional technical lemmas and their proofs. Hereafter we use $\widetilde{C}_i$ with $i = 1, 2, \ldots$ to denote some generic positive or nonnegative constants whose values may vary from line to line. For any set $\mathcal{G}$, denote by $|\mathcal{G}|$ its cardinality.

C Additional Numerical Results

In this section, we report additional numerical results, which include all screening results for all settings in studies 1 and 2 and additional selection results.

C.1 Screening Results for Studies 1 and 2

For the screening step, we employed several recent feature screening procedures: the sure independence screening (Fan and Lv, 2008), feature screening via distance correlation (Li et al., 2012), variable selection via sliced inverse regression (Jiang and Liu, 2014), and interaction pursuit via distance correlation (Kong et al., 2017), respectively. Since this paper focuses on interaction models with one single response, we can also consider the method of interaction pursuit via Pearson correlation for screening, which is exactly the same as interaction pursuit via distance correlation, except for the placement of distance correlation by Pearson correlation when identifying the variables in $\mathcal{A}$ and $\mathcal{B}$. The method in Jiang and Liu (2014) is an iterative procedure that alternates between a large-scale variable screening step and a moderate-scale variable selection step when the dimensionality p is large. Since all other screening methods are non-iterative, in this section we compare the initial screening step of Jiang and Liu's method with other methods. We use the full iterative method in Jiang and Liu (2014) when comparing the variable selection performance. Each method in Fan and Lv (2008), Li et al. (2012), and the initial screening step of Jiang and Liu (2014) returns a set of variables without distinguishing between important main effects and active interaction variables. Thus for each of those methods, we construct interactions using all possible pairwise interactions of the recruited variables. By doing so, the strong heredity assumption is enforced. We would like to remark that the resulting feature screening procedures are different from their original versions.

In other words, we include the following five methods to assess the variable screening performance: SIS2, the sure independence screening; DC-SIS2, feature screening via distance correlation; SIRI*2, variable selection via sliced inverse regression; IPDC, interaction pursuit via distance correlation; IP, interaction pursuit via the Pearson correlation. For the first three methods, we construct interactions using the recruited variables. For the third method, only the initial screening step of the method in Jiang and Liu (2014) is used to recruit variables.

Table 3: The percentages of retaining each important interaction or main effect, and all important ones (All) by all the screening methods for models 1-4.

Method	Model 1				Model 2				Model 3				Model 4		
	X_1	X_5	X_1X_5	All	X_1	X_{10}	X_1X_5	All	X_{10}	X_{15}	X_1X_5	All	X_1X_5	$X_{10}X_{15}$	All
$(n,p,\rho) = (300, 5000, 0)$															
SIS2	1.00	1.00	1.00	1.00	1.00	1.00	0.14	0.14	1.00	1.00	0.02	0.02	0.00	0.01	0.00
DC-SIS2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.10	0.10	0.18	0.20	0.01
SIRI*2	1.00	1.00	1.00	1.00	1.00	1.00	0.93	0.93	1.00	1.00	0.32	0.32	0.63	0.65	0.43
IPDC	1.00	1.00	1.00	1.00	1.00	1.00	0.98	0.98	1.00	1.00	1.00	1.00	1.00	1.00	1.00
IP	1.00	1.00	0.98	0.98	1.00	1.00	0.98	0.98	1.00	1.00	0.97	0.97	0.86	0.85	0.71
$(n,p,\rho) = (300, 5000, 0.5)$															
SIS2	1.00	1.00	1.00	1.00	1.00	1.00	0.09	0.09	1.00	1.00	0.02	0.02	0.00	0.01	0.00
DC-SIS2	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.16	0.16	0.32	0.25	0.05
SIRI*2	1.00	1.00	1.00	1.00	1.00	1.00	0.92	0.92	1.00	1.00	0.34	0.34	0.70	0.57	0.36
IPDC	1.00	1.00	0.99	0.99	1.00	1.00	0.99	0.99	1.00	1.00	0.99	0.99	1.00	1.00	1.00
IP	1.00	1.00	0.99	0.99	1.00	1.00	0.94	0.94	1.00	1.00	0.97	0.97	0.81	0.89	0.70

Table 3 lists the comparison results for all screening methods in recovering each important interaction or main effect, and retaining all important ones for models 1-4. Table 4 lists the comparison results for model 5. For model 1 satisfying the strong heredity

Table 4: The percentages of retaining each important interaction or main effect, and all important ones (All) by all the screening methods for model 5.

	X_1	X_{10}	X_{20}	X_{30}	X_{40}	X_1X_5	X_1X_{10}	X_5X_{15}	$X_{10}X_{15}$	All
					$(n,p,\rho)=(300,5000,0)$					
SIS2	0.96	0.96	0.98	0.99	0.99	0.09	0.92	0.01	0.05	0.01
DC-SIS2	0.96	0.99	0.99	1.00	0.99	0.26	0.95	0.09	0.42	0.08
SIRI*2	0.90	0.95	0.90	0.94	0.93	0.32	0.85	0.18	0.50	0.08
IPDC	0.97	1.00	0.99	1.00	0.99	0.79	0.85	0.85	0.95	0.73
IP	0.95	1.00	0.97	0.98	0.97	0.42	0.56	0.55	0.64	0.25
					$(n,p,\rho)=(300,5000,0.5)$					
SIS2	0.96	0.88	1.00	0.93	0.94	0.10	0.84	0.02	0.06	0.01
DC-SIS2	1.00	0.97	1.00	0.98	0.99	0.33	0.97	0.19	0.35	0.19
SIRI*2	0.90	0.93	0.97	0.91	0.89	0.46	0.85	0.22	0.45	0.16
IPDC	1.00	0.99	1.00	0.98	0.96	0.81	0.85	0.90	0.95	0.73
IP	0.94	0.97	0.97	0.90	0.93	0.46	0.50	0.60	0.64	0.21

Table 5: The percentages of retaining each important interaction or main effect, and all important ones (All) by all the screening methods for study 2.

$(n,p,\rho)=(300,10000,0)$				$(n,p,\rho)=(400,10000,0)$			
	X_1X_5	$X_{10}X_{15}$	All		X_1X_5	$X_{10}X_{15}$	All
SIS2	0.00	0.00	0.00	SIS2	0.00	0.00	0.00
DC-SIS2	0.08	0.13	0.00	DC-SIS2	0.32	0.30	0.09
SIRI*2	0.50	0.51	0.28	SIRI*2	0.84	0.76	0.65
IPDC	0.98	0.99	0.97	IPDC	1.00	1.00	1.00
IP	0.76	0.83	0.59	IP	0.97	0.98	0.95
$(n,p,\rho)=(300,10000,0.5)$				$(n,p,\rho)=(400,10000,0.5)$			
	X_1X_5	$X_{10}X_{15}$	All		X_1X_5	$X_{10}X_{15}$	All
SIS2	0.00	0.00	0.00	SIS2	0.00	0.01	0.00
DC-SIS2	0.10	0.16	0.00	DC-SIS2	0.39	0.36	0.17
SIRI*2	0.53	0.55	0.28	SIRI*2	0.81	0.78	0.64
IPDC	1.00	0.99	0.99	IPDC	1.00	1.00	1.00
IP	0.82	0.83	0.65	IP	0.96	0.95	0.91

assumption, all methods performed rather similarly and all retaining percentages were either equal or close to 100%. The last four methods performed similarly and improved over the sure independence screening method in model 2 in which the weak heredity assumption holds. In models 3 and 4, the method of interaction pursuit via distance correlation and the method of interaction pursuit with the Pearson correlation significantly outperformed all other methods in detecting interactions across all settings, showing their advantage when the heredity assumption is not satisfied. We also observe that the sure independence screening method failed to detect interactions, whereas the method of variable selection via sliced inverse regression improved over the method of feature screening via distance correlation in these two models. Model 5 was designed to examine the robustness of each method at the presence of more main effects and interactions. The advantages of interaction pursuit via distance correlation and interaction pursuit remain in this model. These results suggest that a separate screening step should be designed specifically for interactions to improve the screening accuracy, which is indeed one of the main innovations of the method of interaction pursuit in Kong et al. (2017). Table 5 shows the same conclusions.

C.2 Additional Selection Results for Studies 1 and 2

Table 6: Variable selection results for study 1 with $(n, p, \rho) = (300, 5000, 0)$. Reported values are medians and robust standard deviations (in parentheses) of three performance measures: PE, prediction error; FS, falsely discovered signs; and Time, running time in seconds. 0^* means that the corresponding value is small than 0.001.

	SIS2-L_1+SICA	DC-SIS2-L_1+SICA	SIRI	IPDC-L_1+SICA	IP-Lasso	IP-L_1+SICA	Oracle
Model 1							
PE	3.1 (0.8)	3.1 (0.7)	2.9 (0.3)	3.1 (0.8)	3.6 (0.4)	3.1 (0.9)	2.9 (0.3)
FS	0 (5.2)	0 (4.5)	0 (0)	0 (2.0)	112.5 (24.6)	0 (6.7)	0 (0)
Time	715.7 (37.5)	712.3 (35.5)	808.1 (38.9)	113.6 (17.2)	7.2 (0.7)	122.5 (11.8)	0*(0*)
Model 2							
PE	19.6 (3.3)	2.1 (0.3)	2.2 (0.0)	2.1 (0.4)	2.4 (0.2)	2.1 (0.2)	2.0 (0.1)
FS	25 (14.9)	1 (4.1)	3 (3.0)	0 (5.4)	96 (21.3)	0 (2.2)	0 (0.0)
Time	734.4 (38.5)	718.2 (37.3)	798.3 (49.3)	107.4 (9.0)	6.8 (0.6)	116.5 (6.9)	0* (0*)
Model 3							
PE	20.8 (2.0)	20.0 (3.2)	13.2 (0.5)	2.1 (0.3)	2.4 (0.2)	2.1 (0.2)	2.0 (0.1)
FS	29.5 (5.2)	27 (13.4)	8 (3.0)	0 (2.7)	98.5 (19.4)	0 (3.4)	0 (0.0)
Time	752.5 (32.8)	746.4 (34.3)	436.6 (9.2)	109.8 (9.5)	6.7 (0.6)	116.1 (8.6)	0* (0*)
Model 4							
PE	36.3 (3.2)	34.2 (13.3)	12.3 (2.7)	1.4 (0.1)	1.6 (9.0)	1.4 (9.7)	1.3 (0.0)
FS	35 (14.9)	28.5 (12.3)	6 (6.7)	0 (0.0)	78 (36.2)	0 (4.1)	0 (0.0)
Time	749.2 (15.3)	752.5 (19.9)	223.3 (137.1)	98.3 (6.9)	6.4 (0.9)	108.2 (11.4)	0* (0*)
Model 5							
PE	63.6 (10.8)	47.0 (18.1)	6.4 (34.4)	2.0 (17.0)	36.4 (31.8)	36.3 (33.6)	2.0 (0.0)
FS	20 (6.7)	16 (12.3)	21 (9.7)	5 (1.5)	120 (35.1)	10 (11.6)	0 (0.0)
Time	744.4 (58.7)	749.1 (41.1)	499.8 (82.4)	112.7 (14.9)	9.7 (2.3)	141.5 (17.2)	0* (0*)

Table 7: Variable selection results for study 2. Reported values are medians and robust standard deviations (in parentheses) of three performance measures: PE, prediction error; FS, falsely discovered signs; and Time, running time in seconds. $0*$ means that the corresponding value is small than 0.001.

	SIS2-L_1+SICA	DC-SIS2-L_1+SICA	SIRI	IPDC-L_1+SICA	IP-Lasso	IP-L_1+SICA	Oracle
			$(n,p,\rho)=(300,10000,0)$				
PE	37.7 (2.0)	37.0 (4.7)	13.3 (8.6)	2.9 (0.4)	3.3 (9.5)	3.1 (11.2)	2.9 (0.2)
FS	34 (13.4)	31.5 (12.7)	6.0 (5.2)	0 (1.3)	79 (43.3)	1 (8.6)	0 (0.0)
Time	791.9 (163.2)	785.2 (177.0)	361.3 (220.1)	100.8 (9.0)	7.8 (1.6)	126.0 (24.1)	0* (0*)
			$(n,p,\rho)=(300,10000,0.5)$				
PE	37.7 (2.4)	36.4 (10.8)	13.1 (8.5)	2.9 (0.4)	3.2 (9.3)	3.0 (10.4)	2.9 (0.2)
FS	35 (13.4)	31 (12.7)	6 (5.2)	0 (0.0)	79.5 (38.1)	0 (6.7)	0 (0.0)
Time	830.0 (144.7)	812.3 (141.0)	653.2 (399.3)	101.0 (7.8)	7.4 (1.7)	121.2 (18.7)	0* (0*)
			$(n,p,\rho)=(400,10000,0)$				
PE	37.6 (3.0)	23.3 (13.2)	12.4 (0.4)	2.9 (0.4)	3.0 (0.2)	2.9 (0.2)	2.9 (0.2)
FS	41 (17.2)	33 (12.7)	8 (7.5)	0 (0.0)	74 (17.2)	0 (0.0)	0 (0.0)
Time	1836.1 (449.1)	1998.0 (460.6)	533.0 (177.9)	175.5 (8.9)	14.4 (3.5)	224.1 (52.0)	0* (0*)
			$(n,p,\rho)=(400,10000,0.5)$				
PE	38.6 (2.2)	22.1 (12.9)	12.6 (8.7)	2.9 (0.4)	3.0 (0.2)	2.9 (0.3)	2.9 (0.2)
FS	44 (5.2)	33.5 (17.5)	8 (5.2)	0 (0.0)	76.5 (16.8)	0 (0.0)	0 (0.0)
Time	1896.2 (468.9)	1906.3 (452.4)	792.1 (384.0)	179.0 (14.0)	14.4 (3.7)	203.8 (54.0)	0* (0*)

D Lemma 1 and Its Proof

D.1 Lemma 1

Lemma 1. *a) Under Conditions 2 and 6, if* $0 \leq \max\{2\kappa_1+4\xi_1, 2\kappa_1+4\xi_2\} < 1$ *and* $E(Y^4)=O(1)$*, then for any* $C>0$*, there exists some constant* $C_1>0$ *depending on* C *such that for* $\log p = o(n^{\alpha_1\eta_1})$ *with* $\eta_1 = \min\{(1-2\kappa_1-4\xi_2)/(8+\alpha_1),\ (1-2\kappa_1-4\xi_1)/(12+\alpha_1)\}$*,*

$$P(\max_{1\leq k\leq p} |\widehat{\omega}_k - \omega_k| \geq Cn^{-\kappa_1}) = o(n^{-C_1}). \tag{D.1}$$

b) Under Conditions 2 and 6, if $0 \leq \max\{2\kappa_2+2\xi_1, 2\kappa_2+2\xi_2\} < 1$ *and* $E(Y^2)=O(1)$*, then for any* $C>0$*, there exists some constant* $C_2>0$ *depending on* C *such that*

$$P(\max_{1\leq j\leq p} |\widehat{\omega}_j^* - \omega_j^*| \geq Cn^{-\kappa_2}) = o(n^{-C_2}) \tag{D.2}$$

for $\log p = o(n^{\alpha_1\eta_2})$ *with* $\eta_2 = \min\{(1-2\kappa_2-2\xi_2)/(4+\alpha_1), (1-2\kappa_2-2\xi_1)/(6+\alpha_1)\}$*.*

c) Under Conditions 2, 6 and 7, and the choices of $\tau = c_2 n^{-\kappa_1}$ *and* $\widetilde{\tau} = c_2 n^{-\kappa_2}$*, if*

$0 \leq \xi_1, \xi_2 < \min\{1/4 - \kappa_1/2, 1/2 - \kappa_2\}$ *and* $E(Y^4) = O(1)$*, then we have*

$$P\Big(\mathcal{I} \subset \widehat{\mathcal{I}} \;\; and \;\; \mathcal{M} \subset \widehat{\mathcal{M}}\Big) = 1 - o\Big(n^{-\min\{C_1, C_2\}}\Big) \tag{D.3}$$

for $\log p = o(n^{\alpha_1 \min\{\eta_1, \eta_2\}})$ *with constants* C_1 *and* C_2 *given in* (D.1) *and* (D.2)*, respectively. In addition, it holds that*

$$\begin{aligned} &P\left(|\widehat{\mathcal{I}}| \leq O\{n^{4\kappa_1}\lambda_{\max}^2(\Sigma^*)\} \;\; and \;\; |\widehat{\mathcal{M}}| \leq O\{n^{2\kappa_1}\lambda_{\max}(\Sigma^*) + n^{2\kappa_2}\lambda_{\max}(\Sigma)\}\right) \\ &\quad = 1 - o\left(n^{-\min\{C_1, C_2\}}\right), \end{aligned} \tag{D.4}$$

where $\lambda_{\max}(\cdot)$ *denotes the largest eigenvalue,* Σ *is the population covariance matrix of the random vector* $(X_1, \ldots, X_p)^T$ *and* Σ^* *is the population covariance matrix of the random vector* $(X_1^*, \ldots, X_p^*)^T$ *with* $X_k^* = \{X_k^2 - E(X_k^2)\}/\{\mathrm{var}(X_k^2)\}^{1/2}$.

Comparing the results from the first two parts of Lemma 1 on interactions and main effects, respectively, we see that interaction screening generally requires more restrictive assumption on dimensionality p. This reflects that the task of interaction screening is intrinsically more challenging than that of main effect screening. In particular, when $\alpha_1 = 2$, the method of interaction pursuit via Pearson correlation can handle ultra-high dimensionality up to

$$\log p = o\left(n^{\min\{(1-2\kappa_1-4\xi_2)/5,\, (1-2\kappa_1-4\xi_1)/7,\, (1-2\kappa_2-2\xi_2)/3,\, (1-2\kappa_2-2\xi_1)/4\}}\right). \tag{D.5}$$

It is worth mentioning that both constants C_1 and C_2 in the probability bounds (D.1)–(D.2) can be chosen arbitrarily large without affecting the order of p and ranges of constants κ_1 and κ_2. We also observe that stronger marginal signal strength for interaction variables and main effects, in terms of smaller values of κ_1 and κ_2, can enable us to tackle higher dimensionality.

The third part of Lemma 1 shows that the method of interaction pursuit via Pearson correlation enjoys the sure screening property for both interaction and main effect screening, and thus Condition 1 holds with $C = \min\{C_1, C_2\}$ and $\eta = \alpha_1 \min\{\eta_1, \eta_2\}$. The third part of lemma 1 also admits an explicit bound on the size of the reduced model after screening. More specifically, an upper bound of the reduced model size is controlled by the choices of both thresholds τ and $\widetilde{\tau}$, and the largest eigenvalues of the two population covariance matrices Σ^* and Σ. If we assume $\lambda_{\max}(\Sigma^*) = O(n^{\xi_3})$ and $\lambda_{\max}(\Sigma) = O(n^{\xi_4})$ for some constants $\xi_3, \xi_4 \geq 0$, then with overwhelming probability the total number of interactions and main effects in the reduced model is at most of a polynomial order of sample size n.

The thresholds $\tau = c_2 n^{-\kappa_1}$ and $\tilde{\tau} = c_2 n^{-\kappa_2}$ given in Lemma 1 depend on unknown constants c_2, κ_1, and κ_2, and thus are unavailable in practice. In real applications, to estimate the set of active interaction variables $\mathcal{A}$, we sort $|\hat{\omega}_k|, 1 \leq k \leq p$, in decreasing order and then retain the top d variables. This strategy is also widely used in the existing literature; see, for example, Fan and Lv (2008), Fan and Song (2010), Li et al. (2012), Barut et al. (2016), and Zhou et al. (2019). The set of main effects $\mathcal{B}$ is estimated similarly except that the marginal utility $|\hat{\omega}_k^*|$ is used. Following the suggestion in Fan and Lv (2008), one may choose the number of retained variables for each of sets $\mathcal{A}$ and $\mathcal{B}$ in a screening procedure as $n-1$ or $[cn/(\log n)]$ with c some positive constant, depending on the available sample size n. The parameter c can be tuned using some data-driven method such as the cross-validation.

