Supplementary Materials to "Large-Scale Multiple Testing for Matrix-Valued Data under Cross-Dependency"

Shiyu Zhang, Xu Han, and Sanat K. Sarkar

Temple University

Supplementary Material

This file contains the theoretical proofs for Theorems 1 & 2 and Propositions 1 & 2 as well as additional figures for numerical results.

S1 Technical Proofs

Lemma 1. *For any matrix $\hat{\mathbf{\Sigma}}$ with eigenvalues $\{\lambda_i\}$ (in non-increasing order) and the corresponding eigenvectors $\{\boldsymbol{\gamma}_i\}$. Let $\widehat{\mathbf{\Sigma}}$ be an estimate of $\mathbf{\Sigma}$, with corresponding eigenvalues $\{\widehat{\lambda}_i\}$ and eigenvectors $\{\widehat{\boldsymbol{\gamma}}_i\}$. We have*

$$|\widehat{\lambda}_i - \lambda_i| \le \|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\| \quad \textit{and} \quad \|\widehat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\| \le \frac{\sqrt{2}\|\widehat{\mathbf{\Sigma}} - \mathbf{\Sigma}\|}{\min(|\widehat{\lambda}_{i-1} - \lambda_i|, |\lambda_i - \widehat{\lambda}_{i+1}|)}.$$

The first result is referred to Weyl's Theorem (Horn & Johnson, 1990) and the second result is called the $\sin\theta$ Theorem (Davis & Kahan, 1970).

Proof of Proposition 1:

We further denote those diagonal elements of $\mathbf{T}^{1/2}$ as $\{\sqrt{T_i}\}_{i=1}^{pq}$, then

$$\begin{aligned} vec(\mathbf{X})|\{T_l\}_{l=1}^{pq} &\sim N_{pq}(\mathbf{T}^{1/2}vec(\boldsymbol{\mu}^\star), \mathbf{T}^{1/2}\boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1 \mathbf{T}^{1/2}) \\ T_i &\sim InverseGamma(\frac{n+m-2}{2}, \frac{n+m-2}{2}), i = 1, \cdots, pq \end{aligned}$$

where we let $\boldsymbol{\mu}^\star = \sqrt{\frac{nm}{n+m}}(\boldsymbol{\mu}_y - \boldsymbol{\mu}_z) \circ \boldsymbol{\Sigma}$.

Let

$$\Delta \equiv \frac{1}{pq} \sum_{l \in \{\text{true null}\}} \mathbf{I}(P_l \leq t|\mathbf{W}) - \frac{1}{pq} \sum_{l \in \{\text{true null}\}} [\Phi(a_l(z_{t/2}+\zeta_l)) + \Phi(a_l(z_{t/2}-\zeta_l))].$$

We want to show that $\Delta = O_p((pq)^{-\delta/2}) + O_p((n+m-2)^{-1/2})$. To prove this, it suffices to show that

$$\begin{aligned} & |\frac{1}{pq} \sum_{l \in \{\text{true null}\}} \mathbf{I}(P_l \leq t|\mathbf{W}) - \frac{1}{pq} \sum_{l \in \{\text{true null}\}} P(P_l \leq t|\mathbf{W})| \\ = \; & O_p((pq)^{-\delta/2}) + O_p((n+m-2)^{-1/2}), \end{aligned} \tag{S1.1}$$

and

$$\begin{aligned} & |\frac{1}{pq} \sum_{i \in \{\text{true null}\}} P(P_i \leq t|\mathbf{W}) \\ & \quad - \frac{1}{pq} \sum_{l \in \{\text{true null}\}} [\Phi(a_l(z_{t/2}+\zeta_l)) + \Phi(a_l(z_{t/2}-\zeta_l))]| \\ = \; & O((n+m-2)^{-1/2}). \end{aligned} \tag{S1.2}$$

