Supplementary Materials for "Joint Estimation of Multiple High-dimensional Precision Matrices"

In this supplementary file, we include additional simulations studies and the proofs of the theorems in the paper.

S.1. ADDITIONAL SIMULATION STUDIES

In addition to two graph models (ER model and WS model) discussed in Section 4, we generate common graph structures based on another two commonly-used graph models, Barabási and Albert model and geometric random graph model. For Barabási and Albert Model, a new vertex is added to the existing graph each time and the new vertex is connected to an existing old vertex with a probability proportional to the degree of the existing vertices plus one. For geometric random graph, p points are dropped on a unit square. Two vertex will be connected with an undirected edge if and only if their corresponding points are closer to each other than a radius of 0.05.

Based on each graph model (ER model or WS model), a common graph structure is generated. The following data generation and evaluation procedure is the same as described in Section 4. Figure S1 shows the ROC curves of all the methods. Clearly, our joint estimation method (MPE) outperforms all other methods. We also applied tuning parameter selection procedure on all these methods and compared the performance of the resulting estimators. Table 1 and Table 2 list the results. our joint estimation method (MPE) has the highest MCC, so that it can recover the graphical structure better.

S.2. PROOFS OF THEOREMS

We first state a lemma which follows from Theorem 1 in Zaïtsev (1987).

Lemma 2. *Let $|\cdot|_K$ denote the Euclidean norm of K dimensional vector. Suppose $X_1, \dots, X_n$ are independent K-dimensional random vectors satisfying $\mathsf{E}X_i = 0$ and $|X_i|_K \le M$ for $1 \le i \le n$. We have for any $\delta > 0$ and $x > \delta$*

$$\mathsf{P}\Big(\Big|\sum_{k=1}^{n} X_k\Big|_K \ge x\Big) \le \mathsf{P}\Big\{|N|_K \ge (x-\delta)/\lambda_{\max}^{1/2}\Big\} + c_1 K^{5/2}\exp(-c_2 K^{-5/2}\delta/M),$$

where $\lambda_{\max}$ is the largest eigenvalue of $\mathsf{Cov}(\sum_{k=1}^{n} X_k)$, N is a K-dimensional standard Gaussian random vector and c_1, c_2 are absolute positive constants.

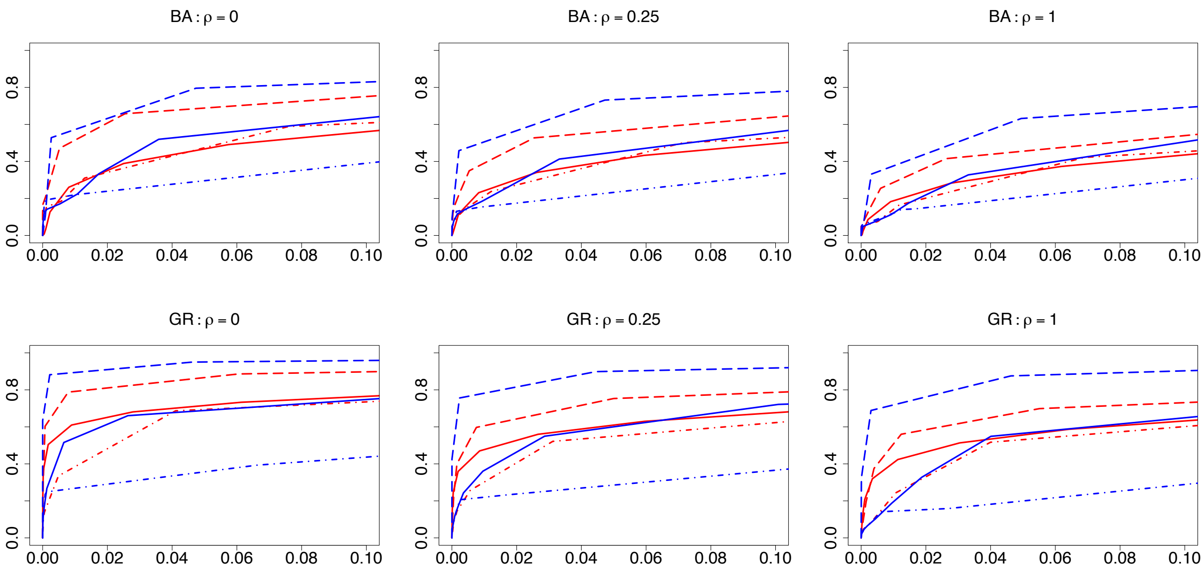

Figure S1: *Receiver operator characteristic curves for graph structure recovery for the simulated Barabási and Albert graphs (first row), and the geometric random graphs (the second row). The x-axis and y-axis of each panel are average false positive rate and average sensitivity across* $K = 3$ *groups. Red solid line:* CLIME*; red dot-dashed line:* GLASSO*; red long-dashed line:* JEMGM*; blue solid line:* FGL*; blue dot-dashed line:* GGL*; blue long-dashed line:* MPE.