It is worth pointing out that our result is weaker than that in Fan and Lv (2008) in terms of growth of dimensionality, where one can allow $\log p = o(n^{1-2\kappa_2})$. This is mainly because they considered linear models without interactions, indicating the intrinsic challenges of feature screening in the presence of interactions. Moreover, our assumptions on the distributions for the covariates and errors are more flexible.

The results in Lemma 1 can be improved in the case when the covariates X_j's and the response Y are uniformly bounded. An application of the proofs for (D.1)–(D.2) in

D.2 and D.3 of the e-companion to this paper yields

$$P\Big(\max_{1\le k\le p}|\widehat{\omega}_k-\omega_k|\ge c_2 n^{-\kappa_1}\Big)\le pC_3\exp(-C_3^{-1}n^{1-2\kappa_1}),$$
$$P\Big(\max_{1\le j\le p}|\widehat{\omega}_j^*-\omega_j^*|\ge c_2 n^{-\kappa_2}\Big)\le pC_3\exp(-C_3^{-1}n^{1-2\kappa_2}),$$

where C_3 is some positive constant. In this case, the method of interaction pursuit via Pearson correlation can handle ultra-high dimensionality $\log p = o(n^{\xi})$ with $\xi = \min\{1-2\kappa_1, 1-2\kappa_2\}$.

D.2 Proof of part a) of Lemma 1

Let $S_{k1}=n^{-1}\sum_{i=1}^n X_{ik}^2Y_i^2$, $S_{k2}=n^{-1}\sum_{i=1}^n X_{ik}^2$, $S_{k3}=n^{-1}\sum_{i=1}^n X_{ik}^4$, and $S_4=n^{-1}\sum_{i=1}^n Y_i^2$. Then ω_k and $\widehat{\omega}_k$ can be written as

$$\omega_k=\frac{E(S_{k1})-E(S_{k2})E(S_4)}{\sqrt{E(S_{k3})-E^2(S_{k2})}}\quad\text{and}\quad \widehat{\omega}_k=\frac{S_{k1}-S_{k2}S_4}{\sqrt{S_{k3}-S_{k2}^2}}.$$

To prove (D.1), the key step is to show that for any positive constant C, there exist some constants $\widetilde{C}_1,\ldots,\widetilde{C}_4>0$ such that the following probability bounds

$$P(\max_{1\le k\le p}|S_{k1}-E(S_{k1})|\ge Cn^{-\kappa_1})\le p\widetilde{C}_1\exp\left(-\widetilde{C}_2 n^{\alpha_1\eta_1}\right)+\widetilde{C}_3\exp\left(-\widetilde{C}_4 n^{\alpha_2\eta_1}\right),\tag{D.6}$$

$$P(\max_{1\le k\le p}|S_{k2}-E(S_{k2})|\ge Cn^{-\kappa_1})\le p\widetilde{C}_1\exp[-\widetilde{C}_2 n^{\alpha_1(1-2\kappa_1)/(4+\alpha_1)}],\tag{D.7}$$

$$P(\max_{1\le k\le p}|S_{k3}-E(S_{k3})|\ge Cn^{-\kappa_1})\le p\widetilde{C}_1\exp[-\widetilde{C}_2 n^{\alpha_1(1-2\kappa_1)/(8+\alpha_1)}],\tag{D.8}$$

$$P(|S_4-E(S_4)|\ge Cn^{-\kappa_1})\le \widetilde{C}_1\exp\left(-\widetilde{C}_2 n^{\alpha_1\zeta_1}\right)+\widetilde{C}_3\exp\left(-\widetilde{C}_4 n^{\alpha_2\zeta_2'}\right)\tag{D.9}$$

hold for all n sufficiently large when $0\le 2\kappa_1+4\xi_1<1$ and $0\le 2\kappa_1+4\xi_2<1$, where $\eta_1=\min\{(1-2\kappa_1-4\xi_2)/(8+\alpha_1),\,(1-2\kappa_1-4\xi_1)/(12+\alpha_1)\}$, $\zeta_1=\min\{(1-2\kappa_1-4\xi_2)/(4+\alpha_1),\,(1-2\kappa_1-4\xi_1)/(8+\alpha_1)\}$, $\zeta_2=\min\{(1-2\kappa_1-2\xi_2)/(4+\alpha_1),\,(1-2\kappa_1-2\xi_1)/(6+\alpha_1)\}$, and $\zeta_2'=\min\{\zeta_2,(1-2\kappa_1)/(4+\alpha_2)\}$. Define $\eta=\min\{\eta_1,(1-2\kappa_1)/(4+\alpha_1),(1-2\kappa_1)/(8+\alpha_1),\zeta_1\}$ and $\zeta=\min\{\eta_1,\zeta_2'\}$. Then $\eta=\eta_1$ and $\zeta=\min\{\eta_1,(1-2\kappa_1)/(4+\alpha_2)\}$. Thus, by Lemmas 8–12, we have

$$P(\max_{1\le k\le p}|\widehat{\omega}_k-\omega_k|\ge Cn^{-\kappa_1})\le p\widetilde{C}_1\exp(-\widetilde{C}_2 n^{\alpha_1\eta})+\widetilde{C}_3\exp(-\widetilde{C}_4 n^{\alpha_2\zeta}). \tag{D.10}$$

Thus, if $\log p=o\{n^{\alpha_1\eta}\}$, the result of the part (a) in Lemma 1 follows immediately.

It thus remains to prove the probability bounds (D.6)–(D.9). Since the proofs of (D.6)–(D.9) are similar, here we focus on (D.6) to save space. Throughout the proof, the same notation $\widetilde{C}$ is used to denote a generic positive constant without loss of generality, which may take different values at each appearance.

Recall that $Y_i=\alpha_0+x_i^T\beta_0+z_i^T\gamma_0+\varepsilon_i=\beta_0+x_{i,\mathcal{B}}^T\beta_{0,\mathcal{B}}+z_{i,\mathcal{I}}^T\gamma_{0,\mathcal{I}}+\varepsilon_i$, where $x_i=(X_{i1},\ldots,X_{ip})^T$, $z_i=(X_{i1}X_{i2},\ldots,X_{i,p-1}X_{i,p})^T$, $x_{i,\mathcal{B}}=(X_{ij},j\in\mathcal{B})^T$, $z_{i,\mathcal{I}}=(X_{ik}X_{i\ell},(k,\ell)\in\mathcal{I})^T$, $\beta_{0,\mathcal{B}}=(\beta_{0,j}\in\mathcal{B})^T$, and $\gamma_{0,\mathcal{I}}=(\gamma_{0,k\ell},(k,\ell)\in\mathcal{I})^T$. To simplify the presentation, we assume that the intercept α_0 is zero without loss of generality. Thus

$$\begin{aligned}S_{k1}=&n^{-1}\sum_{i=1}^n X_{ik}^2Y_i^2=n^{-1}\sum_{i=1}^n X_{ik}^2(x_{i,\mathcal{B}}^T\beta_{0,\mathcal{B}}+z_{i,\mathcal{I}}^T\gamma_{0,\mathcal{I}}+\varepsilon_i)^2\\ =&n^{-1}\sum_{i=1}^n X_{ik}^2(x_{i,\mathcal{B}}^T\beta_{0,\mathcal{B}}+z_{i,\mathcal{I}}^T\gamma_{0,\mathcal{I}})^2+2n^{-1}\sum_{i=1}^n X_{ik}^2(x_{i,\mathcal{B}}^T\beta_{0,\mathcal{B}}+z_{i,\mathcal{I}}^T\gamma_{0,\mathcal{I}})\varepsilon_i+n^{-1}\sum_{i=1}^n X_{ik}^2\varepsilon_i^2\\ \triangleq&S_{k1,1}+2S_{k1,2}+S_{k1,3}.\end{aligned}$$

Similarly, $E(S_{k1})$ can be written as $E(S_{k1})=E(S_{k1,1})+2E(S_{k1,2})+E(S_{k1,3})$. So $S_{k1}-E(S_{k1})$ can be expressed as $S_{k1}-E(S_{k1})=[S_{k1,1}-E(S_{k1,1})]+2[S_{k1,2}-E(S_{k1,2})]+[S_{k1,3}-E(S_{k1,3})]$. By the triangle inequality and the union bound we have

$$\begin{aligned}P(\max_{1\le k\le p}|S_{k1}-E(S_{k1})|\ge Cn^{-\kappa_1})\le&P(\bigcup_{j=1}^3\{\max_{1\le k\le p}|S_{k1,j}-E(S_{k1,j})|\ge Cn^{-\kappa_1}/4\})\\ \le&\sum_{j=1}^3 P(\max_{1\le k\le p}|S_{k1,j}-E(S_{k1,j})|\ge Cn^{-\kappa_1}/4).\end{aligned} \tag{D.11}$$

In what follows, we will provide details on deriving an exponential tail probability bound for each term on the right hand side above. To enhance readability, we split the proof into three steps.

Step 1. We start with the first term $\max_{1\le k\le p}|S_{k1,1}-E(S_{k1,1})|$. Define the event $\Omega_i=\{|X_{ij}|\le M_1 \text{ for all } j\in\mathcal{M}\cup\{k\}\}$ with $\mathcal{M}=\mathcal{A}\cup\mathcal{B}$ and M_1 a large positive number that will be specified later. Let $T_{k1}=n^{-1}\sum_{i=1}^n X_{ik}^2(x_{i,\mathcal{B}}^T\beta_{0,\mathcal{B}}+z_{i,\mathcal{I}}^T\gamma_{0,\mathcal{I}})^2\mathbb{I}_{\Omega_i}$ and $T_{k2}=n^{-1}\sum_{i=1}^n X_{ik}^2(x_{i,\mathcal{B}}^T\beta_{0,\mathcal{B}}+z_{i,\mathcal{I}}^T\gamma_{0,\mathcal{I}})^2\mathbb{I}_{\Omega_i^c}$, where $\mathbb{I}(\cdot)$ is the indicator function and Ω_i^c is the complement of the set Ω_i. Then

$$S_{k1,1}-E(S_{k1,1})=[T_{k1}-E(T_{k1})]+T_{k2}-E(T_{k2}). \tag{D.12}$$

Note that $E(T_{k2})=E[X_{1k}^2(x_{1,\mathcal{B}}^T\beta_{0,\mathcal{B}}+z_{1,\mathcal{I}}^T\gamma_{0,\mathcal{I}})^2\mathbb{I}_{\Omega_1^c}]$. By the fact $(a+b)^2\le 2(a^2+b^2)$ for two real numbers a and b, the Cauchy-Schwarz inequality, and Condition 6, we have

$$(x_{1,\mathcal{B}}^T\beta_{0,\mathcal{B}}+z_{1,\mathcal{I}}^T\gamma_{0,\mathcal{I}})^2\le 2[(x_{1,\mathcal{B}}^T\beta_{0,\mathcal{B}})^2+(z_{1,\mathcal{I}}^T\gamma_{0,\mathcal{I}})^2]\le 2C_0^2(s_2\|x_{1,\mathcal{B}}\|^2+s_1\|z_{1,\mathcal{I}}\|^2), \tag{D.13}$$

where C_0 is some positive constant and $\|\cdot\|$ denotes the Euclidean norm. This ensures that $E(T_{k2})$ is bounded by $2C_0^2[s_2E(X_{1k}^2\|x_{1,\mathcal{B}}\|^2\mathbb{I}_{\Omega_1^c})+s_1E(X_{1k}^2\|z_{1,\mathcal{I}}\|^2\mathbb{I}_{\Omega_1^c})]$. By the Cauchy-Schwarz inequality, the union bound, and the inequality $(a+b)^2\le 2(a^2+b^2)$, we obtain that

$$\begin{aligned}
E(X_{1k}^2\|x_{1,\mathcal{B}}\|^2\mathbb{I}_{\Omega_1^c}) &\le \left[E(X_{1k}^4\|x_{1,\mathcal{B}}\|^4)P(\Omega_1^c)\right]^{1/2}\le\left\{\left[s_2\sum_{j\in\mathcal{B}}E(X_{1k}^4X_{1j}^4)\right]P(\Omega_1^c)\right\}^{1/2}\\
&\le\left\{2^{-1}s_2\sum_{j\in\mathcal{B}}[E(X_{1k}^8)+E(X_{1j}^8)]\right\}^{1/2}\left[\sum_{j\in\mathcal{M}\cup\{k\}}P(|X_{ij}|>M_1)\right]^{1/2}\\
&\le\widetilde{C}s_2(1+s_2+2s_1)^{1/2}\exp[-M_1^{\alpha_1}/(2c_1)]
\end{aligned}$$

for some positive constant $\widetilde{C}$, where the last inequality follows from Condition 2 and Lemma 3. Similarly, we have $E(X_{1k}^2\|z_{1,\mathcal{I}}\|^2\mathbb{I}_{\Omega_1^c})\le\widetilde{C}s_1(1+s_2+2s_1)^{1/2}\exp[-M_1^{\alpha_1}/(2c_1)]$.

This together with the above inequalities entails that

$$0 \leq E(T_{k2}) \leq 2C_0^2 \widetilde{C}(s_1^2 + s_2^2)(1 + s_2 + 2s_1)^{1/2} \exp[-M_1^{\alpha_1}/(2c_1)].$$

If we choose $M_1 = n^{\eta_1}$ with $\eta_1 > 0$, then by Condition 6, for any positive constant C, when n is sufficiently large,

$$\begin{aligned} |E(T_{k2})| &\leq 2C_0^2 \widetilde{C}(n^{2\xi_1} + n^{2\xi_2})(1 + n^{\xi_2} + 2n^{\xi_1})^{1/2} \exp[-n^{\alpha_1 \eta_1}/(2c_1)] \\ &< Cn^{-\kappa_1}/12 \end{aligned} \tag{D.14}$$

holds uniformly for all $1 \leq k \leq p$. The above inequality together with (D.12) ensures that

$$\begin{aligned} &P(\max_{1 \leq k \leq p} |S_{k1,1} - E(S_{k1,1})| \geq Cn^{-\kappa_1}/4) \\ \leq &P(\max_{1 \leq k \leq p} |T_{k1} - E(T_{k1})| \geq Cn^{-\kappa_1}/12) + P(\max_{1 \leq k \leq p} |T_{k2}| \geq Cn^{-\kappa_1}/12) \end{aligned} \tag{D.15}$$

for all n sufficiently large. Thus we only need to establish the probability bound for each term on the right hand side of (D.15).

First consider $\max_{1 \leq k \leq p} |T_{k1} - E(T_{k1})|$. Using similar arguments for proving (D.13), we have $(x_{i,\mathcal{B}}^T \beta_{0,\mathcal{B}} + z_{i,\mathcal{I}}^T \gamma_{0,\mathcal{I}})^2 \leq 2C_0^2(s_2 \|x_{i,\mathcal{B}}\|^2 + s_1 \|z_{i,\mathcal{I}}\|^2)$ and thus

$$0 \leq X_{ik}^2 (x_{i,\mathcal{B}}^T \beta_{0,\mathcal{B}} + z_{i,\mathcal{I}}^T \gamma_{0,\mathcal{I}})^2 \mathbb{I}_{\Omega_i} \leq 2C_0^2 X_{ik}^2 (s_2 \|x_{i,\mathcal{B}}\|^2 + s_1 \|z_{i,\mathcal{I}}\|^2) \mathbb{I}_{\Omega_i} \leq 2C_0^2 M_1^4 (s_2^2 + s_1^2 M_1^2).$$

For any $\delta > 0$, by Hoeffding's inequality, we obtain

$$\begin{aligned} P(|T_{k1} - E(T_{k1})| \geq \delta) \leq &2 \exp\left[-\frac{n\delta^2}{2C_0^4 M_1^8 (s_2^2 + s_1^2 M_1^2)^2}\right] \leq 2 \exp\left[-\frac{n\delta^2}{4C_0^4 M_1^8 (s_2^4 + s_1^4 M_1^4)}\right] \\ \leq &2 \exp\left(-\frac{n\delta^2}{8C_0^4 M_1^8 s_2^4}\right) + 2 \exp\left(-\frac{n\delta^2}{8C_0^4 M_1^{12} s_1^4}\right), \end{aligned}$$

where we have used the fact that $(a+b)^2 \leq 2(a^2 + b^2)$ for any real numbers a and b, and

$\exp[-c/(a+b)] \leq \exp[-c/(2a)] + \exp[-c/(2b)]$ for any $a, b, c > 0$. Recall that $M_1 = n^{\eta_1}$. Under Condition 6, taking $\delta = Cn^{-\kappa_1}/12$ gives that

$$P(\max_{1\leq k\leq p} |T_{k1} - E(T_{k1})| \geq Cn^{-\kappa_1}/12) \leq \sum_{k=1}^{p} P(|T_{k1} - E(T_{k1})| \geq Cn^{-\kappa_1}/12)$$
$$\leq 2p\exp\left(-\widetilde{C}n^{1-2\kappa_1-8\eta_1-4\xi_2}\right) + 2p\exp\left(-\widetilde{C}n^{1-2\kappa_1-12\eta_1-4\xi_1}\right). \tag{D.16}$$

Next, consider $\max_{1\leq k\leq p} |T_{k2}|$. Recall that $T_{k2} = n^{-1}\sum_{i=1}^{n} X_{ik}^2 (x_{i,\mathcal{B}}^T \beta_{0,\mathcal{B}} + z_{i,\mathcal{I}}^T \gamma_{0,\mathcal{I}})^2 \mathbb{I}_{\Omega_i^c} \geq 0$. By Markov's inequality, for any $\delta > 0$, we have $P(|T_{k2}| \geq \delta) \leq \delta^{-1}E(|T_{k2}|) = \delta^{-1}E(T_{k2})$. In view of the first inequality in (D.14), taking $\delta = Cn^{-\kappa_1}/12$ leads to

$$P(|T_{k2}| \geq Cn^{-\kappa_1}/12) \leq 24C^{-1}C_0^2\widetilde{C}n^{\kappa_1}(n^{2\xi_1} + n^{2\xi_2})(1 + n^{\xi_2} + 2n^{\xi_1})^{1/2}\exp[-n^{\alpha_1\eta_1}/(2c_1)]$$

for all $1 \leq k \leq p$. Therefore,

$$P(\max_{1\leq k\leq p} |T_{k2}| \geq Cn^{-\kappa_1}/12) \leq \sum_{k=1}^{p} P(|T_{k2}| \geq Cn^{-\kappa_1}/12)$$
$$\leq 24pC^{-1}C_0^2\widetilde{C}n^{\kappa_1}(n^{2\xi_1} + n^{2\xi_2})(1 + n^{\xi_2} + 2n^{\xi_1})^{1/2}\exp[-n^{\alpha_1\eta_1}/(2c_1)]. \tag{D.17}$$

Combining (D.15), (D.16), and (D.17) yields that for sufficiently large n,

$$P(\max_{1\leq k\leq p} |S_{k1,1} - E(S_{k1,1})| \geq Cn^{-\kappa_1}/4)$$
$$\leq 2p\exp\left(-\widetilde{C}n^{1-2\kappa_1-8\eta_1-4\xi_2}\right) + 2p\exp\left(-\widetilde{C}n^{1-2\kappa_1-12\eta_1-4\xi_1}\right)$$
$$+ 24pC^{-1}C_0^2\widetilde{C}n^{\kappa_1}(n^{2\xi_1} + n^{2\xi_2})(1 + n^{\xi_2} + 2n^{\xi_1})^{1/2}\exp[-n^{\alpha_1\eta_1}/(2c_1)]. \tag{D.18}$$

To balance the three terms on the right hand side of (D.18), we choose $\eta_1 = \min\{(1 - 2\kappa_1 - 4\xi_2)/(8 + \alpha_1), (1 - 2\kappa_1 - 4\xi_1)/(12 + \alpha_1)\} > 0$ and the probability bound (D.18) becomes

$$P(\max_{1\leq k\leq p} |S_{k1,1} - E(S_{k1,1})| \geq Cn^{-\kappa_1}/4) \leq p\widetilde{C}_5\exp\left(-\widetilde{C}_6 n^{\alpha_1\eta_1}\right) \tag{D.19}$$

for all n sufficiently large, where $\widetilde{C}_5$ and $\widetilde{C}_6$ are two positive constants.