To show (S1.1), note that

$$\begin{aligned}
& Var((pq)^{-1} \sum_{l \in \{\text{true null}\}} \mathbf{I}(P_l \leq t|\mathbf{W})) \\
= & \frac{1}{(pq)^2} \sum_{i \in \{\text{true null}\}} Var(\mathbf{I}(P_i \leq t|\mathbf{W})) \\
& + \frac{1}{(pq)^2} \sum_{i,j \in \{\text{true null}\}, i \neq j} cov(\mathbf{I}(P_i \leq t|\mathbf{W}), \mathbf{I}(P_j \leq t|\mathbf{W})) \\
= & O((pq)^{-1}) + O((pq)^{-1}(\sum_{l=h+1}^{pq} \theta_l^2)^{1/2}) + O((n+m-2)^{-1})
\end{aligned}$$

based on the proof of Theorem 2 in Fan and Han (2017). By the condition in Proposition 1, $(pq)^{-1}(\sum_{l=h+1}^{pq} \theta_l^2)^{1/2} = O((pq)^{-\delta})$. Therefore, expression (S1.1) is $O_p((pq)^{-\delta/2}) + O_p((n+m-2)^{-1/2})$. By Lemma 4 in Fan and Han (2017), the conclusion in expression (S1.2) is also correct. Combining with the condition that $R(t)^{-1} = O_p((pq)^{-(1-\zeta)})$, the proof is now complete.

Proof of Theorem 1:

Define an infeasible estimator $\widetilde{\mathbf{W}}_2 = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T vec(\mathbf{X})$. Denote $\mathrm{FDP}_2(t)$ as the formula in $\mathrm{FDP}_{A,1}(t)$ with using the infeasible estimator $\widetilde{\mathbf{W}}_2$. Define the infeasible estimator $\widetilde{\mathbf{W}}_1 = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T vec(\widetilde{\mathbf{X}})$. Denote $\mathrm{FDP}_1(t)$ as the formula in $\mathrm{FDP}_{A,1}(t)$ by using the infeasible estimator $\widetilde{\mathbf{W}}_1$. Then we have

$$\begin{aligned}
\widehat{\mathrm{FDP}}_1(t) - \mathrm{FDP}_{A,1}(t) &= \widehat{\mathrm{FDP}}_1(t) - \mathrm{FDP}_2(t) + \mathrm{FDP}_2(t) - \mathrm{FDP}_1(t) \\
&\quad + \mathrm{FDP}_1(t) - \mathrm{FDP}_{A,1}(t).
\end{aligned}$$

We will first evaluate $\widehat{\mathrm{FDP}}_1(t) - \mathrm{FDP}_2(t)$. Define

$$\begin{aligned}\Delta_1 &= \sum_{l=1}^{pq} [\Phi(\widehat{a}_l(z_{t/2} + \widehat{\mathbf{f}}_l^T \widehat{\mathbf{W}})) - \Phi(a_l(z_{t/2} + \mathbf{f}_l^T \widetilde{\mathbf{W}}_2))],\\ \Delta_2 &= \sum_{l=1}^{pq} [\Phi(\widehat{a}_l(z_{t/2} - \widehat{\mathbf{f}}_l^T \widehat{\mathbf{W}})) - \Phi(a_l(z_{t/2} - \mathbf{f}_l^T \widetilde{\mathbf{W}}_2))].\end{aligned}$$

Then we have $\widehat{\mathrm{FDP}}_1(t) - \mathrm{FDP}_2(t) = R(t)^{-1}[\Delta_1 + \Delta_2]$.