Table 1: Simulation results for data generated based on the Barabási and Albert graph with different ratios of the number of individual-specific edges to the number of shared edges. Results are based on 100 replications. The numbers in the brackets are standard errors. CLIME: method of Cai et al. (2011); GLASSO: graphical Lasso; JEMGM: method of Guo et al. (2011); FGLand GGL: methods of Danaher et al. (2014); MPE: proposed method.

Model(ρ)	Method	Performance					
		L_1	L_2	L_F	SEN	SPE	MCC
BA(0)	CLIME	18.87(0.14)	6.31(0.06)	28.76(0.72)	0.26(0.02)	0.99(0.00)	0.25(0.01)
	GLASSO	18.48(1.23)	6.42(0.31)	22.22(0.68)	0.11(0.03)	1.00(0.00)	0.28(0.02)
	JEMGM	13.34(1.36)	5.39(0.36)	22.06(3.15)	0.24(0.08)	1.00(0.00)	0.40(0.03)
	FGL	17.79(0.69)	6.23(0.21)	21.30(0.48)	0.14(0.09)	0.99(0.04)	0.28(0.04)
	GGL	18.11(0.81)	6.39(0.19)	22.09(0.44)	0.15(0.03)	1.00(0.00)	0.34(0.03)
	MPE	17.11(1.27)	5.54(0.10)	27.68(1.38)	0.53(0.05)	0.99(0.08)	0.60(0.06)
BA(1/4)	CLIME	18.45(0.17)	6.41(0.06)	30.15(0.76)	0.23(0.02)	0.99(0.00)	0.24(0.02)
	GLASSO	18.25(0.61)	6.47(0.22)	23.93(0.61)	0.07(0.02)	1.00(0.00)	0.21(0.01)
	JEMGM	13.55(0.81)	5.27(0.30)	22.19(2.38)	0.22(0.07)	1.00(0.00)	0.36(0.03)
	FGL	18.16(0.78)	6.43(0.29)	23.29(0.65)	0.12(0.13)	0.99(0.06)	0.22(0.03)
	GGL	18.01(0.42)	6.42(0.16)	23.89(0.39)	0.11(0.02)	1.00(0.00)	0.26(0.02)
	MPE	16.80(1.85)	5.69(0.10)	30.06(1.95)	0.46(0.06)	0.99(0.09)	0.57(0.06)
BA(1)	CLIME	21.97(0.24)	7.18(0.07)	35.38(0.75)	0.18(0.02)	0.99(0.00)	0.21(0.01)
	GLASSO	21.66(0.54)	7.27(0.12)	29.12(0.30)	0.04(0.00)	1.00(0.00)	0.19(0.01)
	JEMGM	18.50(1.46)	6.73(0.45)	34.36(3.91)	0.07(0.02)	1.00(0.00)	0.23(0.02)
	FGL	21.85(0.81)	7.13(0.50)	28.11(1.31)	0.11(0.20)	0.97(0.09)	0.18(0.02)
	GGL	21.72(0.83)	7.30(0.19)	28.95(0.33)	0.05(0.01)	1.00(0.00)	0.20(0.02)
	MPE	19.39(0.28)	6.31(0.09)	34.53(0.72)	0.33(0.02)	1.00(0.00)	0.47(0.02)

Table 2: Simulation results for data generated based on the geometric random graph with different ratios of the number of individual-specific edges to the number of shared edges. Results are based on 100 replications. The numbers in the brackets are standard errors. CLIME: method of Cai et al. (2011); GLASSO: graphical Lasso; JEMGM: method of Guo et al. (2011); FGLand GGL: methods of Danaher et al. (2014); MPE: proposed method.