Step 2. We establish the probability bound for $\max_{1\le k\le p}|S_{k1,2}-E(S_{k1,2})|$. Define the event $\Psi_i=\{|X_{ij}|\le M_2 \text{ for all } j\in\mathcal{M}\cup\{k\}\}$ with $\mathcal{M}=\mathcal{A}\cup\mathcal{B}$ and let

$$T_{k3}=n^{-1}\sum_{i=1}^n X_{ik}^2(x_{i,\mathcal{B}}^T\beta_{0,\mathcal{B}}+z_{i,\mathcal{I}}^T\gamma_{0,\mathcal{I}})\varepsilon_i\mathbb{I}_{\Psi_i}\mathbb{I}(|\varepsilon_i|\le M_3),$$
$$T_{k4}=n^{-1}\sum_{i=1}^n X_{ik}^2(x_{i,\mathcal{B}}^T\beta_{0,\mathcal{B}}+z_{i,\mathcal{I}}^T\gamma_{0,\mathcal{I}})\varepsilon_i\mathbb{I}_{\Psi_i}\mathbb{I}(|\varepsilon_i|> M_3),$$
$$T_{k5}=n^{-1}\sum_{i=1}^n X_{ik}^2(x_{i,\mathcal{B}}^T\beta_{0,\mathcal{B}}+z_{i,\mathcal{I}}^T\gamma_{0,\mathcal{I}})\varepsilon_i\mathbb{I}_{\Psi_i^c},$$

where M_2 and M_3 are two large positive numbers which will be specified later. Then $S_{k1,2}=T_{k3}+T_{k4}+T_{k5}$. Similarly, $E(S_{k1,2})$ can be written as $E(S_{k1,2})=E(T_{k3})+E(T_{k4})+E(T_{k5})$. Since ε_1 has mean zero and is independent of $X_{1,1},\dots,X_{1,p}$, we have $E(T_{k5})=E[X_{1k}^2(x_{1,\mathcal{B}}^T\beta_{0,\mathcal{B}}+z_{1,\mathcal{I}}^T\gamma_{0,\mathcal{I}})\varepsilon_1\mathbb{I}_{\Psi_1^c}]=E[X_{1k}^2(x_{1,\mathcal{B}}^T\beta_{0,\mathcal{B}}+z_{1,\mathcal{I}}^T\gamma_{0,\mathcal{I}})\mathbb{I}_{\Psi_1^c}]E(\varepsilon_1)=0$. Thus $S_{k1,2}-E(S_{k1,2})$ can be expressed as

$$S_{k1,2}-E(S_{k1,2})=[T_{k3}-E(T_{k3})]+T_{k4}+T_{k5}-E(T_{k4}). \tag{D.20}$$

Note that $E(T_{k4})=E[X_{1k}^2(x_{1,\mathcal{B}}^T\beta_{0,\mathcal{B}}+z_{1,\mathcal{I}}^T\gamma_{0,\mathcal{I}})\varepsilon_1\mathbb{I}_{\Psi_1}\mathbb{I}(|\varepsilon_1|>M_3)]$. Thus

$$|E(T_{k4})|\le E[X_{1k}^2|x_{1,\mathcal{B}}^T\beta_{0,\mathcal{B}}+z_{1,\mathcal{I}}^T\gamma_{0,\mathcal{I}}|\mathbb{I}_{\Psi_1}|\varepsilon_1|\mathbb{I}(|\varepsilon_1|>M_3)].$$

It follows from the triangle inequality and Condition 6 that

$$X_{1k}^2|x_{1,\mathcal{B}}^T\beta_{0,\mathcal{B}}+z_{1,\mathcal{I}}^T\gamma_{0,\mathcal{I}}|\mathbb{I}_{\Psi_1}\le X_{1k}^2(|x_{1,\mathcal{B}}^T\beta_{0,\mathcal{B}}|+|z_{1,\mathcal{I}}^T\gamma_{0,\mathcal{I}}|)\mathbb{I}_{\Psi_1}\le C_0M_2^3(s_2+s_1M_2^2). \tag{D.21}$$

for all $1\le k\le p$ and some positive constant C_0. By the Cauchy-Schwarz inequality, Condition 2, and Lemma 3, we have

$$E[|\varepsilon_1|\mathbb{I}(|\varepsilon_1|>M_3)]\le[E(\varepsilon_1^2)P(|\varepsilon_1|>M_3)]^{1/2}\le\widetilde{C}\exp[-M_3^{\alpha_2}/(2c_1)]. \tag{D.22}$$

This together with the above inequalities entails that

$$|E(T_{k4})| \le C_0 M_2^3 (s_2 + s_1 M_2) E[|\varepsilon_1| \mathbb{I}(|\varepsilon_1| > M_3)] \le C_0 \widetilde{C} M_2^3 (s_2 + s_1 M_2) \exp[-M_3^{\alpha_2}/(2c_1)].$$

If we choose $M_2 = n^{\eta_2}$ and $M_3 = n^{\eta_3}$ with $\eta_2 > 0$ and $\eta_3 > 0$, then under Condition 6, for any positive constant C, when n is sufficiently large,

$$|E(T_{k4})| \le C_0 \widetilde{C} n^{3\eta_2} (n^{\xi_2} + n^{\xi_1 + \eta_2}) \exp[-n^{\alpha_2 \eta_3}/(2c_1)] \le C n^{-\kappa_1}/16$$

holds uniformly for all $1 \le k \le p$. This together with (D.20) ensures that

$$\begin{aligned} P(\max_{1\le k\le p} |S_{k1,2} - E(S_{k1,2})| \ge C n^{-\kappa_1}/4) \le P(\max_{1\le k\le p} |T_{k3} - E(T_{k3})| \ge C n^{-\kappa_1}/16) \\ + P(\max_{1\le k\le p} |T_{k4}| \ge C n^{-\kappa_1}/16) + P(\max_{1\le k\le p} |T_{k5}| \ge C n^{-\kappa_1}/16) \end{aligned} \tag{D.23}$$

for all n sufficiently large. In what follows, we will provide details on establishing the probability bound for each term on the right hand side of (D.23).

First consider $\max_{1\le k\le p} |T_{k3} - E(T_{k3})|$. In view of (D.21), we have $|X_{ik}^2 (x_{i,\mathcal{B}}^T \beta_{0,\mathcal{B}} + z_{i,\mathcal{I}}^T \gamma_{0,\mathcal{I}}) \varepsilon_i \mathbb{I}_{\Psi_i} \cdot \mathbb{I}(|\varepsilon_i| \le M_3)| \le C_0 M_2^3 M_3 (s_2 + s_1 M_2)$. For any $\delta > 0$, by Hoeffding's inequality, it holds that

$$\begin{aligned} P(|T_{k3} - E(T_{k3})| \ge \delta) \le& 2 \exp\left[-\frac{n\delta^2}{2C_0^2 M_2^6 M_3^2 (s_2 + s_1 M_2)^2} \right] \\ \le& 2 \exp\left[-\frac{n\delta^2}{4C_0^2 M_2^6 M_3^2 (s_2^2 + s_1^2 M_2^2)} \right] \\ \le& 2 \exp\left(-\frac{n\delta^2}{8C_0^2 M_2^6 M_3^2 s_2^2} \right) + 2 \exp\left(-\frac{n\delta^2}{8C_0^2 M_2^8 M_3^2 s_1^2} \right), \end{aligned}$$

where we have used the fact that $\exp[-c/(a+b)] \le \exp[-c/(2a)] + \exp[-c/(2b)]$ for any

$a, b, c > 0$. Recall that $M_2 = n^{\eta_2}$ and $M_3 = n^{\eta_3}$. Thus, taking $\delta = Cn^{-\kappa_1}/16$ gives

$$\begin{aligned}
&P(\max_{1\le k\le p} |T_{k3} - E(T_{k3})| \ge Cn^{-\kappa_1}/16) \le \sum_{k=1}^{p} P(|T_{k3} - E(T_{k3})| \ge Cn^{-\kappa_1}/16) \\
\le& 2p \exp\left(-\widetilde{C}n^{1-2\kappa_1-6\eta_2-2\eta_3-2\xi_2}\right) + 2p \exp\left(-\widetilde{C}n^{1-2\kappa_1-8\eta_2-2\eta_3-2\xi_1}\right). \qquad \text{(D.24)}
\end{aligned}$$

Next we handle $\max_{1\le k\le p} |T_{k4}|$. Using similar arguments as for proving (D.21), we have $X_{ik}^2 |x_{i,\mathcal{B}}^T \beta_{0,\mathcal{B}} + z_{i,\mathcal{I}}^T \gamma_{0,\mathcal{I}}| \mathbb{I}_{\Psi_i} \le C_0 M_2^3 (s_2 + s_1 M_2)$ for all $1 \le i \le n$ and $1 \le k \le p$ and thus

$$\max_{1\le k\le p} |T_{k4}| \le C_0 M_2^3 (s_2 + s_1 M_2) n^{-1} \sum_{i=1}^{n} |\varepsilon_i| \mathbb{I}(|\varepsilon_i| > M_3).$$

It follows from Markov's inequality and (D.22) that

$$\begin{aligned}
P(\max_{1\le k\le p} |T_{k4}| \ge \delta) \le& P\left\{C_0 M_2^3 (s_2 + s_1 M_2) n^{-1} \sum_{i=1}^{n} |\varepsilon_i| \mathbb{I}(|\varepsilon_i| > M_3) \ge \delta\right\} \\
\le& \delta^{-1} E\left[C_0 M_2^3 (s_2 + s_1 M_2) n^{-1} \sum_{i=1}^{n} |\varepsilon_i| \mathbb{I}(|\varepsilon_i| > M_3)\right] \\
=& \delta^{-1} C_0 M_2^3 (s_2 + s_1 M_2) E[|\varepsilon_1| \mathbb{I}(|\varepsilon_1| > M_3)] \\
\le& \delta^{-1} C_0 \widetilde{C} M_2^3 (s_2 + s_1 M_2) \exp[-M_3^{\alpha_2}/(2c_1)].
\end{aligned}$$

Recall that $M_2 = n^{\eta_2}$ and $M_3 = n^{\eta_3}$. Thus, taking $\delta = Cn^{-\kappa_1}/16$ results in

$$P(\max_{1\le k\le p} |T_{k4}| \ge Cn^{-\kappa_1}/16) \le 16C^{-1} C_0 \widetilde{C} n^{3\eta_2+\kappa_1} (n^{\xi_2} + n^{\xi_1+2\eta_2}) \exp[-n^{\alpha_2\eta_3}/(2c_1)]. \qquad \text{(D.25)}$$

We next consider $\max_{1\le k\le p} |T_{k5}|$. Since $|T_{k5}| \le n^{-1} \sum_{i=1}^{n} X_{ik}^2 |(x_{i,\mathcal{B}}^T \beta_{0,\mathcal{B}} + z_{i,\mathcal{I}}^T \gamma_{0,\mathcal{I}}) \varepsilon_i| \mathbb{I}_{\Psi_i^c}$, by Markov's inequality we have

$$
\begin{aligned}
P(|T_{k5}| \geq \delta) \leq & P\left\{n^{-1}\sum_{i=1}^{n} X_{ik}^2 |(x_{i,\mathcal{B}}^T\beta_{0,\mathcal{B}} + z_{i,\mathcal{I}}^T\gamma_{0,\mathcal{I}})\varepsilon_i| \mathbb{I}_{\Psi_i^c} \geq \delta\right\} \\
\leq & \delta^{-1}E\left[n^{-1}\sum_{i=1}^{n} X_{ik}^2 |(x_{i,\mathcal{B}}^T\beta_{0,\mathcal{B}} + z_{i,\mathcal{I}}^T\gamma_{0,\mathcal{I}})\varepsilon_i| \mathbb{I}_{\Psi_i^c}\right] \\
= & \delta^{-1}E[X_{1k}^2 |(x_{1,\mathcal{B}}^T\beta_{0,\mathcal{B}} + z_{1,\mathcal{I}}^T\gamma_{0,\mathcal{I}})\varepsilon_1| \mathbb{I}_{\Psi_1^c}].
\end{aligned}
$$

It follows from the Cauchy-Schwarz inequality and (D.13) that

$$
\begin{aligned}
& E[X_{1k}^2 |(x_{1,\mathcal{B}}^T\beta_{0,\mathcal{B}} + z_{1,\mathcal{I}}^T\gamma_{0,\mathcal{I}})\varepsilon_1| \mathbb{I}_{\Psi_i^c}] \leq \{E[X_{1k}^4 (x_{1,\mathcal{B}}^T\beta_{0,\mathcal{B}} + z_{1,\mathcal{I}}^T\gamma_{0,\mathcal{I}})^2 \varepsilon_1^2] P(\Psi_1^c)\}^{1/2} \\
\leq & \{2C_0^2 \left[s_2 E(X_{1k}^4 \|x_{1,\mathcal{B}}\|^2 \varepsilon_1^2) + s_1 E(X_{1k}^4 \|z_{1,\mathcal{I}}\|^2 \varepsilon_1^2)\right] P(\Psi_1^c)\}^{1/2}.
\end{aligned}
$$

Applying the Cauchy-Schwarz inequality again gives

$$
\begin{aligned}
E(X_{1k}^4 \|x_{1,\mathcal{B}}\|^2 \varepsilon_1^2) \leq & \left[E(X_{1k}^8 \|x_{1,\mathcal{B}}\|^4) E(\varepsilon_1^4)\right]^{1/2} \leq \left[s_2 \sum_{j\in\mathcal{B}} E(X_{1k}^8 X_{1j}^4)\right]^{1/2} \left[E(\varepsilon_1^4)\right]^{1/2} \\
\leq & \left\{2^{-1} s_2 \sum_{j\in\mathcal{B}} [E(X_{1k}^{16}) + E(X_{1j}^8)]\right\}^{1/2} \left[E(\varepsilon_1^4)\right]^{1/2} \leq \widetilde{C} s_2,
\end{aligned}
$$

where the last inequality follows from Condition 2 and Lemma 3. Similarly, we can show that $E(X_{1k}^4 \|z_{1,\mathcal{I}}\|^2 \varepsilon_1^2) \leq \widetilde{C} s_1$. By Condition 2 and the union bound, we deduce $P(\Psi_1^c) = P(|X_{ij}| > M_2 \text{ for some } j \in \mathcal{M} \cup \{k\}) \leq (1 + 2s_1 + s_2) c_1 \exp(-M_2^{\alpha_1}/c_1)$. This together with the above inequalities entails that

$$
P(|T_{k5}| \geq \delta) \leq \delta^{-1}\{2C_0^2 \widetilde{C}(s_1^2 + s_2^2)(1 + 2s_1 + s_2) c_1 \exp(-M_2^{\alpha_1}/c_1)\}^{1/2}.
$$

Recall that $M_2 = n^{\eta_2}$. Under Condition 6, taking $\delta = Cn^{-\kappa_1}/16$ yields

$$
\begin{aligned}
& P(\max_{1\leq k\leq p} |T_{k5}| \geq Cn^{-\kappa_1}/16) \leq \sum_{k=1}^{p} P(|T_{k5}| \geq Cn^{-\kappa_1}/16) \\
\leq & 16pC^{-1} n^{\kappa_1} \{2C_0^2 \widetilde{C} c_1 (n^{2\xi_1} + n^{2\xi_2})(1 + 2n^{\xi_1} + n^{\xi_2})\}^{1/2} \exp[-n^{\alpha_1\eta_2}/(2c_1)]. \qquad \text{(D.26)}
\end{aligned}
$$

Combining (D.23), (D.24), (D.25), and (D.26) yields that for sufficiently large n,

$$
\begin{aligned}
&P(\max_{1\le k\le p}|S_{k1,2}-E(S_{k1,2})|\ge Cn^{-\kappa_1}/4)\\
\le&2p\exp\left(-\widetilde{C}n^{1-2\kappa_1-6\eta_2-2\eta_3-2\xi_2}\right)+2p\exp\left(-\widetilde{C}n^{1-2\kappa_1-8\eta_2-2\eta_3-2\xi_1}\right)\\
&+16pC^{-1}n^{\kappa_1}\{2C_0^2\widetilde{C}c_1(n^{2\xi_1}+n^{2\xi_2})(1+2n^{\xi_1}+n^{\xi_2})\}^{1/2}\exp[-n^{\alpha_1\eta_2}/(2c_1)]\\
&+16C^{-1}C_0\widetilde{C}n^{3\eta_2+\kappa_1}(n^{\xi_2}+n^{\xi_1+\eta_2})\exp[-n^{\alpha_2\eta_3}/(2c_1)]. \qquad (D.27)
\end{aligned}
$$

Let $\eta_2=\eta_3=\min\{(1-2\kappa_1-2\xi_2)/(8+\alpha_1),(1-2\kappa_1-2\xi_1)/(10+\alpha_1)\}$. Then (D.27) becomes

$$
P(\max_{1\le k\le p}|S_{k1,2}-E(S_{k1,2})|\ge Cn^{-\kappa_1}/4)\le p\widetilde{C}_7\exp\left(-\widetilde{C}_8n^{\alpha_1\eta_2}\right)+\widetilde{C}_9\exp[-\widetilde{C}_{10}n^{\alpha_2\eta_2}]. \qquad (D.28)
$$

for all n sufficiently large, where $\widetilde{C}_7$, $\widetilde{C}_8$, $\widetilde{C}_9$, and $\widetilde{C}_{10}$ are some positive constants.