We further let $\Delta_1 = \sum_{l=1}^{pq} \Delta_{1l}$, where

$$\begin{aligned}\Delta_{1l} &= \Phi(\widehat{a}_l(z_{t/2} + \widehat{\mathbf{f}}_l^T \widehat{\mathbf{W}})) - \Phi(\widehat{a}_l(z_{t/2} + \mathbf{f}_l^T \widetilde{\mathbf{W}}_2))\\ &\quad + \Phi(\widehat{a}_l(z_{t/2} + \mathbf{f}_l^T \widetilde{\mathbf{W}}_2)) - \Phi(a_l(z_{t/2} + \mathbf{f}_l^T \widetilde{\mathbf{W}}_2))\\ &\equiv \Delta_{11l} + \Delta_{12l},\end{aligned}$$

where the first part focuses on the difference between $\widehat{\mathbf{f}}_l^T \widehat{\mathbf{W}}$ and $\mathbf{f}_l^T \widetilde{\mathbf{W}}_2$ and the second part focuses on the difference between $\widehat{a}_l$ and a_l.

For Δ_{11l}, by the mean value theorem, there exists ζ_l between $\widehat{\mathbf{f}}_l^T \widehat{\mathbf{W}}$ and $\mathbf{f}_l^T \widetilde{\mathbf{W}}_2$ such that $\Delta_{11l} = \phi(\widehat{a}_l(z_{t/2} + \zeta_l))\widehat{a}_l(\widehat{\mathbf{f}}_l^T \widehat{\mathbf{W}} - \mathbf{f}_l^T \widetilde{\mathbf{W}}_2)$. By the condition that $\widehat{a}_l$ is bounded and so is $\phi(\widehat{a}_l(z_{t/2} + \zeta_l))\widehat{a}_l$. Then

$$\begin{aligned}\sum_{l=1}^{pq} |\widehat{\mathbf{f}}_l^T \widehat{\mathbf{W}} - \mathbf{f}_l^T \widetilde{\mathbf{W}}_2| &= \mathbf{1}^T |\widehat{\mathbf{F}} \widehat{\mathbf{W}} - \mathbf{F} \widetilde{\mathbf{W}}_2|\\ &= \mathbf{1}^T | \sum_{k=1}^{h} (\widehat{\boldsymbol{\rho}}_k \widehat{\boldsymbol{\rho}}_k^T - \boldsymbol{\rho}_k \boldsymbol{\rho}_k^T) vec(\mathbf{X})|\\ &\leq \sqrt{pq} \| \sum_{k=1}^{h} (\widehat{\boldsymbol{\rho}}_k \widehat{\boldsymbol{\rho}}_k^T - \boldsymbol{\rho}_k \boldsymbol{\rho}_k^T) \| \cdot \| vec(\mathbf{X}) \|.\end{aligned}$$

For the eigenvectors,

$$\|\sum_{k=1}^{h}(\widehat{\boldsymbol{\rho}}_k\widehat{\boldsymbol{\rho}}_k^T-\boldsymbol{\rho}_k\boldsymbol{\rho}_k^T)\| \leq \sum_{k=1}^{h}[\|\widehat{\boldsymbol{\rho}}_k(\widehat{\boldsymbol{\rho}}_k-\boldsymbol{\rho}_k)^T\|+\|(\widehat{\boldsymbol{\rho}}_k-\boldsymbol{\rho}_k)\widehat{\boldsymbol{\rho}}_k^T\|] \leq 2\sum_{k=1}^{h}\|\widehat{\boldsymbol{\rho}}_k-\boldsymbol{\rho}_k\|.$$

For $\|vec(\mathbf{X})\|$, note that $vec(\mathbf{X}) = \mathbf{T}^{1/2}vec(\widetilde{\mathbf{X}})$, where $\mathbf{T}^{1/2} = \text{diag}(\sqrt{T_i})$ and

$$T_i \sim InverseGamma(\frac{n+m-2}{2}, \frac{n+m-2}{2})$$

for $i = 1, \cdots, pq$ independent of $vec(\widetilde{\mathbf{X}})$. Therefore,

$$E\|vec(\mathbf{X})\|^2 = \sum_{i=1}^{pq} E(T_i(vec(\widetilde{\mathbf{X}}))_i^2) = \sum_{i=1}^{pq} ET_iE[vec(\widetilde{\mathbf{X}})]_i^2.$$