Model(ρ)	Method	Performance					
		L_1	L_2	L_F	SEN	SPE	MCC
GR(0)	CLIME	5.48(0.11)	3.84(0.11)	19.43(0.89)	0.49(0.10)	1.00(0.00)	0.56(0.05)
	GLASSO	5.56(0.16)	3.83(0.21)	15.47(0.55)	0.19(0.10)	1.00(0.00)	0.30(0.03)
	JEMGM	4.98(0.17)	3.16(0.17)	12.40(0.52)	0.71(0.10)	0.99(0.00)	0.63(0.07)
	FGL	5.79(0.20)	4.00(0.16)	15.00(0.57)	0.30(0.14)	1.00(0.00)	0.39(0.04)
	GGL	5.52(0.13)	3.80(0.12)	15.41(0.29)	0.24(0.06)	1.00(0.00)	0.40(0.03)
	MPE	4.77(0.24)	3.57(0.18)	18.32(1.05)	0.81(0.11)	1.00(0.00)	0.81(0.02)
GR(1/4)	CLIME	6.92(0.13)	4.54(0.09)	22.88(0.71)	0.41(0.06)	0.99(0.00)	0.45(0.04)
	GLASSO	7.05(0.13)	4.51(0.18)	19.11(0.68)	0.19(0.06)	1.00(0.00)	0.27(0.02)
	JEMGM	6.19(0.19)	3.64(0.14)	15.45(0.69)	0.55(0.08)	0.99(0.00)	0.52(0.02)
	FGL	7.16(0.15)	4.46(0.10)	18.86(0.36)	0.24(0.05)	1.00(0.00)	0.31(0.02)
	GGL	7.14(0.10)	4.59(0.11)	19.27(0.38)	0.22(0.05)	1.00(0.00)	0.32(0.02)
	MPE	6.28(0.11)	4.15(0.07)	21.94(0.54)	0.76(0.02)	1.00(0.00)	0.76(0.02)
GR(1)	CLIME	7.84(0.14)	4.66(0.09)	26.75(0.74)	0.38(0.05)	0.99(0.00)	0.39(0.02)
	GLASSO	8.98(0.39)	4.52(0.12)	23.07(0.77)	0.19(0.07)	0.99(0.00)	0.21(0.04)
	JEMGM	7.61(0.24)	3.83(0.18)	19.05(0.89)	0.52(0.08)	0.99(0.00)	0.46(0.01)
	FGL	9.16(0.29)	4.56(0.08)	23.27(0.49)	0.16(0.05)	0.99(0.00)	0.19(0.03)
	GGL	8.95(0.27)	4.59(0.06)	23.73(0.30)	0.14(0.03)	0.99(0.00)	0.18(0.02)
	MPE	7.04(0.13)	4.31(0.08)	25.31(0.53)	0.69(0.02)	1.00(0.00)	0.72(0.01)

Proof of Theorem 1. Suppose that the true $\mathbf{\Omega}^{(k)}$ belong to the above feasible set, that is

$$\max_{ij}\Big\{\sum_{k=1}^{K} w_k|(\hat{\Sigma}^{(k)}\mathbf{\Omega}^{(k)}-I)_{ij}|^2\Big\}^{1/2}\le\lambda_n. \tag{S1}$$

We have

$$\begin{aligned}
&\max_{i,j}\Big\{\sum_{k=1}^{K} w_k|(\hat{\mathbf{\Omega}}_1^{(k)}-\mathbf{\Omega}^{(k)})_{ij}|^2\Big\}^{1/2}\\
=&\max_{i,j}\Big[\sum_{k=1}^{K} w_k|\{(\mathbf{\Omega}^{(k)}\hat{\Sigma}^{(k)}-I)\hat{\mathbf{\Omega}}_1^{(k)}-\mathbf{\Omega}^{(k)}(\hat{\Sigma}^{(k)}\hat{\mathbf{\Omega}}_1^{(k)}-I)\}_{ij}|^2\Big]^{1/2}\\
\le&\max_{i,j}\Big[\sum_{k=1}^{K} w_k|\{(\mathbf{\Omega}^{(k)}\hat{\Sigma}^{(k)}-I)\hat{\mathbf{\Omega}}_1^{(k)}\}_{ij}|^2\Big]^{1/2}+\max_{i,j}\Big[\sum_{k=1}^{K} w_k|\{\mathbf{\Omega}^{(k)}(\hat{\Sigma}^{(k)}\hat{\mathbf{\Omega}}_1^{(k)}-I)\}_{ij}|^2\Big]^{1/2}\\
=:&I_1+I_2.
\end{aligned}$$