Step 3. We establish the probability bound for $\max_{1\le k\le p}|S_{k1,3}-E(S_{k1,3})|$. Define

$$
\begin{aligned}
T_{k6}&=n^{-1}\sum_{i=1}^{n}X_{ik}^2\varepsilon_i^2\mathbb{I}(|X_{ik}|\le M_4)\mathbb{I}(|\varepsilon_i|\le M_5),\\
T_{k7}&=n^{-1}\sum_{i=1}^{n}X_{ik}^2\varepsilon_i^2\mathbb{I}(|X_{ik}|\le M_4)\mathbb{I}(|\varepsilon_i|> M_5),\\
T_{k8}&=n^{-1}\sum_{i=1}^{n}X_{ik}^2\varepsilon_i^2\mathbb{I}(|X_{ik}|> M_4),
\end{aligned}
$$

where M_4 and M_5 are two large positive numbers whose values will be specified later. Then $S_{k1,3}=T_{k6}+T_{k7}+T_{k8}$. Similarly, $E(S_{k1,3})$ can be written as $E(S_{k1,3})=E(T_{k6})+E(T_{k7})+E(T_{k8})$ with $E(T_{k6})=E[X_{1k}^2\varepsilon_1^2\mathbb{I}(|X_{1k}|\le M_4)\mathbb{I}(|\varepsilon_1|\le M_5)]$, $E(T_{k7})=E[X_{1k}^2\varepsilon_1^2\mathbb{I}(|X_{1k}|\le M_4)\mathbb{I}(|\varepsilon_1|> M_5)]$, and $E(T_{k8})=E[X_{1k}^2\varepsilon_1^2\mathbb{I}(|X_{1k}|> M_4)]$. Thus $S_{k1,3}-E(S_{k1,3})$ can be expressed as

$$
S_{k1,3}-E(S_{k1,3})=[T_{k6}-E(T_{k6})]+T_{k7}+T_{k8}-[E(T_{k7})+E(T_{k8})]. \qquad (D.29)
$$

First consider the last two terms $E(T_{k7})$ and $E(T_{k8})$. It follows from $0 \leq X_{1k}^2 \varepsilon_1^2 \mathbb{I}(|X_{1k}| \leq M_4)\mathbb{I}(|\varepsilon_1| > M_5) \leq M_4^2 \varepsilon_1^2 \mathbb{I}(|\varepsilon_1| > M_5)$ that

$$0 \leq E(T_{k7}) \leq M_4^2 E[\varepsilon_1^2 \mathbb{I}(|\varepsilon_1| > M_5)]. \tag{D.30}$$

An application of the Cauchy-Schwarz inequality leads to $E[\varepsilon_1^2 \mathbb{I}(|\varepsilon_1| > M_5)] \leq [E(\varepsilon_1^4) P(|\varepsilon_1| > M_5)]^{1/2}$. By Condition 2 and Lemma 3, we have

$$E[\varepsilon_1^2 \mathbb{I}(|\varepsilon_1| > M_5)] \leq \{E(\varepsilon_1^4) c_1\}^{1/2} \exp(-c_1^{-1} M_5^{\alpha_2}/2) \leq \widetilde{C} \exp[-M_5^{\alpha_2}/(2c_1)] \tag{D.31}$$

Combining (D.30) with (D.31) yields

$$|E(T_{k7})| \leq \widetilde{C} M_4^2 \exp[-M_5^{\alpha_2}/(2c_1)]. \tag{D.32}$$

Similarly, by the Cauchy-Schwarz inequality and Lemma 3 we obtain

$$\begin{aligned} |E(T_{k8})| &= E[X_{1k}^2 \varepsilon_1^2 \mathbb{I}(|X_{1k}| > M_4)] \leq \{E(X_{1k}^4 \varepsilon_1^4) P(|X_{1k}| > M_4)]\}^{1/2} \\ &\leq \Big\{\frac{c_1}{2}[E(X_{1k}^8) + E(\varepsilon_1^8)]\Big\}^{1/2} \exp[-M_4^{\alpha_1}/(2c_1)] \leq \widetilde{C} \exp[-M_4^{\alpha_1}/(2c_1)]. \end{aligned} \tag{D.33}$$

Combining (D.32) and (D.33) results in

$$|E(T_{k7}) + E(T_{k8})| \leq \widetilde{C} M_4^2 \exp[-M_5^{\alpha_2}/(2c_1)] + \widetilde{C} \exp[-M_4^{\alpha_1}/(2c_1)].$$

If we choose $M_4 = n^{\eta_4}$ and $M_5 = n^{\eta_5}$ with $\eta_4 > 0$ and $\eta_5 > 0$, then for any positive constant C, when n is sufficiently large,

$$|E(T_{k7}) + E(T_{k8})| \leq \widetilde{C} n^{2\eta_4} \exp[-n^{\alpha_2 \eta_5}/(2c_1)] + \widetilde{C} \exp[-n^{\alpha_1 \eta_4}/(2c_1)] < Cn^{-\kappa_1}/16$$

holds uniformly for all $1 \leq k \leq p$. The above inequality together with (D.29) ensures that

$$
\begin{aligned}
&P(\max_{1\le k\le p}|S_{k1,3}-E(S_{k1,3})|\ge Cn^{-\kappa_1}/4)\\
&\le P(\max_{1\le k\le p}|T_{k6}-E(T_{k6})|\ge Cn^{-\kappa_1}/16)+P(\max_{1\le k\le p}|T_{k7}|\ge Cn^{-\kappa_1}/16)\\
&\quad+P(\max_{1\le k\le p}|T_{k8}|\ge Cn^{-\kappa_1}/16) \qquad \text{(D.34)}
\end{aligned}
$$

for all n sufficiently large.

In what follows, we will provide details on establishing the probability bound for each term on the right hand side of (D.34). First consider $\max_{1\le k\le p}|T_{k6}-E(T_{k6})|$. Since $0\le X_{ik}^2\varepsilon_i^2\mathbb{I}(|X_{ik}|\le M_4)\mathbb{I}(|\varepsilon_i|\le M_5)\le M_4^2M_5^2$, by Hoeffding's inequality, we have for any $\delta>0$ that

$$
P(|T_{k6}-E(T_{k6})|\ge\delta)\le 2\exp\left(-\frac{2n\delta^2}{M_4^4M_5^4}\right)=2\exp\left(-2n^{1-4\eta_4-4\eta_5}\delta^2\right),
$$

by noting that $M_4=n^{\eta_4}$ and $M_5=n^{\eta_5}$. Thus, taking $\delta=Cn^{-\kappa_1}/16$ gives

$$
\begin{aligned}
&P(\max_{1\le k\le p}|T_{k6}-E(T_{k6})|\ge Cn^{-\kappa_1}/16)\le\sum_{k=1}^{p}P(|T_{k6}-E(T_{k6})|\ge Cn^{-\kappa_1}/16)\\
&\le 2p\exp\left(-\widetilde{C}n^{1-2\kappa_1-4\eta_4-4\eta_5}\right). \qquad \text{(D.35)}
\end{aligned}
$$

Next we handle $\max_{1\le k\le p}|T_{k7}|$. Since $\max_{1\le k\le p}|T_{k7}|\le n^{-1}M_4^2\sum_{i=1}^n\varepsilon_i^2\mathbb{I}(|\varepsilon_i|>M_5)$, it follows from Markov's inequality and (D.31) that for any $\delta>0$,

$$
\begin{aligned}
P(\max_{1\le k\le p}|T_{k7}|\ge\delta)\le&P\{n^{-1}M_4^2\sum_{i=1}^{n}\varepsilon_i^2\mathbb{I}(|\varepsilon_i|>M_5)\ge\delta\}\\
\le&\delta^{-1}E[n^{-1}M_4^2\sum_{i=1}^{n}\varepsilon_i^2\mathbb{I}(|\varepsilon_i|>M_5)]\\
=&\delta^{-1}M_4^2E[\varepsilon_1^2\mathbb{I}(|\varepsilon_1|>M_5)]\le\widetilde{C}\delta^{-1}M_4^2\exp[-M_5^{\alpha_2}/(2c_1)].
\end{aligned}
$$

Recall that $M_4=n^{\eta_4}$ and $M_5=n^{\eta_5}$. Setting $\delta=Cn^{-\kappa_1}/16$ in the above inequality entails

$$P(\max_{1\le j\le p}|T_{k7}|\ge Cn^{-\kappa_1}/16)\le 16C^{-1}\widetilde{C}n^{2\eta_4+\kappa_1}\exp[-n^{\alpha_2\eta_5}/(2c_1)]. \tag{D.36}$$

We then consider $\max_{1\le k\le p}|T_{k8}|$. By Markov's inequality and (D.33), for any $\delta>0$,

$$\begin{aligned}P(|T_{k8}|\ge\delta)&\le\delta^{-1}E[n^{-1}\sum_{i=1}^{n}X_{ik}^2\varepsilon_i^2\mathbb{I}(|X_{ik}|>M_4)]=\delta^{-1}E[X_{1k}^2\varepsilon_1^2\mathbb{I}(|X_{1k}|>M_4)]\\&\le\delta^{-1}\widetilde{C}\exp[-M_4^{\alpha_1}/(2c_1)]. \end{aligned}\tag{D.37}$$

Recall that $M_4=n^{\eta_1}$. In view of (D.37), taking $\delta=Cn^{-\kappa_1}/16$ leads to

$$P(\max_{1\le k\le p}|T_{k8}|\ge Cn^{-\kappa_1}/16)\le\sum_{k=1}^{p}P(|T_{k8}|\ge Cn^{-\kappa_1}/16)\le\ 16pC^{-1}\widetilde{C}n^{\kappa_1}\exp[-n^{2\eta_4}/(2c_1)]. \tag{D.38}$$

Combining (D.34), (D.35), (D.36) with (D.38) yields that for sufficiently large n,

$$\begin{aligned}&P(\max_{1\le k\le p}|S_{k1,3}-E(S_{k1,3})|\ge Cn^{-\kappa_1}/4)\le 2p\exp\left(-\widetilde{C}n^{1-2\kappa_1-4\eta_4-4\eta_5}\right)\\&\quad+16pC^{-1}\widetilde{C}n^{\kappa_1}\exp[-n^{\alpha_1\eta_4}/(2c_1)]+16C^{-1}\widetilde{C}n^{2\eta_4+\kappa_1}\exp[-n^{\alpha_2\eta_5}/(2c_1)]. \end{aligned}\tag{D.39}$$

Let $\eta_4=\eta_5=(1-2\kappa_1)/(8+\alpha_1)$. Then (D.39) becomes

$$P(\max_{1\le k\le p}|S_{k1,3}-E(S_{k1,3})|\ge Cn^{-\kappa_1}/4)\le p\widetilde{C}_{11}\exp[-\widetilde{C}_{12}n^{\alpha_1\eta_4}]+\widetilde{C}_{13}\exp[-\widetilde{C}_{14}n^{\alpha_2\eta_4}] \tag{D.40}$$

for all n sufficiently large, where $\widetilde{C}_{11}$, $\widetilde{C}_{12}$, $\widetilde{C}_{13}$, and $\widetilde{C}_{14}$ are some positive constants.

Since $0<\eta_1<\eta_2=\eta_3$ and $\eta_1\le\eta_4$, it follows from (D.11), (D.19), (D.28), and (D.40) that there exist some positive constants $\widetilde{C}_1,\dots,\widetilde{C}_4$ such that

$$P(\max_{1\le k\le p}|S_{k1}-E(S_{k1})|\ge Cn^{-\kappa_1})\le p\widetilde{C}_1\exp\left(-\widetilde{C}_2n^{\alpha_1\eta_1}\right)+\widetilde{C}_3\exp\left(-\widetilde{C}_4n^{\alpha_2\eta_1}\right)$$

for all n sufficiently large. This concludes the proof of part a) of Lemma 1.

D.3 Proof of part b) of Lemma 1

We recall that $\omega_j^* = E(X_j Y)$ and $\widehat{\omega}_j^* = n^{-1} \sum_{i=1}^n X_{ij} Y_i$. Note that $Y_i = \beta_0 + x_i^T \beta_0 + z_i^T \gamma_0 + \varepsilon_i = \beta_0 + x_{i,\mathcal{B}}^T \beta_{0,\mathcal{B}} + z_{i,\mathcal{I}}^T \gamma_{0,\mathcal{I}} + \varepsilon_i$, where $x_i = (X_{i1}, \ldots, X_{ip})^T$, $z_i = (X_{i1}X_{i2}, \ldots, X_{i,p-1}X_{i,p})^T$, $x_{i,\mathcal{B}} = (X_{i\ell}, \ell \in \mathcal{B})^T$, $z_{i,\mathcal{I}} = (X_{ik}X_{i\ell}, (k,\ell) \in \mathcal{I})^T$, $\boldsymbol{\beta}_{0,\mathcal{B}} = (\beta_\ell^0, \ell \in \mathcal{B})^T$, and $\boldsymbol{\gamma}_{0,\mathcal{I}} = (\gamma_{k\ell}, (k,\ell) \in \mathcal{I})^T$. To simplify the proof, we assume that the intercept α_0 is zero without loss of generality. Thus

$$
\begin{aligned}
\widehat{\omega}_j^* &= n^{-1} \sum_{i=1}^n X_{ij} Y_i = n^{-1} \sum_{i=1}^n X_{ij} (x_{i,\mathcal{B}}^T \beta_{0,\mathcal{B}} + z_{i,\mathcal{I}}^T \gamma_{0,\mathcal{I}}) + n^{-1} \sum_{i=1}^n X_{ij} \varepsilon_i \\
&\triangleq S_{j1} + S_{j2}.
\end{aligned}
$$

Similarly, ω_j^* can be written as $\omega_j^* = E(X_j Y) = E(S_{j1}) + E(S_{j2})$. So $\widehat{\omega}_j^* - \omega_j^*$ can be expressed as $\widehat{\omega}_j^* - \omega_j^* = [S_{j1} - E(S_{j1})] + [S_{j2} - E(S_{j2})]$. By the triangle inequality and the union bound, it holds that

$$
\begin{aligned}
&P(\max_{1 \le j \le p} |\widehat{\omega}_j^* - \omega_j^*| \ge C n^{-\kappa_2}) \\
\le &P(\max_{1 \le j \le p} |S_{j1} - E(S_{j1})| \ge C n^{-\kappa_2}/2) + P(\max_{1 \le j \le p} |S_{j2} - E(S_{j2})| \ge C n^{-\kappa_2}/2). \qquad \text{(D.41)}
\end{aligned}
$$

In what follows, we will provide details on deriving an exponential tail probability bound for each term on the right hand side above. To enhance readability, we split the proof into two steps.

Step 1. We start with the first term $\max_{1 \le k \le p} |S_{j1} - E(S_{j1})|$. Define the event $\Phi_i = \{|X_{i\ell}| \le M_6 \text{ for all } \ell \in \mathcal{M} \cup \{j\}\}$ with $\mathcal{M} = \mathcal{A} \cup \mathcal{B}$ and M_6 a large positive number that will be specified later. Let $T_{j1} = n^{-1} \sum_{i=1}^n X_{ij} (x_{i,\mathcal{B}}^T \beta_{0,\mathcal{B}} + z_{i,\mathcal{I}}^T \gamma_{0,\mathcal{I}}) \mathbb{I}_{\Phi_i}$ and $T_{j2} = n^{-1} \sum_{i=1}^n X_{ij} (x_{i,\mathcal{B}}^T \beta_{0,\mathcal{B}} + z_{i,\mathcal{I}}^T \gamma_{0,\mathcal{I}}) \mathbb{I}_{\Phi_i^c}$, where $\mathbb{I}(\cdot)$ is the indicator function and Φ_i^c is the complement of the set Φ_i. Then an application of the triangle inequality yields

$$
\begin{aligned}
|S_{j1} - E(S_{j1})| =& |[T_{j1} - E(T_{j1})] + T_{j2} - E(T_{j2})| \leq |T_{j1} - E(T_{j1})| + |T_{j2}| + |E(T_{j2})| \\
\leq& |T_{j1} - E(T_{j1})| + |T_{j2}| + E(|T_{j2}|). \qquad \text{(D.42)}
\end{aligned}
$$

Note that $|T_{j2}| \leq n^{-1} \sum_{i=1}^{n} |X_{ij}(x_{i,\mathcal{B}}^T \beta_{0,\mathcal{B}} + z_{i,\mathcal{I}}^T \gamma_{0,\mathcal{I}})| \mathbb{I}_{\Phi_i^c}$ and thus $E(|T_{j2}|) \leq E[|X_{1j}(x_{1,\mathcal{B}}^T \beta_{0,\mathcal{B}} + z_{1,\mathcal{I}}^T \gamma_{0,\mathcal{I}})| \mathbb{I}_{\Phi_1^c}]$. By the triangle inequality and Condition 6, we have

$$
|X_{1j}(x_{1,\mathcal{B}}^T \beta_{0,\mathcal{B}} + z_{1,\mathcal{I}}^T \gamma_{0,\mathcal{I}})| \leq C_0(|X_{1j}| \|x_{1,\mathcal{B}}\|_1 + |X_{1j}| \|z_{1,\mathcal{I}}\|_1), \qquad \text{(D.43)}
$$

which ensures that $E(|T_{j2}|)$ is bounded by $C_0[E(|X_{1j}| \|x_{1,\mathcal{B}}\|_1 \mathbb{I}_{\Omega_1^c}) + E(|X_{1j}| \|z_{1,\mathcal{I}}\|_1 \mathbb{I}_{\Omega_1^c})]$. Here $\|\cdot\|_1$ is the L_1 norm. By the Cauchy-Schwarz inequality and the triangular inequality, we deduce

$$
\begin{aligned}
E(|X_{1j}| \|x_{1,\mathcal{B}}\|_1 \mathbb{I}_{\Phi_1^c}) \leq& \left[E(X_{1j}^2 \|x_{1,\mathcal{B}}\|_1^2) P(\Phi_1^c)\right]^{1/2} \leq \left\{ \left[s_2 \sum_{\ell \in \mathcal{B}} E(X_{1j}^2 X_{1\ell}^2) \right] P(\Phi_1^c) \right\}^{1/2} \\
\leq& \left\{ 2^{-1} s_2 \sum_{\ell \in \mathcal{B}} [E(X_{1j}^4) + E(X_{1\ell}^4)] \right\}^{1/2} \left[\sum_{\ell \in \mathcal{M} \cup \{j\}} P(|X_{i\ell}| > M_6) \right]^{1/2} \\
\leq& \widetilde{C} s_2 (1 + s_2 + 2s_1)^{1/2} \exp[-M_6^{\alpha_1}/(2c_1)]
\end{aligned}
$$

for some positive constant $\widetilde{C}$, where the last inequality follows from Condition 2 and Lemma 3. Similarly, we have $E(|X_{1j}| \|z_{1,\mathcal{I}}\|_1 \mathbb{I}_{\Phi_1^c}) \leq \widetilde{C} s_1 (1 + s_2 + 2s_1)^{1/2} \exp[-M_6^{\alpha_1}/(2c_1)]$. This together with the above inequalities entails that

$$
E(|T_{j2}|) \leq C_0 \widetilde{C} (s_1 + s_2)(1 + s_2 + 2s_1)^{1/2} \exp[-M_6^{\alpha_1}/(2c_1)].
$$

If we choose $M_6 = n^{\eta_6}$ with $\eta_6 > 0$, then by Condition 6, for any positive constant C, when n is sufficiently large,

$$
E(|T_{j2}|) \leq C_0 \widetilde{C} (n^{\xi_1} + n^{\xi_2})(1 + n^{\xi_2} + 2n^{\xi_1})^{1/2} \exp[-n^{\alpha_1 \eta_6}/(2c_1)] < C n^{-\kappa_2}/6 \qquad \text{(D.44)}
$$

holds uniformly for all $1 \leq j \leq p$. The above inequality together with (D.42) ensures that

$$
\begin{aligned}
&P(\max_{1\leq j\leq p} |S_{j1} - E(S_{j1})| \geq Cn^{-\kappa_2}/2) \\
\leq &P(\max_{1\leq j\leq p} |T_{j1} - E(T_{j1})| \geq Cn^{-\kappa_2}/6) + P(\max_{1\leq j\leq p} |T_{j2}| \geq Cn^{-\kappa_2}/6) \qquad \text{(D.45)}
\end{aligned}
$$

for all n is sufficiently large. Thus we only need to establish the probability bound for each term on the right hand side of (D.45).