Since $ET_i = \frac{n+m-2}{n+m-4}$ and $E[vec(\widetilde{\mathbf{X}})]_i^2 = (vec(\boldsymbol{\mu}^\star))_i^2 + 1$. Therefore,

$$E\|vec(\mathbf{X})\|^2 = \frac{n+m-2}{n+m-4}(\|vec(\boldsymbol{\mu}^\star)\|^2 + pq).$$

This implies that $\|vec(\mathbf{X})\| = O_p(\|vec(\boldsymbol{\mu}^\star)\| + (pq)^{1/2})$.

Next we evaluate Δ_{12l}. By the mean value theorem, there exists $a_l^* \in (a_l, \widehat{a}_l)$ such that $\Delta_{12l} = \phi(a_l^*(z_{t/2} + \mathbf{f}_l^T\widetilde{\mathbf{W}}_2))(\widehat{a}_l - a_l)(z_{t/2} + \mathbf{f}_l^T\widetilde{\mathbf{W}}_2)$. Since both a_l and $\widehat{a}_l$ are greater than 1, we have $a_l^* > 1$ and hence $\phi(a_l^*(z_{t/2} + \mathbf{f}_l^T\widetilde{\mathbf{W}}))|z_{t/2} + \mathbf{f}_l^T\widetilde{\mathbf{W}}|$ is bounded. Therefore,

$$|\sum_{l=1}^{pq}\Delta_{12l}| \leq C_1\sum_{l=1}^{pq}|\widehat{a}_l - a_l|$$

for some positive constant C_1.

For the difference between $\widehat{a}_l$ and a_l, apply the mean value theorem, we have $|\widehat{a}_l - a_l| \leq C_2(\|\widehat{\mathbf{f}}_l\|^2 - \|\mathbf{f}_l\|^2)$. We define $\boldsymbol{\rho}_k = (\rho_{1k}, \cdots, \rho_{pq,k})^T$ and

$\widehat{\boldsymbol{\rho}}_k = (\widehat{\rho}_{1k}, \cdots, \widehat{\rho}_{pq,k})^T$, then

$$\begin{aligned}\sum_{l=1}^{pq}[\|\widehat{\mathbf{f}}_l\|^2 - \|\mathbf{f}_l\|^2] &= \sum_{l=1}^{pq}|\sum_{k=1}^{h}(\widehat{\theta}_k - \theta_k)\widehat{\rho}_{lk}^2 + \sum_{k=1}^{h}\theta_k(\widehat{\rho}_{lk}^2 - \rho_{lk}^2)| \\ &\leq \sum_{k=1}^{h}|\widehat{\theta}_k - \theta_k|\sum_{l=1}^{pq}\widehat{\rho}_{lk}^2 + \sum_{k=1}^{h}\theta_k\sum_{l=1}^{pq}|\widehat{\rho}_{lk}^2 - \rho_{lk}^2|.\end{aligned}$$

Note that $\sum_{l=1}^{pq}\widehat{\rho}_{lk}^2 = 1$ by the definition of eigenvectors.

Furthermore, we have

$$\begin{aligned}\sum_{l=1}^{pq}|\widehat{\rho}_{lk}^2 - \rho_{lk}^2| &\leq \{\sum_{l=1}^{pq}(\widehat{\rho}_{lk} - \rho_{lk})^2\sum_{l=1}^{pq}(\widehat{\rho}_{lk} + \rho_{lk})^2\}^{1/2} \\ &\leq \|\widehat{\boldsymbol{\rho}}_k - \boldsymbol{\rho}_k\|\{2\sum_{l=1}^{pq}(\widehat{\rho}_{lk}^2 + \rho_{lk}^2)\}^{1/2} \\ &= 2\|\widehat{\boldsymbol{\rho}}_k - \boldsymbol{\rho}_k\|.\end{aligned}$$