Note that

$$\{(\mathbf{\Omega}^{(k)}\hat{\Sigma}^{(k)}-I)\hat{\mathbf{\Omega}}_1^{(k)}\}_{ij}=\delta_{i\cdot}^{(k)}\hat{\omega}_{1\cdot j}^{(k)},$$

where $\delta_{i\cdot}^{(k)}=:(\delta_{i1}^{(k)},\ldots,\delta_{ip}^{(k)})$ is the i-th row of $\mathbf{\Omega}^{(k)}\hat{\Sigma}^{(k)}-I$ and $\hat{\omega}_{1\cdot j}^{(k)}=(\hat{\omega}_{11j}^{(k)},\ldots,\hat{\omega}_{1pj}^{(k)})^T$ is the j-th column of $\hat{\mathbf{\Omega}}_1^{(k)}$. We have

$$\begin{aligned}
I_1 &\le \max_{i,j}\Big(\sum_{k=1}^{K} w_k\sum_{1\le l,m\le p}\delta_{il}^{(k)}\delta_{im}^{(k)}\hat{\omega}_{1lj}^{(k)}\hat{\omega}_{1mj}^{(k)}\Big)^{1/2}\\
&\le \max_{i,j}\Big(\sum_{1\le l,m\le p}\sum_{k=1}^{K} w_k|\delta_{il}^{(k)}\delta_{im}^{(k)}||\hat{\omega}_{1lj}^{(k)}\hat{\omega}_{1mj}^{(k)}|\Big)^{1/2}.
\end{aligned}$$

Assume that $w_K|\delta_{il}^{(K)}\delta_{im}^{(K)}|\le\cdots\le w_1|\delta_{il}^{(1)}\delta_{im}^{(1)}|$. Since by (S1),

$$\sum_{k=1}^{K} w_k|\delta_{il}^{(k)}\delta_{im}^{(k)}|\le 2^{-1}\sum_{k=1}^{K} w_k(|\delta_{il}^{(k)}|^2+|\delta_{im}^{(k)}|^2)\le\max_{i,j}\Big(\sum_{k=1}^{K} w_k|\delta_{ij}^{(k)}|^2\Big)\le\lambda_n^2,$$

we have

$$\max_{i,l,m} w_k|\delta_{il}^{(k)}\delta_{im}^{(k)}|\le k^{-1}\max_{i,l,m}\sum_{j=1}^{k} w_j|\delta_{il}^{(j)}\delta_{im}^{(j)}|\le\lambda_n^2/k.$$

Therefore

$$I_1 \le \max_{i,j}\Big(\sum_{1\le l,m\le p}\sum_{k=1}^{K} k^{-1}|\hat{\omega}_{1lj}^{(k)}\hat{\omega}_{1mj}^{(k)}|\Big)^{1/2}\lambda_n$$

$$\leq \quad \Big(\sum_{k=1}^{K} k^{-1}\hat{M}_n^2\Big)^{1/2}\lambda_n \leq (\log K)^{1/2}\hat{M}_n\lambda_n, \tag{S2}$$

where $\hat{M}_n = \max_{1\leq k\leq K} \|\hat{\boldsymbol{\Omega}}_1^{(k)}\|_{l_1}$. Similarly, we can show that

$$I_2 \leq (\log K)^{1/2} M_n\lambda_n. \tag{S3}$$

By the definition of $\hat{\boldsymbol{\Omega}}_1^{(k)}$, we have $\hat{M}_n \leq M_n$.

So it suffices to prove (S1) holds with probability greater than $1-O(p^{-\epsilon})$. Assume a new set of random variables $\tilde{X}_l^{(k)}$ independently follows $N(0,\Sigma^{(k)})$. Let $Y_{lij}^{(k)} = w_k^{1/2}n_k^{-1}\{(\tilde{X}_l^{(k)}\tilde{X}_l^{(k)\prime}\boldsymbol{\Omega}^{(k)})_{ij} - \tilde{e}_{ij}\}$ and $Y_{lij} = (Y_{lij}^{(1)},\ldots,Y_{lij}^{(K)})$, where $\tilde{e}_{ij} = e_{ij}n_k/(n_k-1)$. When $l \geq n_k - 1$, we set $Y_{lij}^{(k)} = 0$. Let $|\cdot|_K$ denotes the Euclidean norm of K dimensional vector. Because $\sum_{l=1}^{n_k}(X_l^{(k)}-\bar{X}^{(k)})(X_l^{(k)}-\bar{X}^{(k)})'$ follows the same Wishart distribution as $\sum_{l=1}^{n_k-1}\tilde{X}_l^{(k)}\tilde{X}_l^{(k)\prime}$, we have $\Big\{\sum_{k=1}^{K} w_k|(\hat{\Sigma}^{(k)}\beta^{(k)}-e_j)_i|^2\Big\}^{1/2}$ follows the same distribution as $\Big|\sum_{l=1}^{n} Y_{lij}\Big|_K$. For $1\leq l\leq n$,$1\leq k\leq K$ and $1\leq i,j\leq p$, let