First consider $\max_{1\leq j\leq p} |T_{j1} - E(T_{j1})|$. Using similar arguments as for proving (D.43), we have

$$
|X_{ij}(x_{i,\mathcal{B}}^T \beta_{0,\mathcal{B}} + z_{i,\mathcal{I}}^T \gamma_{0,\mathcal{I}})\mathbb{I}_{\Phi_i}| \leq C_0(|X_{ij}|\|x_{i,\mathcal{B}}\|_1 + |X_{ij}|\|z_{i,\mathcal{I}}\|_1)\mathbb{I}_{\Phi_i} \leq C_0(s_2 M_6^2 + s_1 M_6^3).
$$

For any $\delta > 0$, an application of Hoeffding's inequality gives

$$
\begin{aligned}
P(|T_{j1} - E(T_{j1})| \geq \delta) \leq &2\exp\left[-\frac{n\delta^2}{2C_0^2 M_6^4 (s_2 + s_1 M_6)^2}\right] \\
\leq &2\exp\left[-\frac{n\delta^2}{4C_0^2 M_6^4 (s_2^2 + s_1^2 M_6^2)}\right] \\
\leq &2\exp\left(-\frac{n\delta^2}{8C_0^2 M_6^4 s_2^2}\right) + 2\exp\left(-\frac{n\delta^2}{8C_0^2 M_6^6 s_1^2}\right),
\end{aligned}
$$

where we have used the fact that $(a+b)^2 \leq 2(a^2+b^2)$ for any real numbers a and b, and $\exp[-c/(a+b)] \leq \exp[-c/(2a)] + \exp[-c/(2b)]$ for any $a, b, c > 0$. Recall that $M_6 = n^{\eta_6}$. Under Condition 6, taking $\delta = Cn^{-\kappa_2}/6$ results in

$$
\begin{aligned}
&P(\max_{1\leq j\leq p} |T_{j1} - E(T_{j1})| \geq Cn^{-\kappa_2}/6) \leq \sum_{j=1}^{p} P(|T_{j1} - E(T_{j1})| \geq Cn^{-\kappa_2}/6) \\
\leq &2p\exp\left(-\widetilde{C}n^{1-2\kappa_2-4\eta_6-2\xi_2}\right) + 2p\exp\left(-\widetilde{C}n^{1-2\kappa_2-6\eta_6-2\xi_1}\right). \qquad \text{(D.46)}
\end{aligned}
$$

Next, consider $\max_{1\leq j\leq p} |T_{j2}|$. By Markov's inequality, for any $\delta > 0$, we have $P(|T_{j2}| \geq \delta) \leq \delta^{-1}E(|T_{j2}|)$. In view of the first inequality in (D.44), taking $\delta = Cn^{-\kappa_2}/6$

gives that

$$P(|T_{j2}| \geq Cn^{-\kappa_2}/6) \leq 6C^{-1}C_0\widetilde{C}n^{\kappa_2}(n^{\xi_1}+n^{\xi_2})(1+n^{\xi_2}+2n^{\xi_1})^{1/2}\exp[-n^{\alpha_1\eta_6}/(2c_1)]$$

for all $1 \leq j \leq p$. Therefore,

$$\begin{aligned} &P(\max_{1\leq j\leq p}|T_{j2}| \geq Cn^{-\kappa_2}/6) \leq \sum_{j=1}^{p} P(|T_{j2}| \geq Cn^{-\kappa_2}/6) \\ \leq & 6pC^{-1}C_0\widetilde{C}n^{\kappa_2}(n^{\xi_1}+n^{\xi_2})(1+n^{\xi_2}+2n^{\xi_1})^{1/2}\exp[-n^{\alpha_1\eta_6}/(2c_1)]. \end{aligned} \tag{D.47}$$

Combining (D.45), (D.46), and (D.47) yields that for sufficiently large n,

$$\begin{aligned} &P(\max_{1\leq j\leq p}|S_{j1}-E(S_{j1})| \geq Cn^{-\kappa_2}/2) \\ &\leq 2p\exp\left(-\widetilde{C}n^{1-2\kappa_2-4\eta_6-2\xi_2}\right) + 2p\exp\left(-\widetilde{C}n^{1-2\kappa_2-6\eta_6-2\xi_1}\right) \\ &\quad + 6pC^{-1}C_0\widetilde{C}n^{\kappa_2}(n^{\xi_1}+n^{\xi_2})(1+n^{\xi_2}+2n^{\xi_1})^{1/2}\exp[-n^{\alpha_1\eta_6}/(2c_1)]. \end{aligned} \tag{D.48}$$

To balance the three terms on the right hand side of (D.48), we choose $\eta_6 = \min\{(1-2\kappa_2-2\xi_2)/(4+\alpha_1), (1-2\kappa_2-2\xi_1)/(6+\alpha_1)\} > 0$ and the probability bound (D.48) then becomes

$$P(\max_{1\leq j\leq p}|S_{j1}-E(S_{j1})| \geq Cn^{-\kappa_2}/2) \leq p\widetilde{C}_1\exp\left(-\widetilde{C}_2 n^{\alpha_1\eta_6}\right) \tag{D.49}$$

for all n sufficiently large, where $\widetilde{C}_1$ and $\widetilde{C}_2$ are two positive constants.

Step 2. We establish the probability bound for $\max_{1\leq j\leq p}|S_{j2}-E(S_{j2})|$. Define

$$\begin{aligned} T_{j3} &= n^{-1}\sum_{i=1}^{n} X_{ij}\varepsilon_i \mathbb{I}(|X_{ij}| \leq M_7)\mathbb{I}(|\varepsilon_i| \leq M_8), \\ T_{j4} &= n^{-1}\sum_{i=1}^{n} X_{ij}\varepsilon_i \mathbb{I}(|X_{ij}| \leq M_7)\mathbb{I}(|\varepsilon_i| > M_8), \\ T_{j5} &= n^{-1}\sum_{i=1}^{n} X_{ij}\varepsilon_i \mathbb{I}(|X_{ij}| > M_7), \end{aligned}$$

where M_7 and M_8 are two large positive numbers whose values will be specified later. Then $S_{j2} = T_{j3} + T_{j4} + T_{j5}$. Similarly, $E(S_{j2})$ can be written as $E(S_{j2}) = E(T_{j3}) + E(T_{j4}) + E(T_{j5})$. Since ε_1 has mean zero and is independent of $X_{1,1}, \ldots, X_{1,p}$, we have $E(T_{j5}) = E[X_{1j}\varepsilon_1 \mathbb{I}(|X_{1j}| > M_7)] = E[X_{1j}\mathbb{I}(|X_{1j}| > M_7)]E(\varepsilon_1) = 0$. Thus $S_{j2} - E(S_{j2})$ can be expressed as $S_{j2} - E(S_{j2}) = [T_{j3} - E(T_{j3})] + T_{j4} + T_{j5} - E(T_{j4})$. An application of the triangle inequality yields

$$\begin{aligned} |S_{j2} - E(S_{j2})| \leq & |T_{j3} - E(T_{j3})| + |T_{j4}| + |T_{j5}| + |E(T_{j4})| \\ \leq & |T_{j3} - E(T_{j3})| + |T_{j4}| + |T_{j5}| + E(|T_{j4}|). \end{aligned} \tag{D.50}$$

First consider the last term $E(|T_{j4}|)$. Note that $|T_{j4}| \leq n^{-1}\sum_{i=1}^{n} |X_{ij}\varepsilon_i|\mathbb{I}(|X_{ij}| \leq M_7)\mathbb{I}(|\varepsilon_i| > M_8)$ and thus

$$E(|T_{j4}|) \leq E[|X_{1j}\varepsilon_1|\mathbb{I}(|X_{1j}| \leq M_7)\mathbb{I}(|\varepsilon_1| > M_8)] \leq M_7 E[|\varepsilon_1|\mathbb{I}(|\varepsilon_1| > M_8)]. \tag{D.51}$$

An application of the Cauchy-Schwarz inequality gives $E[|\varepsilon_1|\mathbb{I}(|\varepsilon_1| > M_8)] \leq [E(\varepsilon_1^2)P(|\varepsilon_1| > M_8)]^{1/2}$. By Condition 2 and Lemma 3, we have

$$E[|\varepsilon_1|\mathbb{I}(|\varepsilon_1| > M_8)] \leq \{E(\varepsilon_1^2)c_1\}^{1/2} \exp(-c_1^{-1}M_8^{\alpha_2}/2) \leq \widetilde{C} \exp[-M_8^{\alpha_2}/(2c_1)] \tag{D.52}$$

Combining (D.51) with (D.52) yields

$$E(|T_{j4}|) \leq \widetilde{C} M_7 \exp[-M_8^{\alpha_2}/(2c_1)]. \tag{D.53}$$

If we choose $M_7 = n^{\eta_7}$ and $M_8 = n^{\eta_8}$ with $\eta_7 > 0$ and $\eta_8 > 0$, then for any positive constant C, when n is sufficiently large,

$$E(|T_{j4}|) \leq \widetilde{C} n^{\eta_7} \exp[-n^{\alpha_2\eta_8}/(2c_1)] < Cn^{-\kappa_2}/8$$

holds uniformly for all $1 \leq j \leq p$. The above inequality together with (D.50) ensures that

$$
\begin{aligned}
&P(\max_{1\le j\le p}|S_{j2}-E(S_{j2})|\ge Cn^{-\kappa_2}/2)\\
\le&P(\max_{1\le j\le p}|T_{j3}-E(T_{j3})|\ge Cn^{-\kappa_2}/8)+P(\max_{1\le j\le p}|T_{j4}|\ge Cn^{-\kappa_2}/8)\\
&+P(\max_{1\le j\le p}|T_{j5}|\ge Cn^{-\kappa_2}/8) \qquad \text{(D.54)}
\end{aligned}
$$

for all n sufficiently large.

In what follows, we will provide details on establishing the probability bound for each term on the right hand side of (D.54). First consider $\max_{1\le j\le p}|T_{j3}-E(T_{j3})|$. Since $|X_{ij}\varepsilon_i\mathbb{I}(|X_{ij}|\le M_7)\mathbb{I}(|\varepsilon_i|\le M_8)|\le M_7M_8$, for any $\delta>0$, by Hoeffding's inequality we obtain

$$
P(|T_{j3}-E(T_{j3})|\ge\delta)\le 2\exp\left(-\frac{n\delta^2}{2M_7^2M_8^2}\right)=2\exp\left(-2^{-1}n^{1-2\eta_7-2\eta_8}\delta^2\right),
$$

by noting that $M_7=n^{\eta_7}$ and $M_8=n^{\eta_8}$. Thus, taking $\delta=Cn^{-\kappa_2}/8$ gives

$$
\begin{aligned}
&P(\max_{1\le j\le p}|T_{j3}-E(T_{j3})|\ge Cn^{-\kappa_2}/8)\le\sum_{j=1}^{p}P(|T_{j3}-E(T_{j3})|\ge Cn^{-\kappa_2}/8)\\
\le&2p\exp\left(-\widetilde{C}n^{1-2\kappa_2-2\eta_7-2\eta_8}\right). \qquad \text{(D.55)}
\end{aligned}
$$

Next we handle $\max_{1\le j\le p}|T_{j4}|$. Since $\max_{1\le j\le p}|T_{j4}|\le n^{-1}M_7\sum_{i=1}^n|\varepsilon_i|\mathbb{I}(|\varepsilon_i|>M_8)$, it follows from Markov's inequality and (D.52) that for any $\delta>0$,

$$
\begin{aligned}
P(\max_{1\le j\le p}|T_{j4}|\ge\delta)\le&P\{n^{-1}M_7\sum_{i=1}^{n}|\varepsilon_i|\mathbb{I}(|\varepsilon_i|>M_8)\ge\delta\}\\
\le&\delta^{-1}E[n^{-1}M_7\sum_{i=1}^{n}|\varepsilon_i|\mathbb{I}(|\varepsilon_i|>M_8)]\\
=&\delta^{-1}M_7E[|\varepsilon_1|\mathbb{I}(|\varepsilon_1|>M_8)]\le\widetilde{C}\delta^{-1}M_7\exp[-M_8^{\alpha_2}/(2c_1)].
\end{aligned}
$$

Recall that $M_7 = n^{\eta_7}$ and $M_8 = n^{\eta_8}$. Setting $\delta = Cn^{-\kappa_2}/8$ in the above inequality entails

$$P(\max_{1\leq j\leq p} |T_{j4}| \geq Cn^{-\kappa_2}/8) \leq 16C^{-1}\widetilde{C}n^{\eta_7+\kappa_2}\exp[-n^{\alpha_2\eta_8}/(2c_1)]. \tag{D.56}$$

We now consider $\max_{1\leq j\leq p}|T_{j5}|$. By the Cauchy-Schwarz inequality and Lemma 3 we deduce that

$$\begin{aligned}
E|T_{j5}| &= E|X_{1j}\varepsilon_1 \mathbb{I}(|X_{1j}| > M_7)| \leq \{E(X_{1j}^2\varepsilon_1^2)P(|X_{1j}| > M_7)\}^{1/2} \\
&\leq \Big\{\frac{c_1}{2}[E(X_{1k}^4) + E(\varepsilon_1^4)]\Big\}^{1/2}\exp[-M_7^{\alpha_1}/(2c_1)] \leq \widetilde{C}\exp[-M_7^{\alpha_1}/(2c_1)].
\end{aligned}$$

An application of Markov's inequality yields

$$P(|T_{j5}| \geq \delta) \leq \delta^{-1}E|T_{j5}| \leq \delta^{-1}\widetilde{C}\exp[-M_7^{\alpha_1}/(2c_1)] \tag{D.57}$$

for any $\delta > 0$. Recall that $M_7 = n^{\eta_7}$. In view of (D.57), taking $\delta = Cn^{-\kappa_2}/8$ gives that

$$P(\max_{1\leq j\leq p}|T_{j5}| \geq Cn^{-\kappa_2}/8) \leq \sum_{j=1}^{p} P(|T_{j5}| \geq Cn^{-\kappa_2}/8) \leq \ 8pC^{-1}\widetilde{C}n^{\kappa_2}\exp[-n^{2\eta_7}/(2c_1)]. \tag{D.58}$$

Combining (D.54), (D.55), (D.56), and (D.58) yields that for sufficiently large n,

$$\begin{aligned}
P(\max_{1\leq j\leq p}|S_{j2} - E(S_{j2})| \geq Cn^{-\kappa_2}/2) &\leq 2p\exp\left(-\widetilde{C}n^{1-2\kappa_2-2\eta_7-2\eta_8}\right) \\
&+ 8pC^{-1}\widetilde{C}n^{\kappa_2}\exp[-n^{\alpha_1\eta_7}/(2c_1)] + 16C^{-1}\widetilde{C}n^{\eta_7+\kappa_2}\exp[-n^{\alpha_2\eta_8}/(2c_1)].
\end{aligned} \tag{D.59}$$

Let $\eta_7 = \eta_8 = (1-2\kappa_2)/(4+\alpha_1)$. Then (D.59) becomes

$$\begin{aligned}
&P(\max_{1\leq j\leq p}|S_{j2} - E(S_{j2})| \geq Cn^{-\kappa_1}/2) \\
\leq &p\widetilde{C}_3\exp[-\widetilde{C}_4 n^{\alpha_1\eta_7}] + \widetilde{C}_5\exp[-\widetilde{C}_6 n^{\alpha_2\eta_7}]
\end{aligned} \tag{D.60}$$

for all n sufficiently large, where $\widetilde{C}_3$, $\widetilde{C}_4$, $\widetilde{C}_5$, and $\widetilde{C}_6$ are some positive constants.

Since $0 < \eta_6 < \eta_7$, it follows from (D.41), (D.49), and (D.60) that

$$\begin{aligned}P(\max_{1\le j\le p}|\widehat{\omega}_j^* - \omega_j^*| \ge Cn^{-\kappa_2}) &\le p\widetilde{C}_1 \exp\left(-\widetilde{C}_2 n^{\alpha_1\eta_6}\right) + p\widetilde{C}_3 \exp[-\widetilde{C}_4 n^{\alpha_1\eta_7}] + \widetilde{C}_5 \exp[-\widetilde{C}_6 n^{\alpha_2\eta_7}] \\ &\le p\widetilde{C}_7 \exp\left(-\widetilde{C}_8 n^{\alpha_1\eta_6}\right) + \widetilde{C}_5 \exp[-\widetilde{C}_6 n^{\alpha_2\eta_6}]\end{aligned}$$

with $\widetilde{C}_7 = \widetilde{C}_1 + \widetilde{C}_3$ and $\widetilde{C}_8 = \min\{\widetilde{C}_2, \widetilde{C}_4\}$ for all n sufficiently large. If $\log p = o(n^{\alpha_1\eta'})$ with $\eta' = \min\{(1-2\kappa_2-2\xi_2)/(4+\alpha_1), (1-2\kappa_2-2\xi_1)/(6+\alpha_1)\} > 0$, then for any positive constant C, there exists some arbitrarily large positive constant C_2 such that

$$P(\max_{1\le j\le p}|\widehat{\omega}_j^* - \omega_j^*| \ge Cn^{-\kappa_2}) \le o(n^{-C_2})$$

for all n sufficiently large, which completes the proof of part b) of Lemma 1.

D.4 Proof of part c) of Lemma 1

The main idea of the proof is to find probability bounds for the two events $\{\mathcal{I} \subset \widehat{\mathcal{I}}\}$ and $\{\mathcal{M} \subset \widehat{\mathcal{M}}\}$, respectively. First note that conditional on the event $\{\mathcal{A} \subset \widehat{\mathcal{A}}\}$, we have $\{\mathcal{I} \subset \widehat{\mathcal{I}}\}$. Thus it holds that

$$P(\mathcal{I} \subset \widehat{\mathcal{I}}) \ge P(\mathcal{A} \subset \widehat{\mathcal{A}}). \tag{D.61}$$

Define the event $\mathcal{E}_1 = \{\max_{k\in\mathcal{A}}|\widehat{\omega}_k - \omega_k| < 2^{-1}c_2 n^{-\kappa_1}\}$. Then, with $\tau = c_2 n^{-\kappa_1}$, the event $\mathcal{E}_1$ ensures that $\mathcal{A} \subset \widehat{\mathcal{A}}$. Thus,

$$P(\mathcal{A} \subset \widehat{\mathcal{A}}) \ge P(\mathcal{E}_1) = 1 - P(\mathcal{E}_1^c) = 1 - P(\max_{k\in\mathcal{A}}|\widehat{\omega}_k - \omega_k| \ge 2^{-1}c_2 n^{-\kappa_1}).$$

Following similar arguments as for proving (D.10), it can be shown that there exist some constants $\widetilde{C}_1 > 0$ and $\widetilde{C}_2 > 0$ such that for all n sufficiently large,

$$P(\max_{k\in\mathcal{A}}|\widehat{\omega}_k - \omega_k| \ge 2^{-1}c_2 n^{-\kappa_1}) \le 2s_1\widetilde{C}_1 \exp[-\widetilde{C}_2 n^{\min\{\alpha_1,\alpha_2\}r_1}]. \tag{D.62}$$

Note that the right hand side of (D.62) can be bounded by $o(n^{-C_1})$ for some arbitrarily large positive constant C_1. This gives

$$P(\mathcal{A} \subset \widehat{\mathcal{A}}) \geq 1 - o(n^{-C_1}). \tag{D.63}$$

Thus combining (D.61) and (D.63) yields

$$P(\mathcal{I} \subset \widehat{\mathcal{I}}) \geq 1 - o(n^{-C_1}). \tag{D.64}$$

Using similar arguments as for proving part b) of Lemma 1 and (D.63), we can show that there exist some positive constants $\widetilde{C}_1$, $\widetilde{C}_2$, and C_2 such that for all n sufficiently large,

$$\begin{aligned} P(\mathcal{B} \subset \widehat{\mathcal{B}}) &\geq P(\max_{j \in \mathcal{B}} |\widehat{\omega}_j^* - \omega_j^*| < 2^{-1} c_2 n^{-\kappa_2}) \geq 1 - s_2 \widetilde{C}_1 \exp(-\widetilde{C}_2 n^{\alpha_1 r_2}) \\ &\geq 1 - o(n^{-C_2}), \end{aligned} \tag{D.65}$$

Combining (D.63) and (D.65) leads to

$$\begin{aligned} P(\mathcal{M} \subset \widehat{\mathcal{M}}) \geq & P(\mathcal{A} \subset \widehat{\mathcal{A}} \ \text{ and } \ \mathcal{B} \subset \widehat{\mathcal{B}}) \geq P(\mathcal{A} \subset \widehat{\mathcal{A}}) + P(\mathcal{B} \subset \widehat{\mathcal{B}}) - 1 \\ \geq & 1 - o(n^{-\min\{C_1, C_2\}}). \end{aligned} \tag{D.66}$$

In view of (D.64) and (D.66), we obtain

$$P(\mathcal{I} \subset \widehat{\mathcal{I}} \ \text{ and } \ \mathcal{M} \subset \widehat{\mathcal{M}}) \geq P(\mathcal{I} \subset \widehat{\mathcal{I}}) + P(\mathcal{M} \subset \widehat{\mathcal{M}}) - 1 \geq 1 - o(n^{-\min\{C_1, C_2\}})$$

for all n sufficiently large. This completes the proof for the first part of part c) of Lemma 1.