Therefore,

$$|\sum_{l=1}^{pq}\Delta_{12l}| \leq C_3(\sum_{k=1}^{h}\theta_k\|\widehat{\boldsymbol{\rho}}_k - \boldsymbol{\rho}_k\| + |\widehat{\theta}_k - \theta_k|).$$

To summarize,

$$|\Delta_1| \leq C_4(pq)^{1/2}\sum_{k=1}^{h}\|\widehat{\boldsymbol{\rho}}_k - \boldsymbol{\rho}_k\|O_p(\|vec(\boldsymbol{\mu}^\star)\| + (pq)^{1/2}) + C_5\sum_{k=1}^{h}[\theta_k\|\widehat{\boldsymbol{\rho}}_k - \boldsymbol{\rho}_k\| + |\widehat{\theta}_k - \theta_k|]. \tag{S1.3}$$

Next we evaluate $|\mathrm{FDP}_2(t) - \mathrm{FDP}_1(t)|$. Note that $|\mathrm{FDP}_2(t) - \mathrm{FDP}_1(t)| = |\sum_{l=1}^{pq}\Delta_{3l} + \Delta_{4l}|$ where

$$\begin{aligned}\Delta_{3l} &= \Phi(a_l(z_{t/2} + \mathbf{f}_l^T\widetilde{\mathbf{W}}_2)) - \Phi(a_l(z_{t/2} + \mathbf{f}_l^T\widetilde{\mathbf{W}}_1)) \\ \Delta_{4l} &= \Phi(a_l(z_{t/2} - \mathbf{f}_l^T\widetilde{\mathbf{W}}_2)) - \Phi(a_l(z_{t/2} - \mathbf{f}_l^T\widetilde{\mathbf{W}}_1)).\end{aligned}$$

To analyze Δ_{3l}, apply the mean value theorem, there exists ψ_l between $\mathbf{f}_l^T\widetilde{\mathbf{W}}_2$ and $\mathbf{f}_l^T\widetilde{\mathbf{W}}_1$ such that $\Delta_{3l} = \phi(a_l(z_{t/2} + \psi_l))a_l\mathbf{f}_l^T(\widetilde{\mathbf{W}}_2 - \widetilde{\mathbf{W}}_1)$. By the condition that a_l is bounded and so is $\phi(a_l(z_{t/2} + \psi_l))a_l$.

$$\begin{aligned}\sum_{l=1}^{pq}|\mathbf{f}_l^T(\widetilde{\mathbf{W}}_2 - \widetilde{\mathbf{W}}_1)| &= \mathbf{1}^T|\mathbf{F}(\widetilde{\mathbf{W}}_2 - \widetilde{\mathbf{W}}_1)| \\ &= \mathbf{1}^T|\sum_{k=1}^{h}\boldsymbol{\rho}_k\boldsymbol{\rho}_k^T(vec(\mathbf{X}) - vec(\widetilde{\mathbf{X}}))| \\ &\leq \sqrt{pq}\|\sum_{k=1}^{h}\boldsymbol{\rho}_k\boldsymbol{\rho}_k^T\| \cdot \|vec(\mathbf{X}) - vec(\widetilde{\mathbf{X}})\|.\end{aligned}$$

For the second term in the last line, $\|\sum_{k=1}^{h}\boldsymbol{\rho}_k\boldsymbol{\rho}_k^T\| = 1$. For the third term in the last line, note that $vec(\mathbf{X}) = \mathbf{V}^{1/2}vec(\widetilde{\mathbf{X}})$. Therefore,

$$E\|vec(\mathbf{X}) - vec(\widetilde{\mathbf{X}})\|^2 = \sum_{l=1}^{pq} E(\sqrt{T_i} - 1)^2(vec(\widetilde{\mathbf{X}}))_i^2 = \sum_{l=1}^{pq} E(\sqrt{T_i} - 1)^2 E(vec(\widetilde{\mathbf{X}}))_i^2.$$