$$\hat{Y}_{lij}^{(k)} = Y_{lij}^{(k)} I\left\{|Y_{lij}^{(k)}| \leq (n\log p)^{-1/2}K^{1/2-a}\right\} - \mathsf{E}Y_{lij}^{(k)} I\left\{|Y_{lij}^{(k)}| \leq (n\log p)^{-1/2}K^{1/2-a}\right\}$$

and $\hat{Y}_{lij} = (\hat{Y}_{lij}^{(1)},\ldots,\hat{Y}_{lij}^{(K)})$. Note that $n\max_{i,j}\Big|\mathsf{E}(Y_{lij}-\hat{Y}_{lij})\Big|_K = 0$. We have for any $\delta>0$,

$$\mathsf{P}\Big(\Big|\sum_{l=1}^{n} Y_{lij}\Big|_K \geq \lambda_n\Big) \leq \mathsf{P}\Big(\Big|\sum_{l=1}^{n}\hat{Y}_{lij}\Big|_K \geq (1-\delta)\lambda_n\Big) + (\max_k n_k)K\max_{1\leq k\leq K}\mathsf{P}\Big\{|Y_{lij}^{(k)}| \geq \left(\frac{K^{1-2a}}{n\log p}\right)^{1/2}\Big\}. \tag{S4}$$

Let $Z_{lij}^{(k)} = (\tilde{X}_l^{(k)}\tilde{X}_l^{(k)\prime}\boldsymbol{\Omega}^{(k)})_{ij} - \tilde{e}_{ij}$. We have for some constant $\eta>0$,

$$\begin{aligned}
&(\max_k n_k)K\max_{1\leq k\leq K}\mathsf{P}\Big\{|Y_{lij}^{(k)}| \geq \left(\frac{K^{1-2a}}{n\log p}\right)^{1/2}\Big\}\\
\leq& Cn\max_{1\leq k\leq K}\mathsf{P}\left\{|Z_{lij}^{(k)}| \geq \left(\frac{n}{K^{2a}\log p}\right)^{1/2}\right\}\\
\leq& C\exp\left\{\log n - \eta\left(\frac{n}{K^{2a}\log p}\right)^{1/2}\right\} = o(1)
\end{aligned}$$

By Condition (C3), it is easy to show that

$$\lambda_{\max}\left\{\sum_{l=1}^{n}\mathsf{Cov}(\hat{Y}_{lij})\right\} \leq \{1+o(1)\}(M_1+1)/n$$

uniformly for $1\leq i,j\leq p$. Therefore it follows from (C1), Lemma 2, the tail probability of Chi-squared distribution and some tedious calculations that

$$\mathsf{P}\Big\{\Big|\sum_{l=1}^{n}\hat{Y}_{lij}\Big|_K \geq (1-\delta)\lambda_n\Big\} \leq C\exp\{-C(\log p - K)\} + C\exp\left\{\frac{5}{2}\log K - C_2K^{a-4}(\log p)\right\} = o(1). \tag{S5}$$

Combining (S4)-(S5), we prove that (S1) holds. $\square$

Proof of Proposition 1. Consider $K = 1$ first. Define $\mathcal{V}(M, M_n) = \{\mathbf{\Omega} : \lambda_{\max}(\mathbf{\Omega})/\lambda_{\min}(\mathbf{\Omega}) \leq M,\ \|\mathbf{\Omega}\|_1 \asymp M_n = o\{(n/\log p)^{1/2}\}\}$. The proof of Proposition 1 follows the proof of Theorem 5 in Ren et al. (2014). Construct $\mathbf{\Omega}_0$ and $\mathbf{\Omega}_m$ in the same way. In the proof, $\max_{1\leq j\leq p}\sum_{i\neq j} I(\omega_{ij} \neq 0) \leq k_{n,p}$. Let $k_{n,p}$ satisfy $k_{n,p}(\log p/n_1)^{1/2} \to \infty$ and $k_{n,p} = o(n/\log p)$. Then $M_n \asymp \|\mathbf{\Omega}_m\|_1 \asymp (\log p/n_1)^{1/2}k_{n,p}$. Following the proof of Theorem 5, Eqn. (59) in Ren et al. (2014) will lead to