We proceed to prove the second part of part c) of Lemma 1. The main idea is to establish the probability bounds for two events $\{|\widehat{\mathcal{A}}| = O[n^{2\kappa_1} \lambda_{\max}(\Sigma^*)]\}$ and $\{|\widehat{\mathcal{B}}| =$

$O[n^{2\kappa_2}\lambda_{\max}(\Sigma)]\}$, respectively. If we can show that

$$P\left\{|\widehat{\mathcal{A}}| = O[n^{2\kappa_1}\lambda_{\max}(\Sigma^*)]\right\} \geq 1 - o(n^{-C_1}), \tag{D.67}$$

$$P\left\{|\widehat{\mathcal{B}}| = O[n^{2\kappa_2}\lambda_{\max}(\Sigma)]\right\} \geq 1 - o(n^{-C_2}) \tag{D.68}$$

with C_1 and C_2 defined in (D.1) and (D.2), respectively, then it holds that

$$P\left\{|\widehat{\mathcal{I}}| = O\left[n^{4\kappa_1}\lambda^2_{\max}(\Sigma^*)\right]\right\} \geq P\left\{|\widehat{\mathcal{A}}| = O\left[n^{2\kappa_1}\lambda_{\max}(\Sigma^*)\right]\right\} \geq 1 - o(n^{-C_1})$$

and

$$\begin{aligned} &P\left\{|\widehat{\mathcal{M}}| = O\left[n^{2\kappa_1}\lambda_{\max}(\Sigma^*) + n^{2\kappa_2}\lambda_{\max}(\Sigma)\right]\right\} \\ \geq &P\left\{|\widehat{\mathcal{A}}| = O[n^{2\kappa_1}\lambda_{\max}(\Sigma^*)] \text{ and } |\widehat{\mathcal{B}}| = O[n^{2\kappa_2}\lambda_{\max}(\Sigma)]\right\} \geq 1 - o(n^{-\min\{C_1,C_2\}}). \end{aligned}$$

Combining these two results yields

$$\begin{aligned} &P\left(|\widehat{\mathcal{I}}| = O\{n^{4\kappa_1}\lambda^2_{\max}(\Sigma^*)\} \text{ and } |\widehat{\mathcal{M}}| = O\{n^{2\kappa_1}\lambda_{\max}(\Sigma^*) + n^{2\kappa_2}\lambda_{\max}(\Sigma)\}\right) \\ =&1 - o\left(n^{-\min\{C_1,C_2\}}\right). \end{aligned}$$

It thus remains to prove (D.67) and (D.68). We begin with showing (D.68). The key step is to show that

$$\sum_{j=1}^{p}(\omega_j^*)^2 = \|E(xY)\|_2^2 \leq \widetilde{C}_3\lambda_{\max}(\Sigma) \tag{D.69}$$

for some constant $\widetilde{C}_3 > 0$, where $x = (X_1, \ldots, X_p)^T$. If so, conditional on the event $\mathcal{E}_2 = \{\max_{1\leq j\leq p}|\widehat{\omega}_j^* - \omega_j^*| \leq 2^{-1}c_2 n^{-\kappa_2}\}$, the number of variables in $\widehat{\mathcal{B}} = \{j : |\widehat{\omega}_j^*| > c_2 n^{-\kappa_2}\}$ cannot exceed the number of variables in $\{j : |\omega_j^*| > 2^{-1}c_2 n^{-\kappa_2}\}$, which is bounded by $4\widetilde{C}_3 c_2^{-2} n^{2\kappa_2}\lambda_{\max}(\Sigma)$. Thus it follows from (D.2) that for all n sufficiently large,

$$P\left\{|\widehat{\mathcal{B}}| \leq 4\widetilde{C}_3 c_2^{-2} n^{2\kappa_2}\lambda_{\max}(\Sigma)\right\} \geq P(\mathcal{E}_2) = 1 - P(\mathcal{E}_2^c) \geq 1 - o(n^{-C_2}). \tag{D.70}$$

Now we further prove (D.69). Let $u_0 = \mathrm{argmin}_{u \in \mathbb{R}^p} E\left(Y - x^T u\right)^2$. Then the first order equation $E[x(Y - x^T u_0)] = 0$ gives $E(xY) = [E(xx^T)]u_0 = \Sigma u_0$. Thus

$$\|E(xY)\|_2^2 = u_0^T \Sigma^2 u_0 \leq \lambda_{\max}(\Sigma) {u_0}^T \Sigma u_0 = \lambda_{\max}(\Sigma) \mathrm{var}\left(x^T u_0\right). \tag{D.71}$$

It follows from the orthogonal decomposition that

$$\mathrm{var}\left(Y\right) = \mathrm{var}\left(x^T u_0\right) + \mathrm{var}\left(Y - x^T u_0\right) \geq \mathrm{var}\left(x^T u_0\right).$$

Since $E^2(Y^2) \leq E(Y^4) = O(1)$, we have $\mathrm{var}(Y) \leq E(Y^2) = O(1)$. Then the above inequality ensures that $\mathrm{var}\left(x^T u_0\right) \leq \widetilde{C}_3$ for some constant $\widetilde{C}_3 > 0$. This together with (D.71) completes the proof of (D.69).

We next prove (D.67). Recall that $Y^* = Y^2$ and $X_k^* = \{X_k^2 - E(X_k^2)\}/\sqrt{\mathrm{var}(X_k^2)}$. Then from the definition of ω_k in (B.1), we have $\omega_k = E(X_k^* Y^*)$. Following similar arguments as for proving (D.69), it can be shown that

$$\sum_{k=1}^{p} \omega_k^2 = \sum_{k=1}^{p} E^2(X_k^* Y^*) = \|E(x^* Y^*)\|_2^2 \leq \widetilde{C}_4 \lambda_{\max}(\Sigma^*), \tag{D.72}$$

where $\widetilde{C}_4$ is some positive constant, $x^* = (X_1^*, \ldots, X_p^*)^T$, and $\Sigma^* = \mathrm{cov}(x^*)$. Then, on the event $\mathcal{E}_3 = \{\max_{1 \leq k \leq p} |\widehat{\omega}_k - \omega_k| \leq 2^{-1} c_2 n^{-\kappa_1}\}$, the cardinality of $\{k : |\widehat{\omega}_k| > c_2 n^{-\kappa_1}\}$ cannot exceed that of $\{k : |\omega_k| > 2^{-1} c_2 n^{-\kappa_1}\}$, which is bounded by $4\widetilde{C}_4 c_2^{-2} n^{2\kappa_1} \lambda_{\max}(\Sigma^*)$. Thus, we have

$$P\left\{|\widehat{\mathcal{A}}| \leq 4\widetilde{C}_4 c_2^{-2} n^{2\kappa_1} \lambda_{\max}(\Sigma^*)\right\} \geq P(\mathcal{E}_3) = 1 - P(\mathcal{E}_3^c) \geq 1 - o(n^{-C_1}),$$

where the last equality follows from (D.1). This concludes the proof of part c) of Lemma 1 and thus Lemma 1 is proved.

E Proofs of Technical Results

E.1 Lemma 2 and its proof

Lemma 2. *Let W_1 and W_2 be two random variables such that $P(|W_1| > t) \leq \widetilde{C}_1 \exp(-\widetilde{C}_2 t^{\alpha_1})$ and $P(|W_2| > t) \leq \widetilde{C}_3 \exp(-\widetilde{C}_4 t^{\alpha_2})$ for all $t > 0$, where α_1, α_2, and $\widetilde{C}_i$'s are some positive constants. Then $P(|W_1 W_2| > t) \leq \widetilde{C}_5 \exp(-\widetilde{C}_6 t^{\alpha_1\alpha_2/(\alpha_1+\alpha_2)})$ for all $t > 0$, with $\widetilde{C}_5 = \widetilde{C}_1 + \widetilde{C}_3$ and $\widetilde{C}_6 = \min\{\widetilde{C}_2, \widetilde{C}_4\}$.*

Proof of Lemma 2. For any $t > 0$, we have

$$
\begin{aligned}
& P(|W_1 W_2| > t) \leq P(|W_1| > t^{\alpha_2/(\alpha_1+\alpha_2)}) + P(|W_2| > t^{\alpha_1/(\alpha_1+\alpha_2)}) \\
& \leq \widetilde{C}_1 \exp(-\widetilde{C}_2 t^{\alpha_1\alpha_2/(\alpha_1+\alpha_2)}) + \widetilde{C}_3 \exp(-\widetilde{C}_4 t^{\alpha_1\alpha_2/(\alpha_1+\alpha_2)}) \\
& \leq \widetilde{C}_5 \exp(-\widetilde{C}_6 t^{\alpha_1\alpha_2/(\alpha_1+\alpha_2)})
\end{aligned}
$$

by setting $\widetilde{C}_5 = \widetilde{C}_1 + \widetilde{C}_3$ and $\widetilde{C}_6 = \min\{\widetilde{C}_2, \widetilde{C}_4\}$.

E.2 Lemma 3 and its proof

Lemma 3. *Let W be a nonnegative random variable such that $P(W > t) \leq \widetilde{C}_1 \exp(-\widetilde{C}_2 t^{\alpha})$ for all $t > 0$, where α and $\widetilde{C}_i$'s are some positive constants. Then it holds that $E(e^{\widetilde{C}_3 W^{\alpha}}) \leq \widetilde{C}_4$, $E(W^{\alpha m}) \leq \widetilde{C}_3^{-m} \widetilde{C}_4 m!$ for any integer $m \geq 0$ with $\widetilde{C}_3 = \widetilde{C}_2/2$ and $\widetilde{C}_4 = 1 + \widetilde{C}_1$, and $E(W^k) \leq \widetilde{C}_5$ for any integer $k \geq 1$, where constant $\widetilde{C}_5$ depends on k and α.*

Proof of Lemma 3. Let $F(t)$ be the cumulative distribution function of W. Then for all $t > 0$, $1 - F(W) = P(W > t) \leq \widetilde{C}_1 \exp(-\widetilde{C}_2 t^{\alpha})$. Recall that W is a nonnegative random variable. Thus, for any $0 < T < \widetilde{C}_2$, by integration by parts we have

$$
\begin{aligned}
E(e^{TW^{\alpha}}) &= -\int_0^{\infty} e^{Tt^{\alpha}} d[1 - F(t)] = 1 + \int_0^{\infty} T\alpha t^{\alpha-1} e^{Tt^{\alpha}} [1 - F(t)]\, dt \\
&\leq 1 + \int_0^{\infty} T\alpha t^{\alpha-1} \cdot \widetilde{C}_1 e^{-(\widetilde{C}_2 - T)t^{\alpha}}\, dt = 1 + \frac{T\widetilde{C}_1}{\widetilde{C}_2 - T}.
\end{aligned}
$$

Then, taking $\widetilde{C}_3 = T = \widetilde{C}_2/2$ and $\widetilde{C}_4 = 1 + \widetilde{C}_1$ proves the first desired result.

Note that $\widetilde{C}_3^m E(W^{\alpha m})/m! \le \sum_{k=0}^{\infty} \widetilde{C}_3^k E(W^{\alpha k})/k! = E(e^{\widetilde{C}_3 W^{\alpha}})$ for any nonnegative integer m. Thus $E(W^{\alpha m}) \le \widetilde{C}_3^{-m}\widetilde{C}_4 m!$, which proves the second desired result.

For any integer $k \ge 1$, there exists an integer $m \ge 1$ such that $k < \alpha m$. Then applying Hölder's inequality gives

$$\begin{aligned} E(W^k) \le & \left\{E[(W^k)^{\alpha m/k}]\right\}^{k/(\alpha m)} \left\{E[1^{\alpha m/(\alpha m-k)}]\right\}^{(\alpha m-k)/(\alpha m)} \\ = & \left\{E(W^{\alpha m})\right\}^{k/(\alpha m)} \le \left(\widetilde{C}_3^{-m}\widetilde{C}_4 m!\right)^{k/(\alpha m)}. \end{aligned}$$

Thus the kth moment of W is bounded by a constant $\widetilde{C}_5$, which depends on k and α. This proves the third desired result.

E.3 Lemma 4 and its proof

Lemma 4. *Let W be a nonnegative random variable with tail probability $P(W > t) \le \widetilde{C}_1 \exp(-\widetilde{C}_2 t^{\alpha})$ for all $t > 0$, where α and $\widetilde{C}_i$'s are some positive constants. If constant $\alpha \ge 1$, then $E(e^{\widetilde{C}_3 W}) \le \widetilde{C}_4$ and $E(W^m) \le \widetilde{C}_3^{-m}\widetilde{C}_4 m!$ for any integer $m \ge 0$ with $\widetilde{C}_3 = \widetilde{C}_2/2$ and $\widetilde{C}_4 = e^{\widetilde{C}_2/2} + \widetilde{C}_1 e^{-\widetilde{C}_2/2}$.*

Proof of Lemma 4. Let $F(t)$ be the cumulative distribution function of nonnegative random variable W. Then $1 - F(t) = P(W > t) \le \widetilde{C}_1 \exp(-\widetilde{C}_2 t^{\alpha})$ for all $t \ge 1$. If $\alpha \ge 1$, then $t \le t^{\alpha}$ for all $t \ge 1$ and thus $1 - F(t) \le \widetilde{C}_1 \exp(-\widetilde{C}_2 t)$ for all $t \ge 1$. Define $\widetilde{C}_3 = \widetilde{C}_2/2$ and $\widetilde{C}_4 = e^{\widetilde{C}_2/2} + \widetilde{C}_1 e^{-\widetilde{C}_2/2}$. By integration by parts, we deduce

$$\begin{aligned} E(e^{\widetilde{C}_3 W}) = & -\int_0^{\infty} e^{\widetilde{C}_3 t} d[1-F(t)] = 1 + \int_0^{\infty} \widetilde{C}_3 e^{\widetilde{C}_3 t}[1-F(t)]dt \\ = & 1 + \int_0^1 \widetilde{C}_3 e^{\widetilde{C}_3 t}[1-F(t)]dt + \int_1^{\infty} \widetilde{C}_3 e^{\widetilde{C}_3 t}[1-F(t)]dt \\ \le & 1 + \int_0^1 \widetilde{C}_3 e^{\widetilde{C}_3 t} dt + \int_1^{\infty} \widetilde{C}_1 \widetilde{C}_3 e^{(\widetilde{C}_3 - \widetilde{C}_2)t} dt = e^{\widetilde{C}_2/2} + \widetilde{C}_1 e^{-\widetilde{C}_2/2} = \widetilde{C}_4, \end{aligned}$$

which proves the first desired result.

Note that $\widetilde{C}_3^m E(W^m)/m! \le \sum_{k=0}^{\infty} \widetilde{C}_3^k E(W^k)/k! = E(e^{\widetilde{C}_3 W})$ for any nonnegative integer m. Thus $E(W^m) \le \widetilde{C}_3^{-m}\widetilde{C}_4 m!$, which proves the second desired result.

E.4 Lemma 5 and its proof

Lemma 5. *For any real numbers* $b_1, b_2 \geq 0$ *and* $\alpha > 0$, *it holds that* $(b_1 + b_2)^\alpha \leq C_\alpha(b_1^\alpha + b_2^\alpha)$ *with* $C_\alpha = 1$ *if* $0 < \alpha \leq 1$ *and* $2^{\alpha-1}$ *if* $\alpha > 1$.

Proof of Lemma 5. We first consider the case of $0 < \alpha \leq 1$. It is trivial if $b_1 = 0$ or $b_2 = 0$. Assume that both b_1 and b_2 are positive. Since $0 < b_1/(b_1 + b_2) < 1$, we have $[b_1/(b_1 + b_2)]^\alpha \geq b_1/(b_1 + b_2)$. Similarly, it holds that $[b_2/(b_1 + b_2)]^\alpha \geq b_2/(b_1 + b_2)$. Combining these two results yields

$$\left(\frac{b_1}{b_1 + b_2}\right)^\alpha + \left(\frac{b_2}{b_1 + b_2}\right)^\alpha \geq \frac{b_1}{b_1 + b_2} + \frac{b_2}{b_1 + b_2} = 1,$$

which implies that $(b_1 + b_2)^\alpha \leq b_1^\alpha + b_2^\alpha$.

Next, we deal with the case of $\alpha > 1$. Since x^α is a convex function on $[0, \infty)$ for a given $\alpha > 1$, we have $[(b_1 + b_2)/2]^\alpha \leq (b_1^\alpha + b_2^\alpha)/2$, which ensures that $(b_1 + b_2)^\alpha \leq 2^{\alpha-1}(b_1^\alpha + b_2^\alpha)$. Combining the two cases above leads to the desired result.

E.5 Lemma 6 and its proof

Lemma 6. *Let* $W_1, \ldots, W_n$ *be independent random variables with tail probability* $P(|W_i| > t) \leq \widetilde{C}_1 \exp(-\widetilde{C}_2 t^\alpha)$ *for all* $t > 0$, *where* α *and* $\widetilde{C}_i$*'s are some positive constants. Then there exist some positive constants* $\widetilde{C}_3$ *and* $\widetilde{C}_4$ *such that*

$$P\{|n^{-1}\sum_{i=1}^{n}(W_i - EW_i)| > \epsilon\} \leq \widetilde{C}_3 \exp(-\widetilde{C}_4 n^{\min\{\alpha,1\}}\epsilon^2) \tag{E.1}$$

for $0 < \epsilon \leq 1$.

Proof of Lemma 6. Define $\widetilde{W}_i = W_i - EW_i$. Then by the triangle inequality and the property of expectation, we have

$$|\widetilde{W}_i| = |W_i - EW_i| \leq |W_i| + |EW_i| \leq |W_i| + E|W_i|. \tag{E.2}$$

Next, we consider two cases.

Case 1: $0 < \alpha \leq 1$. It follows from Lemma 3 that $E(e^{T|W_i|^\alpha}) \leq 1+\widetilde{C}_1$ and $E|W_i| \leq C_0$ for all $1 \leq i \leq n$, where $T = \widetilde{C}_2/2$ and C_0 is some positive constant. In view of (E.2) and by Lemma 5, we have $|\widetilde{W}_i|^\alpha \leq (|W_i| + E|W_i|)^\alpha \leq |W_i|^\alpha + (E|W_i|)^\alpha$. This ensures

$$E(e^{T|\widetilde{W}_i|^\alpha}) \leq e^{T(E|W_i|)^\alpha} E(e^{T|W_i|^\alpha}) \leq e^{TC_0^\alpha}(1+\widetilde{C}_1).$$

Thus, by the Chernoff bound arguments we can show that there exist some positive constants $\widetilde{C}_5$ and $\widetilde{C}_6$ such that

$$P(|n^{-1}\sum_{i=1}^n [W_i - EW_i]| > \epsilon) = P(|n^{-1}\sum_{i=1}^n \widetilde{W}_i| > \epsilon) \leq \widetilde{C}_5 \exp\left(-\widetilde{C}_6 n^\alpha \epsilon^2\right) \tag{E.3}$$

for any $0 < \epsilon \leq 1$.