Since $E(\sqrt{T_i} - 1)^2 = O((n + m - 2)^{-1})$, we have

$$\|vec(\mathbf{X}) - vec(\widetilde{\mathbf{X}})\| = O_p((n + m - 2)^{-1/2}(\|vec(\boldsymbol{\mu}^\star)\| + (pq)^{1/2})).$$

We can apply similar analysis to Δ_{4l}. Therefore, we have

$$|\mathrm{FDP}_2(t) - \mathrm{FDP}_1(t)| = \frac{1}{R(t)} O_p((n+m-2)^{-1/2}(pq)^{1/2}(\|vec(\boldsymbol{\mu}^\star)\| + (pq)^{1/2})).$$

Next we want to analyze $|\mathrm{FDP}_1(t) - \mathrm{FDP}_{A,1}(t)|$. With similar arguments as above, we can show that

$$|\mathrm{FDP}_1(t) - \mathrm{FDP}_{A,1}(t)| = O(|\mathbf{1}^T\mathbf{F}(\widetilde{\mathbf{W}}_1 - \mathbf{W})|/R(t)).$$

The infeasible least squares estimator $\widetilde{\mathbf{W}}_1$ is

$$\begin{aligned}\widetilde{\mathbf{W}}_1 = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T vec(\widetilde{\mathbf{X}}) &= (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T(vec(\boldsymbol{\mu}^\star) + \mathbf{F}\mathbf{W} + \boldsymbol{\epsilon}) \\ &= (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T vec(\boldsymbol{\mu}^\star) + \mathbf{W} + (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\boldsymbol{\epsilon}.\end{aligned}$$

For the third term, note that $E(\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\boldsymbol{\epsilon} = 0$ because the mean of $\boldsymbol{\epsilon}$ is 0. More interestingly,

$$var((\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\boldsymbol{\epsilon}) = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T var(\boldsymbol{\epsilon})\mathbf{F}(\mathbf{F}^T\mathbf{F})^{-1},$$

but the columns of $\mathbf{F}$ are orthogonal to $var(\boldsymbol{\epsilon})$. Therefore, $var((\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\boldsymbol{\epsilon}) = 0$. Consequentially, we have shown that $(\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\boldsymbol{\epsilon} = 0$.

Hence,

$$\begin{aligned}|\mathbf{1}^T\mathbf{F}(\widetilde{\mathbf{W}}_1 - \mathbf{W})| &= |\mathbf{1}^T(\sum_{k=1}^{h}\boldsymbol{\rho}_k\boldsymbol{\rho}_k^T)vec(\boldsymbol{\mu}^\star)| \\ &\leq (pq)^{1/2}\|\sum_{k=1}^{h}\boldsymbol{\rho}_k\boldsymbol{\rho}_k^T\|\|vec(\boldsymbol{\mu}^\star)\| = (pq)^{1/2}\|vec(\boldsymbol{\mu}^\star)\|.\end{aligned}$$

Therefore, we have shown that

$$|\mathrm{FDP}_1(t) - \mathrm{FDP}_{A,1}(t)| = O_p((pq)^{\zeta}((pq)^{-1/2}\|vec(\boldsymbol{\mu}^\star)\|)).$$

Next, we will show that

$$\|\widehat{\boldsymbol{\Sigma}}_2 \otimes \widehat{\boldsymbol{\Sigma}}_1 - \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1\| = O_p(pq(n+m)^{-1/2}).$$

By triangular inequality,

$$
\begin{aligned}
& \|\widehat{\mathbf{\Sigma}}_2 \otimes \widehat{\mathbf{\Sigma}}_1 - \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1\| \\
= \ & \|\widehat{\mathbf{\Sigma}}_2 \otimes \widehat{\mathbf{\Sigma}}_1 - \widehat{\mathbf{\Sigma}}_2 \otimes \mathbf{\Sigma}_1 + \widehat{\mathbf{\Sigma}}_2 \otimes \mathbf{\Sigma}_1 - \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1\| \\
\leq \ & \|\widehat{\mathbf{\Sigma}}_2 \otimes (\widehat{\mathbf{\Sigma}}_1 - \mathbf{\Sigma}_1)\| + \|(\widehat{\mathbf{\Sigma}}_2 - \mathbf{\Sigma}_2) \otimes \mathbf{\Sigma}_1\|.
\end{aligned}
$$