$$\inf_{1\leq m\leq m^*}|\omega_{11}^{(m)} - \omega_{11}^{(0)}| \geq C_3 M_n(\log p/n_1)^{1/2}.$$

$$\inf_{1\leq m\leq m^*}|\omega_{12}^{(m)} - \omega_{12}^{(0)}| \geq C_4 M_n(\log p/n_1)^{1/2}.$$

Following the rest of the proof, for some constant $c_1 > 0$ and $\alpha_1 > 0$, we have

$$\inf_{i,j}\inf_{\hat{\omega}_{ij}}\sup_{\mathbf{\Omega}\in\mathcal{V}(M,M_n)} \mathsf{P}\left\{|\hat{\omega}_{ij} - \omega_{ij}| \geq c_1 M_n(\log p/n_1)^{1/2}\right\} \geq \alpha_1. \tag{S6}$$

Now consider the case $K \geq 1$. For each $\mathbf{\Omega}^{(k)} \in \mathcal{V}(M, M_n)$ and separate estimation method, (S6) holds. Therefore, for some c_0 satisfying $c_0 K\log p/n \leq \inf_k \log p/n_k$,

$$\begin{aligned}
&\inf_{i,j}\inf_{\tilde{\mathcal{U}}}\sup_{\mathcal{U}}\mathsf{P}\left\{\left\{\sum_{k=1}^K w_k|(\tilde{\mathbf{\Omega}}^{(k)} - \mathbf{\Omega}^{(k)})_{i,j}|^2\right\}^{1/2} \geq c_0 M_n(K\log p/n)^{1/2}\right\}\\
\geq\ &\inf_{i,j}\inf_{\tilde{\mathcal{U}}}\sup_{\mathcal{U}}\mathsf{P}\left\{\inf_k|\hat{\omega}_{ij}^{(k)} - \omega_{ij}^{(k)}| \geq c_0 M_n(K\log p/n)^{1/2}\right\}\\
\geq\ &\inf_{i,j}\inf_{\tilde{\mathcal{U}}}\sup_{\mathcal{U}}\prod_k \mathsf{P}\left\{|\hat{\omega}_{ij}^{(k)} - \omega_{ij}^{(k)}| \geq c_1 M_n(\log p/n_k)^{1/2}\right\}\\
\geq\ &\alpha_1^K,
\end{aligned}$$

which leads to Proposition 1.

□

Proof of Theorem 2. Suppose that

$$\max_{i,j}\left\{\sum_{k=1}^K w_k|(\hat{\mathbf{\Omega}}^{(k)} - \mathbf{\Omega}^{(k)})_{ij}|^2\right\}^{1/2} \leq CM_n\left(\frac{\log K\cdot\log p}{n}\right)^{1/2}.$$

For $i \in S_j^c$, $\max_{i,j}\left\{\sum_{k=1}^K w_k|(\hat{\mathbf{\Omega}}^{(k)})_{ij}|^2\right\}^{1/2} \leq CM_n(\log K\log p/n)^{1/2}$. Thus $(\breve{\mathbf{\Omega}}^{(k)})_{ij} = 0$ for $i \in S_j^c$. It yields that

$$\sum_{i=1}^p\left\{\sum_{k=1}^K w_k(\breve{\mathbf{\Omega}}^{(k)} - \mathbf{\Omega}^{(k)})_{ij}^2\right\}^{1/2} \leq \sum_{i\in S_j}\left\{\sum_{k=1}^K w_k(\breve{\mathbf{\Omega}}^{(k)} - \mathbf{\Omega}^{(k)})_{ij}^2\right\}^{1/2} + \sum_{i\in S_j^c}\left\{\sum_{k=1}^K w_k(\breve{\mathbf{\Omega}}^{(k)})_{ij}^2\right\}^{1/2}$$

$$\leq \quad CM_n s_0(p) \left(\frac{\log K \cdot \log p}{n} \right)^{1/2}.$$

Theorem 2 then follows from Theorem 1. $\square$