Case 2: $\alpha > 1$. In view of (E.2), it follows from Lemma 5 and Jensen's inequality that for each integer $m \geq 2$,

$$\begin{aligned} &E(|\widetilde{W}_i|^m) \leq E[(|W_i| + E|W_i|)^m] \leq 2^{m-1} E\left[|W_i|^m + (E|W_i|)^m\right] \\ =&2^{m-1}[E(|W_i|^m) + (E|W_i|)^m] \leq 2^{m-1}[E(|W_i|^m) + E(|W_i|^m)] = 2^m E(|W_i|^m). \end{aligned} \tag{E.4}$$

Recall that $P(|W_i| > t) \leq \widetilde{C}_1 \exp(-\widetilde{C}_2 t^\alpha)$ for all $t > 0$ and $\alpha > 1$. By Lemma 4, there exist some positive constants $\widetilde{C}_7$ and $\widetilde{C}_8$ such that $E(|W_i|^m) \leq m!\widetilde{C}_7^m \widetilde{C}_8$. This together with (E.4) gives

$$E(|\widetilde{W}_i|^m) \leq m!(2\widetilde{C}_7)^{m-2}(8\widetilde{C}_7^2\widetilde{C}_8)/2$$

for all $m \geq 2$. Thus an application of Bernstein's inequality yields

$$\begin{aligned} &P\{|n^{-1}\sum_{i=1}^n (W_i - EW_i)| > \epsilon\} = P(|n^{-1}\sum_{i=1}^n \widetilde{W}_i| > \epsilon) \\ &\leq 2\exp\left(-\frac{n\epsilon^2}{16\widetilde{C}_7^2\widetilde{C}_8 + 4\widetilde{C}_7\epsilon}\right) \leq 2\exp\left(-\frac{n\epsilon^2}{16\widetilde{C}_7^2\widetilde{C}_8 + 4\widetilde{C}_7}\right) \end{aligned} \tag{E.5}$$

for any $0 < \epsilon < 1$. Let $\widetilde{C}_3 = \max\{\widetilde{C}_5, 2\}$ and $\widetilde{C}_4 = \min\{\widetilde{C}_6, (16\widetilde{C}_7^2\widetilde{C}_8 + 4\widetilde{C}_7)^{-1}\}$. Combining (E.3) and (E.5) completes the proof of Lemma 6.

E.6 Lemma 7 and its proof

Lemma 7. *Assume that for each $1 \leq j \leq p$, $X_{1j}, \ldots, X_{nj}$ are n i.i.d. random variables satisfying $P(|X_{1j}| > t) \leq \widetilde{C}_1 \exp(-\widetilde{C}_2 t^{\alpha_1})$ for any $t > 0$, where $\widetilde{C}_1, \widetilde{C}_2$ and α_1 are some positive constants. Then for any $0 < \epsilon < 1$, we have*

$$P\left\{\left|n^{-1}\sum_{i=1}^{n}[X_{ij}X_{ik} - E(X_{ij}X_{ik})]\right| > \epsilon\right\} \leq \widetilde{C}_3 \exp(-\widetilde{C}_4 n^{\min\{\alpha_1/2,1\}}\epsilon^2), \tag{E.6}$$

$$P\left\{\left|n^{-1}\sum_{i=1}^{n}[X_{ij}X_{ik}X_{i\ell} - E(X_{ij}X_{ik}X_{i\ell})]\right| > \epsilon\right\} \leq \widetilde{C}_5 \exp(-\widetilde{C}_6 n^{\min\{\alpha_1/3,1\}}\epsilon^2), \tag{E.7}$$

$$\begin{aligned} &P\left\{\left|n^{-1}\sum_{i=1}^{n}[X_{ik}X_{i\ell}X_{ik'}X_{i\ell'} - E(X_{ik}X_{i\ell}X_{ik'}X_{i\ell'})]\right| > \epsilon\right\} \\ &\quad \leq \widetilde{C}_7 \exp(-\widetilde{C}_8 n^{\min\{\alpha_1/4,1\}}\epsilon^2), \end{aligned} \tag{E.8}$$

where $1 \leq j, k, \ell, k', \ell' \leq p$ and $\widetilde{C}_i$'s are some positive constants.

Proof of Lemma 7. The proofs for inequalities (E.6)–(E.8) are similar. To save space, we only show the inequality (E.8) here. Since $P(|X_{ij}| > t) \leq \widetilde{C}_1 \exp(-\widetilde{C}_2 t^{\alpha_1})$ for all $t > 0$ and all i and j, it follows from Lemma 2 that $X_{ik}X_{i\ell}X_{ik'}X_{i\ell'}$ admits tail probability $P(|X_{ik}X_{i\ell}X_{ik'}X_{i\ell'}| > t) \leq 4\widetilde{C}_1 \exp(-\widetilde{C}_2 t^{\alpha_1/4})$. By Lemma 6, there exist some positive constants $\widetilde{C}_3$ and $\widetilde{C}_4$ such that

$$\begin{aligned} &P(|n^{-1}\sum_{i=1}^{n}[X_{ik}X_{i\ell}X_{ik'}X_{i\ell'} - E(X_{ik}X_{i\ell}X_{ik'}X_{i\ell'})]| > \epsilon) \\ &\quad \leq \widetilde{C}_3 \exp\left(-\widetilde{C}_4 n^{\min\{\alpha_1/4,1\}}\epsilon^2\right) \end{aligned}$$

for any $0 < \epsilon < 1$, which concludes the proof of (E.8).

E.7 Lemma 8 and its proof

Lemma 8. *Let A_j's with $j \in \mathcal{D} \subset \{1, \ldots, p\}$ satisfy $\max_{j\in\mathcal{D}} |A_j| \le L_3$ for some constant $L_3 > 0$, and $\widehat{A}_j$ be an estimate of A_j based on a sample of size n for each $j \in \mathcal{D}$. Assume that for any constant $C > 0$, there exist constants $\widetilde{C}_1, \widetilde{C}_2 > 0$ such that*

$$P\left(\max_{j\in\mathcal{D}} |\widehat{A}_j - A_j| \ge Cn^{-\kappa_1}\right) \le |\mathcal{D}|\widetilde{C}_1 \exp\left\{-\widetilde{C}_2 n^{f(\kappa_1)}\right\}$$

with $f(\kappa_1)$ some function of κ_1. Then for any constant $C > 0$, there exist constants $\widetilde{C}_3, \widetilde{C}_4 > 0$ such that

$$P\left(\max_{j\in\mathcal{D}} |\widehat{A}_j^2 - A_j^2| \ge Cn^{-\kappa_1}\right) \le |\mathcal{D}|\widetilde{C}_3 \exp\left\{-\widetilde{C}_4 n^{f(\kappa_1)}\right\}.$$

Proof of Lemma 8. Note that $\max_{j\in\mathcal{D}} |\widehat{A}_j^2 - A_j^2| \le \max_{j\in\mathcal{D}} |\widehat{A}_j(\widehat{A}_j - A_j)| + \max_{j\in\mathcal{D}} |(\widehat{A}_j - A_j)A_j|$. Therefore, for any positive constant C,

$$\begin{aligned}P(\max_{j\in\mathcal{D}} |\widehat{A}_j^2 - A_j^2| \ge Cn^{-\kappa_1}) \le\ & P(\max_{j\in\mathcal{D}} |\widehat{A}_j(\widehat{A}_j - A_j)| \ge Cn^{-\kappa_1}/2) \\ & + P(\max_{j\in\mathcal{D}} |(\widehat{A}_j - A_j)A_j| \ge Cn^{-\kappa_1}/2). \qquad \text{(E.9)}\end{aligned}$$

We first deal with the second term on the right hand side of (E.9). Since $\max_{j\in\mathcal{D}} |A_j| \le L_3$, we have

$$\begin{aligned}&P(\max_{j\in\mathcal{D}} |(\widehat{A}_j - A_j)A_j| \ge Cn^{-\kappa_1}/2) \le P(\max_{j\in\mathcal{D}} |\widehat{A}_j - A_j|L_3 \ge Cn^{-\kappa_1}/2) \\ =&P\{\max_{j\in\mathcal{D}} |\widehat{A}_j - A_j| \ge (2L_3)^{-1}Cn^{-\kappa_1}\} \le |\mathcal{D}|\widetilde{C}_1 \exp\left\{-\widetilde{C}_2 n^{f(\kappa_1)}\right\}, \qquad \text{(E.10)}\end{aligned}$$

where $\widetilde{C}_1$ and $\widetilde{C}_2$ are two positive constants.

Next, we consider the first term on the right hand side of (E.9). Note that

$$
\begin{aligned}
& P(\max_{j\in\mathcal{D}} |\widehat{A}_j(\widehat{A}_j - A_j)| \geq Cn^{-\kappa_1}/2) \\
\leq & P(\max_{j\in\mathcal{D}} |\widehat{A}_j(\widehat{A}_j - A_j)| \geq Cn^{-\kappa_1}/2, \max_{j\in\mathcal{D}} |\widehat{A}_j| \geq L_3 + Cn^{-\kappa_1}/2) \\
& \quad + P(\max_{j\in\mathcal{D}} |\widehat{A}_j(\widehat{A}_j - A_j)| \geq Cn^{-\kappa_1}/2, \max_{j\in\mathcal{D}} |\widehat{A}_j| < L_3 + Cn^{-\kappa_1}/2) \\
\leq & P(\max_{j\in\mathcal{D}} |\widehat{A}_j| \geq L_3 + Cn^{-\kappa_1}/2) \qquad \text{(E.11)} \\
& \quad + P(\max_{j\in\mathcal{D}} |\widehat{A}_j(\widehat{A}_j - A_j)| \geq Cn^{-\kappa_1}/2, \max_{j\in\mathcal{D}} |\widehat{A}_j| < L_3 + C) \\
\leq & P(\max_{j\in\mathcal{D}} |\widehat{A}_j| \geq L_3 + Cn^{-\kappa_1}/2) + P(\max_{j\in\mathcal{D}} |(L_3 + C)(\widehat{A}_j - A_j)| \geq Cn^{-\kappa_1}/2). \qquad \text{(E.12)}
\end{aligned}
$$

Let us bound the two terms on the right hand side of (E.11) one by one. Since $\max_{j\in\mathcal{D}} |A_j| \leq L_3$, we have

$$
\begin{aligned}
& P(\max_{j\in\mathcal{D}} |\widehat{A}_j| \geq L_3 + Cn^{-\kappa_1}/2) \leq P(\max_{j\in\mathcal{D}} |\widehat{A}_j - A_j| + \max_{j\in\mathcal{D}} |A_j| \geq L_3 + Cn^{-\kappa_1}/2) \\
\leq & P(\max_{j\in\mathcal{D}} |\widehat{A}_j - A_j| \geq 2^{-1}Cn^{-\kappa_1}) \leq |\mathcal{D}|\widetilde{C}_5 \exp\left\{-\widetilde{C}_6 n^{f(\kappa_1)}\right\}, \qquad \text{(E.13)}
\end{aligned}
$$

where $\widetilde{C}_5$ and $\widetilde{C}_6$ are two positive constants. It also holds that

$$
\begin{aligned}
& P(\max_{j\in\mathcal{D}} |(L_3 + C)(\widehat{A}_j - A_j)| \geq Cn^{-\kappa_1}/2) \\
& = P\{\max_{j\in\mathcal{D}} |\widehat{A}_j - A_j| \geq (2L_3 + 2C)^{-1}Cn^{-\kappa_1}\} \\
& \leq |\mathcal{D}|\widetilde{C}_7 \exp\left\{-\widetilde{C}_8 n^{f(\kappa_1)}\right\},
\end{aligned}
$$

where $\widetilde{C}_7$ and $\widetilde{C}_8$ are two positive constants. This, together with (E.9)–(E.13), entails

$$
\begin{aligned}
& P(\max_{j\in\mathcal{D}} |\widehat{A}_j^2 - A_j^2| \geq Cn^{-\kappa_1}) \leq |\mathcal{D}|\widetilde{C}_1 \exp\left\{-\widetilde{C}_2 n^{f(\kappa_1)}\right\} + |\mathcal{D}|\widetilde{C}_5 \exp\left\{-\widetilde{C}_6 n^{f(\kappa_1)}\right\} \\
& + |\mathcal{D}|\widetilde{C}_7 \exp\left\{-\widetilde{C}_8 n^{f(\kappa_1)}\right\} \leq |\mathcal{D}|\widetilde{C}_3 \exp\left\{-\widetilde{C}_4 n^{f(\kappa_1)}\right\},
\end{aligned}
$$

where $\widetilde{C}_3 = \widetilde{C}_1 + \widetilde{C}_5 + \widetilde{C}_7 > 0$ and $\widetilde{C}_4 = \min\{\widetilde{C}_2, \widetilde{C}_6, \widetilde{C}_8\} > 0$.

E.8 Lemma 9 and its proof

Lemma 9. *Let $\widehat{A}_j$ and $\widehat{B}_j$ be estimates of A_j and B_j, respectively, based on a sample of size n for each $j \in \mathcal{D} \subset \{1, \ldots, p\}$. Assume that for any constant $C > 0$, there exist constants $\widetilde{C}_1, \ldots, \widetilde{C}_8 > 0$ except $\widetilde{C}_3, \widetilde{C}_7 \geq 0$ such that*

$$P\left(\max_{j\in\mathcal{D}} |\widehat{A}_j - A_j| \geq Cn^{-\kappa_1}\right) \leq |\mathcal{D}|\widetilde{C}_1 \exp\left\{-\widetilde{C}_2 n^{f(\kappa_1)}\right\} + \widetilde{C}_3 \exp\left\{-\widetilde{C}_4 n^{g(\kappa_1)}\right\},$$
$$P\left(\max_{j\in\mathcal{D}} |\widehat{B}_j - B_j| \geq Cn^{-\kappa_1}\right) \leq |\mathcal{D}|\widetilde{C}_5 \exp\left\{-\widetilde{C}_6 n^{f(\kappa_1)}\right\} + \widetilde{C}_7 \exp\left\{-\widetilde{C}_8 n^{g(\kappa_1)}\right\}$$

with $f(\kappa_1)$ and $g(\kappa_1)$ some functions of κ_1. Then for any constant $C > 0$, there exist constants $\widetilde{C}_9, \ldots, \widetilde{C}_{12} > 0$ except $\widetilde{C}_{11} \geq 0$ such that

$$\begin{aligned} &P\left\{\max_{j\in\mathcal{D}} |(\widehat{A}_j - \widehat{B}_j) - (A_j - B_j)| \geq Cn^{-\kappa_1}\right\} \leq |\mathcal{D}|\widetilde{C}_9 \exp\left\{-\widetilde{C}_{10} n^{f(\kappa_1)}\right\} \\ &\quad + \widetilde{C}_{11} \exp\left\{-\widetilde{C}_{12} n^{g(\kappa_1)}\right\}. \end{aligned}$$

Proof of Lemma 9. Note that $\max_{j\in\mathcal{D}} |(\widehat{A}_j - \widehat{B}_j) - (A_j - B_j)| \leq \max_{j\in\mathcal{D}} |\widehat{A}_j - A_j| + \max_{j\in\mathcal{D}} |\widehat{B}_j - B_j|$. Thus, for any positive constant C,

$$\begin{aligned} &P(\max_{j\in\mathcal{D}} |(\widehat{A}_j - \widehat{B}_j) - (A_j - B_j)| \geq Cn^{-\kappa_1}) \\ \leq &P(\max_{j\in\mathcal{D}} |\widehat{A}_j - A_j| \geq Cn^{-\kappa_1}/2) + P(\max_{j\in\mathcal{D}} |\widehat{B}_j - B_j| \geq Cn^{-\kappa_1}/2) \\ \leq &|\mathcal{D}|\widetilde{C}_1 \exp\left\{-\widetilde{C}_2 n^{f(\kappa_1)}\right\} + \widetilde{C}_3 \exp\left\{-\widetilde{C}_4 n^{g(\kappa_1)}\right\} + |\mathcal{D}|\widetilde{C}_5 \exp\left\{-\widetilde{C}_6 n^{f(\kappa_1)}\right\} \\ &+ \widetilde{C}_7 \exp\left\{-\widetilde{C}_8 n^{g(\kappa_1)}\right\} \\ \leq &|\mathcal{D}|\widetilde{C}_9 \exp\left\{-\widetilde{C}_{10} n^{f(\kappa_1)}\right\} + \widetilde{C}_{11} \exp\left\{-\widetilde{C}_{12} n^{g(\kappa_1)}\right\}, \end{aligned}$$

where $\widetilde{C}_9 = \widetilde{C}_1 + \widetilde{C}_5 > 0$, $\widetilde{C}_{10} = \min\{\widetilde{C}_2, \widetilde{C}_6\} > 0$, $\widetilde{C}_{11} = \widetilde{C}_3 + \widetilde{C}_7 \geq 0$, and $\widetilde{C}_{12} = \min\{\widetilde{C}_4, \widetilde{C}_8\} > 0$.

E.9 Lemma 10 and its proof

Lemma 10. *Let B_j's with $j \in \mathcal{D} \subset \{1, \ldots, p\}$ satisfy $\min_{j\in\mathcal{D}} B_j \geq L_4$ for some constant $L_4 > 0$, and $\widehat{B}_j$ be an estimate of B_j based on a sample of size n for each $j \in \mathcal{D}$. Assume that for any constant $C > 0$, there exist constants $\widetilde{C}_1, \widetilde{C}_2 > 0$ such that*

$$P\left(\max_{j\in\mathcal{D}} |\widehat{B}_j - B_j| \geq Cn^{-\kappa_1}\right) \leq |\mathcal{D}|\widetilde{C}_1 \exp\left\{-\widetilde{C}_2 n^{f(\kappa_1)}\right\}.$$

Then for any constant $C > 0$, there exist constants $\widetilde{C}_3, \widetilde{C}_4 > 0$ such that

$$P\left(\max_{j\in\mathcal{D}} \left|\sqrt{\widehat{B}_j} - \sqrt{B_j}\right| \geq Cn^{-\kappa_1}\right) \leq |\mathcal{D}|\widetilde{C}_3 \exp\left\{-\widetilde{C}_4 n^{f(\kappa_1)}\right\}.$$

Proof of Lemma 10. Since $\min_{j\in\mathcal{D}} B_j \geq L_4 > 0$, there exists some constant L_0 such that $0 < L_0 < L_4$. Note that, for any positive constant C,

$$\begin{aligned}
&P(\max_{j\in\mathcal{D}} |\sqrt{\widehat{B}_j} - \sqrt{B_j}| \geq Cn^{-\kappa_1}) \\
\leq &P(\max_{j\in\mathcal{D}} |\sqrt{\widehat{B}_j} - \sqrt{B_j}| \geq Cn^{-\kappa_1}, \min_{j\in\mathcal{D}} |\widehat{B}_j| \leq L_4 - L_0 n^{-\kappa_1}) \\
&\quad + P(\max_{j\in\mathcal{D}} |\sqrt{\widehat{B}_j} - \sqrt{B_j}| \geq Cn^{-\kappa_1}, \min_{j\in\mathcal{D}} |\widehat{B}_j| > L_4 - L_0 n^{-\kappa_1}) \\
\leq &P(\min_{j\in\mathcal{D}} |\widehat{B}_j| \leq L_4 - L_0 n^{-\kappa_1}) \\
&\quad + P(\max_{j\in\mathcal{D}} \frac{|\widehat{B}_j - B_j|}{|\sqrt{\widehat{B}_j} + \sqrt{B_j}|} \geq Cn^{-\kappa_1}, \min_{j\in\mathcal{D}} |\widehat{B}_j| > L_4 - L_0). \qquad \text{(E.14)}
\end{aligned}$$

Consider the first term on the right hand side of (E.14). For any positive constant C, we have

$$\begin{aligned}
&P(\min_{j\in\mathcal{D}} |\widehat{B}_j| \leq L_4 - L_0 n^{-\kappa_1}) \leq P(\min_{j\in\mathcal{D}} |B_j| - \max_{j\in\mathcal{D}} |\widehat{B}_j - B_j| \leq L_4 - L_0 n^{-\kappa_1}) \\
\leq &P(\max_{j\in\mathcal{D}} |\widehat{B}_j - B_j| \geq L_0 n^{-\kappa_1}) \leq |\mathcal{D}|\widetilde{C}_1 \exp\left\{-\widetilde{C}_2 n^{f(\kappa_1)}\right\}, \qquad \text{(E.15)}
\end{aligned}$$

by noticing that $\min_{j\in\mathcal{D}} B_j \geq L_4$, where $\widetilde{C}_1$ and $\widetilde{C}_2$ are some positive constants.

Next consider the second term on the right hand side of (E.14). Recall that $\min_{j\in\mathcal{D}} B_j \geq$

L_4. Then, for any positive constant C,

$$\begin{aligned}&P(\max_{j\in\mathcal{D}}\frac{|\widehat{B}_j-B_j|}{|\sqrt{\widehat{B}_j}+\sqrt{B_j}|}\geq Cn^{-\kappa_1},\min_{j\in\mathcal{D}}|\widehat{B}_j|>L_4-L_0)\\ \leq& P\{\max_{j\in\mathcal{D}}|\widehat{B}_j-B_j|\geq C(\sqrt{L_4-L_0}+\sqrt{L_4})n^{-\kappa_1}\}\leq|\mathcal{D}|\widetilde{C}_5\exp\left\{-\widetilde{C}_6n^{f(\kappa_1)}\right\},\end{aligned} \tag{E.16}$$

where $\widetilde{C}_5$ and $\widetilde{C}_6$ are some positive constants. Combining (E.14), (E.15), and (E.16) gives

$$P(\max_{j\in\mathcal{D}}|\sqrt{\widehat{B}_j}-\sqrt{B_j}|\geq Cn^{-\kappa_1})\leq|\mathcal{D}|\widetilde{C}_3\exp\left\{-\widetilde{C}_4n^{f(\kappa_1)}\right\}, \tag{E.17}$$

where $\widetilde{C}_3=\widetilde{C}_1+\widetilde{C}_5$ and $\widetilde{C}_4=\min\{\widetilde{C}_2,\widetilde{C}_6\}$.

E.10 Lemma 11 and its proof

Lemma 11. *Let A_j's with $j\in\mathcal{D}\subset\{1,\dots,p\}$ and B satisfy $\max_{j\in\mathcal{D}}|A_j|\leq L_5$ and $|B|\leq L_6$ for some constants $L_5,L_6>0$, and $\widehat{A}_j$ and $\widehat{B}$ be estimates of A_j and B, respectively, based on a sample of size n for each $j\in\mathcal{D}$. Assume that for any constant $C>0$, there exist constants $\widetilde{C}_1,\dots,\widetilde{C}_8>0$ except $\widetilde{C}_3\geq 0$ such that*

$$\begin{aligned}&P\left(\max_{j\in\mathcal{D}}|\widehat{A}_j-A_j|\geq Cn^{-\kappa_1}\right)\leq|D|\widetilde{C}_1\exp\left\{-\widetilde{C}_2n^{f(\kappa_1)}\right\}+\widetilde{C}_3\exp\left\{-\widetilde{C}_4n^{g(\kappa_1)}\right\},\\ &P\left(|\widehat{B}-B|\geq Cn^{-\kappa_1}\right)\leq\widetilde{C}_5\exp\left\{-\widetilde{C}_6n^{f(\kappa_1)}\right\}+\widetilde{C}_7\exp\left\{-\widetilde{C}_8n^{g(\kappa_1)}\right\}\end{aligned}$$

with $f(\kappa_1)$ and $g(\kappa_1)$ some functions of κ_1. Then for any constant $C>0$, there exist constants $\widetilde{C}_9,\dots,\widetilde{C}_{12}>0$ such that

$$\begin{aligned}P\left(\max_{j\in\mathcal{D}}|\widehat{A}_j\widehat{B}-A_jB|\geq Cn^{-\kappa_1}\right)\leq&|D|\widetilde{C}_9\exp\left\{-\widetilde{C}_{10}n^{f(\kappa_1)}\right\}\\ &+\widetilde{C}_{11}\exp\left\{-\widetilde{C}_{12}n^{g(\kappa_1)}\right\}.\end{aligned}$$

Proof of Lemma 11. Note that $\max_{j\in\mathcal{D}}|\widehat{A}_j\widehat{B}-A_jB|\leq\max_{j\in\mathcal{D}}|\widehat{A}_j(\widehat{B}-B)|+$

$\max_{j\in\mathcal{D}}|(\widehat{A}_j - A_j)B|$. Therefore, for any positive constant C,

$$\begin{aligned}P(\max_{j\in\mathcal{D}}|\widehat{A}_j\widehat{B} - A_jB| \geq Cn^{-\kappa_1}) \leq\ & P(\max_{j\in\mathcal{D}}|\widehat{A}_j(\widehat{B} - B)| \geq Cn^{-\kappa_1}/2)\\ &+ P(\max_{j\in\mathcal{D}}|(\widehat{A}_j - A_j)B| \geq Cn^{-\kappa_1}/2). \qquad \text{(E.18)}\end{aligned}$$

We first deal with the second term on the right hand side of (E.18). Since $|B| \leq L_6$, we have

$$\begin{aligned}&P(\max_{j\in\mathcal{D}}|(\widehat{A}_j - A_j)B| \geq Cn^{-\kappa_1}/2) \leq P(\max_{j\in\mathcal{D}}|\widehat{A}_j - A_j|L_6 \geq Cn^{-\kappa_1}/2)\\ &= P\{\max_{j\in\mathcal{D}}|\widehat{A}_j - A_j| \geq (2L_6)^{-1}Cn^{-\kappa_1}\}\\ &\leq |\mathcal{D}|\widetilde{C}_1 \exp\left\{-\widetilde{C}_2 n^{f(\kappa_1)}\right\} + \widetilde{C}_3 \exp\left\{-\widetilde{C}_4 n^{g(\kappa_1)}\right\} \qquad \text{(E.19)}\end{aligned}$$

with constants $\widetilde{C}_1, \widetilde{C}_2, \widetilde{C}_4 > 0$ and $\widetilde{C}_3 \geq 0$.

Next, we consider the first term on the right hand side of (E.18). Note that

$$\begin{aligned}&P(\max_{j\in\mathcal{D}}|\widehat{A}_j(\widehat{B} - B)| \geq Cn^{-\kappa_1}/2)\\ \leq\ & P(\max_{j\in\mathcal{D}}|\widehat{A}_j(\widehat{B} - B)| \geq Cn^{-\kappa_1}/2, \max_{j\in\mathcal{D}}|\widehat{A}_j| \geq L_5 + Cn^{-\kappa_1}/2)\\ &+ P(\max_{j\in\mathcal{D}}|\widehat{A}_j(\widehat{B} - B)| \geq Cn^{-\kappa_1}/2, \max_{j\in\mathcal{D}}|\widehat{A}_j| < L_5 + Cn^{-\kappa_1}/2)\\ \leq\ & P(\max_{j\in\mathcal{D}}|\widehat{A}_j| \geq L_5 + Cn^{-\kappa_1}/2)\\ &+ P(\max_{j\in\mathcal{D}}|\widehat{A}_j(\widehat{B} - B)| \geq Cn^{-\kappa_1}/2, \max_{j\in\mathcal{D}}|\widehat{A}_j| < L_5 + C)\\ \leq\ & P(\max_{j\in\mathcal{D}}|\widehat{A}_j| \geq L_5 + Cn^{-\kappa_1}/2) + P\{(L_5 + C)|\widehat{B} - B| \geq Cn^{-\kappa_1}/2\}. \qquad \text{(E.20)}\end{aligned}$$

We will bound the two terms on the right hand side of (E.20) separately. Since $\max_{j\in\mathcal{D}}|A_j| \leq$

L_5, it holds that

$$
\begin{aligned}
&P(\max_{j\in\mathcal{D}}|\widehat{A}_j| \geq L_5 + Cn^{-\kappa_1}/2) \leq P(\max_{j\in\mathcal{D}}|\widehat{A}_j - A_j| + \max_{j\in\mathcal{D}}|A_j| \geq L_5 + Cn^{-\kappa_1}/2) \\
&\leq P\{\max_{j\in\mathcal{D}}|\widehat{A}_j - A_j| \geq 2^{-1}Cn^{-\kappa_1}\} \\
&\leq |\mathcal{D}|\widetilde{C}_5 \exp\left\{-\widetilde{C}_6 n^{f(\kappa_1)}\right\} + \widetilde{C}_7 \exp\left\{-\widetilde{C}_8 n^{g(\kappa_1)}\right\}, \qquad \text{(E.21)}
\end{aligned}
$$

where $\widetilde{C}_5, \widetilde{C}_6, \widetilde{C}_8 > 0$ and $\widetilde{C}_7 \geq 0$ are some constants. We also have that

$$
\begin{aligned}
&P((L_5 + C)|\widehat{B} - B| \geq Cn^{-\kappa_1}/2) = P\{|\widehat{B} - B| \geq (2L_5 + 2C)^{-1}Cn^{-\kappa_1}\} \\
&\leq \widetilde{C}_{13} \exp\left\{-\widetilde{C}_{14} n^{f(\kappa_1)}\right\} + \widetilde{C}_{15} \exp\left\{-\widetilde{C}_{16} n^{g(\kappa_1)}\right\},
\end{aligned}
$$

where $\widetilde{C}_{13}, \ldots, \widetilde{C}_{16}$ are some positive constants. This, together with (E.18)–(E.21), entails that

$$
\begin{aligned}
&P(\max_{j\in\mathcal{D}}|\widehat{A}_j\widehat{B} - A_j B| \geq Cn^{-\kappa_1}) \\
\leq &|\mathcal{D}|\widetilde{C}_1 \exp\left\{-\widetilde{C}_2 n^{f(\kappa_1)}\right\} + \widetilde{C}_3 \exp\left\{-\widetilde{C}_4 n^{g(\kappa_1)}\right\} + |\mathcal{D}|\widetilde{C}_5 \exp\left\{-\widetilde{C}_6 n^{f(\kappa_1)}\right\} \\
&+ \widetilde{C}_7 \exp\left\{-\widetilde{C}_8 n^{g(\kappa_1)}\right\} + \widetilde{C}_{13} \exp\left\{-\widetilde{C}_{14} n^{f(\kappa_1)}\right\} + \widetilde{C}_{15} \exp\left\{-\widetilde{C}_{16} n^{g(\kappa_1)}\right\} \\
\leq &|D|\widetilde{C}_9 \exp\left\{-\widetilde{C}_{10} n^{f(\kappa_1)}\right\} + \widetilde{C}_{11} \exp\left\{-\widetilde{C}_{12} n^{g(\kappa_1)}\right\},
\end{aligned}
$$

where $\widetilde{C}_9 = \widetilde{C}_1 + \widetilde{C}_5 + \widetilde{C}_{13} > 0$, $\widetilde{C}_{10} = \min\{\widetilde{C}_2, \widetilde{C}_6, \widetilde{C}_{14}\} > 0$, $\widetilde{C}_{11} = \widetilde{C}_3 + \widetilde{C}_7 + \widetilde{C}_{15} > 0$, and $\widetilde{C}_{12} = \min\{\widetilde{C}_4, \widetilde{C}_8, \widetilde{C}_{16}\} > 0$.

E.11 Lemma 12 and its proof

Lemma 12. *Let A_j's and B_j's with $j \in \mathcal{D} \subset \{1, \ldots, p\}$ satisfy $\max_{j\in\mathcal{D}}|A_j| \leq L_7$ and $\min_{j\in\mathcal{D}}|B_j| \geq L_8$ for some constants $L_7, L_8 > 0$, and $\widehat{A}_j$ and $\widehat{B}_j$ be estimates of A_j and B_j, respectively, based on a sample of size n for each $j \in \mathcal{D}$. Assume that for any constant*

$C > 0$, there exist constants $\widetilde{C}_1, \ldots, \widetilde{C}_6 > 0$ such that

$$P\left(\max_{j\in\mathcal{D}} |\widehat{A}_j - A_j| \geq Cn^{-\kappa_1}\right) \leq |\mathcal{D}|\widetilde{C}_1 \exp\left\{-\widetilde{C}_2 n^{f(\kappa_1)}\right\} + \widetilde{C}_3 \exp\left\{-\widetilde{C}_4 n^{g(\kappa_1)}\right\},$$
$$P\left(\max_{j\in\mathcal{D}} |\widehat{B}_j - B_j| \geq Cn^{-\kappa_1}\right) \leq |D|\widetilde{C}_5 \exp\left\{-\widetilde{C}_6 n^{f(\kappa_1)}\right\}$$

with $f(\kappa_1)$ and $g(\kappa_1)$ some functions of κ_1. Then for any constant $C > 0$, there exist constants $\widetilde{C}_7, \ldots, \widetilde{C}_{10} > 0$ such that

$$P\left(\max_{j\in\mathcal{D}} \left|\widehat{A}_j/\widehat{B}_j - A_j/B_j\right| \geq Cn^{-\kappa_1}\right) \leq |\mathcal{D}|\widetilde{C}_7 \exp\left\{-\widetilde{C}_8 n^{f(\kappa_1)}\right\}$$
$$+ \widetilde{C}_9 \exp\left\{-\widetilde{C}_{10} n^{g(\kappa_1)}\right\}.$$

Proof of Lemma 12. Since $\min_{j\in\mathcal{D}} B_j \geq L_8 > 0$, there exists some constant L_0 such that $0 < L_0 < L_8$. Note that, for any positive constant C,

$$\begin{aligned}
&P(\max_{j\in\mathcal{D}} |\frac{\widehat{A}_j}{\widehat{B}_j} - \frac{A_j}{B_j}| \geq Cn^{-\kappa_1}) \\
\leq &P(\max_{j\in\mathcal{D}} |\frac{\widehat{A}_j}{\widehat{B}_j} - \frac{A_j}{B_j}| \geq Cn^{-\kappa_1}, \min_{j\in\mathcal{D}} |\widehat{B}_j| \leq L_8 - L_0 n^{-\kappa_1}) \\
&\quad + P(\max_{j\in\mathcal{D}} |\frac{\widehat{A}_j}{\widehat{B}_j} - \frac{A_j}{B_j}| \geq Cn^{-\kappa_1}, \min_{j\in\mathcal{D}} |\widehat{B}_j| > L_8 - L_0 n^{-\kappa_1}) \\
\leq &P(\min_{j\in\mathcal{D}} |\widehat{B}_j| \leq L_8 - L_0 n^{-\kappa_1}) + P(\max_{j\in\mathcal{D}} |\frac{\widehat{A}_j}{\widehat{B}_j} - \frac{A_j}{B_j}| \geq Cn^{-\kappa_1}, \min_{j\in\mathcal{D}} |\widehat{B}_j| > L_8 - L_0).
\end{aligned} \tag{E.22}$$

Let us consider the first term on the right hand side of (E.22). Since $\min_{j\in\mathcal{D}} B_j \geq L_8$, it holds that for any positive constant C,

$$\begin{aligned}
P(\min_{j\in\mathcal{D}} |\widehat{B}_j| \leq L_8 - L_0 n^{-k}) &\leq P(\min_{j\in\mathcal{D}} |B_j| - \max_{j\in\mathcal{D}} |\widehat{B}_j - B_j| \leq L_8 - L_0 n^{-\kappa_1}) \\
&\leq P(\max_{j\in\mathcal{D}} |\widehat{B}_j - B_j| \geq L_0 n^{-\kappa_1}) \leq |\mathcal{D}|\widetilde{C}_1 \exp\left\{-\widetilde{C}_2 n^{f(\kappa_1)}\right\},
\end{aligned} \tag{E.23}$$

where $\widetilde{C}_1$ and $\widetilde{C}_2$ are some positive constants.

The second term on the right hand side of (E.22) can be bounded as

$$
\begin{aligned}
&P(\max_{j\in\mathcal{D}}|\frac{\widehat{A}_j}{\widehat{B}_j}-\frac{A_j}{B_j}|\geq Cn^{-\kappa_1},\ \min_{j\in\mathcal{D}}|\widehat{B}_j|>L_8-L_0)\\
\leq &P(\max_{j\in\mathcal{D}}|\frac{\widehat{A}_j}{\widehat{B}_j}-\frac{A_j}{\widehat{B}_j}|\geq Cn^{-\kappa_1}/2,\ \min_{j\in\mathcal{D}}|\widehat{B}_j|>L_8-L_0)\\
&\quad+P(\max_{j\in\mathcal{D}}|\frac{A_j}{\widehat{B}_j}-\frac{A_j}{B_j}|\geq Cn^{-\kappa_1}/2,\ \min_{j\in\mathcal{D}}|\widehat{B}_j|>L_8-L_0)\\
\leq &P\{\max_{j\in\mathcal{D}}|\widehat{A}_j-A_j|\geq 2^{-1}(L_8-L_0)Cn^{-\kappa_1}\}\\
&\quad+P\{\max_{j\in\mathcal{D}}|\widehat{B}_j-B_j|\geq (2L_7)^{-1}(L_8-L_0)L_8Cn^{-\kappa_1}\}\\
\leq &|\mathcal{D}|\widetilde{C}_3\exp\left\{-\widetilde{C}_4n^{f(\kappa_1)}\right\}+\widetilde{C}_5\exp\left\{-\widetilde{C}_6n^{g(\kappa_1)}\right\}+|\mathcal{D}|\widetilde{C}_{11}\exp\left\{-\widetilde{C}_{12}n^{f(\kappa_1)}\right\}, \qquad \text{(E.24)}
\end{aligned}
$$

where $\widetilde{C}_3,\ldots,\widetilde{C}_6$, $\widetilde{C}_{11}$, and $\widetilde{C}_{12}$ are some positive constants. Combining (E.22)–(E.24) results in

$$
\begin{aligned}
P(\max_{j\in\mathcal{D}}|\widehat{A}_j/\widehat{B}_j-A_j/B_j|\geq Cn^{-\kappa_1})\leq |\mathcal{D}|\widetilde{C}_7\exp\left\{-\widetilde{C}_8n^{f(\kappa_1)}\right\}\\
+\widetilde{C}_9\exp\left\{-\widetilde{C}_{10}n^{g(\kappa_1)}\right\},
\end{aligned}
$$

where $\widetilde{C}_7=\widetilde{C}_1+\widetilde{C}_3+\widetilde{C}_{11}>0$, $\widetilde{C}_8=\min\{\widetilde{C}_2,\widetilde{C}_4,\widetilde{C}_{12}\}>0$, $\widetilde{C}_9=\widetilde{C}_5>0$, and $\widetilde{C}_{10}=\widetilde{C}_6>0$. This completes the proof of Lemma 12.