By the property of the Kronecker product, the operator norm of the product of two matrices $\mathbf{A}$ and $\mathbf{B}$ will be the product of the largest eigenvalue of $\mathbf{A}$ and the largest eigenvalue of $\mathbf{B}$. Therefore, the last line is

$$
\|\widehat{\mathbf{\Sigma}}_2\| \times \|\widehat{\mathbf{\Sigma}}_1 - \mathbf{\Sigma}_1\| + \|\widehat{\mathbf{\Sigma}}_2 - \mathbf{\Sigma}_2\| \times \|\mathbf{\Sigma}_1\|.
$$

We denote the (i,j)th element in $\widehat{\mathbf{\Sigma}}_1$ as $\widehat{\mathbf{\Sigma}}_{1,(i,j)}$, then

$$
\begin{aligned}
& \widehat{\mathbf{\Sigma}}_{1,(i,j)} \\
= \ & \frac{1}{q}\sum_{s=1}^{q}\frac{1}{n+m-2}\Big[\sum_{l=1}^{n}(\mathbf{Y}_{l,(i,s)} - \frac{1}{n}\sum_{l=1}^{n}\mathbf{Y}_{l,(i,s)})(\mathbf{Y}_{l,(j,s)} - \frac{1}{n}\sum_{l=1}^{n}\mathbf{Y}_{l,(j,s)})/(\widehat{\sigma}_{is}\widehat{\sigma}_{js}) \\
& \qquad + \sum_{k=1}^{m}(\mathbf{Z}_{k,(i,s)} - \frac{1}{m}\sum_{k=1}^{m}\mathbf{Z}_{k,(i,s)})(\mathbf{Z}_{k,(j,s)} - \frac{1}{m}\sum_{k=1}^{m}\mathbf{Z}_{k,(j,s)})/(\widehat{\sigma}_{is}\widehat{\sigma}_{js})\Big].
\end{aligned}
$$

Note that

$$\begin{aligned}
&\sum_{l=1}^{n}(\mathbf{Y}_{l,(i,s)}-\frac{1}{n}\sum_{l=1}^{n}\mathbf{Y}_{l,(i,s)})(\mathbf{Y}_{l,(j,s)}-\frac{1}{n}\sum_{l=1}^{n}\mathbf{Y}_{l,(j,s)})/(\widehat{\sigma}_{is}\widehat{\sigma}_{js})\\
= &\sum_{l=1}^{n}\big[\mathbf{Y}_{l,(i,s)}-\boldsymbol{\mu}_{is}-(\frac{1}{n}\sum_{l=1}^{n}\mathbf{Y}_{l,(i,s)}-\boldsymbol{\mu}_{is})\big]\\
&\qquad\times\big[\mathbf{Y}_{l,(j,s)}-\boldsymbol{\mu}_{js}-(\frac{1}{n}\sum_{l=1}^{n}\mathbf{Y}_{l,(j,s)}-\boldsymbol{\mu}_{js})\big]/(\widehat{\sigma}_{is}\widehat{\sigma}_{js})\\
= &\sum_{l=1}^{n}(\mathbf{Y}_{l,(i,s)}-\boldsymbol{\mu}_{is})(\mathbf{Y}_{l,(j,s)}-\boldsymbol{\mu}_{js})/(\widehat{\sigma}_{is}\widehat{\sigma}_{js})\\
&\qquad-\frac{1}{n}\sum_{l=1}^{n}(\mathbf{Y}_{l,(i,s)}-\boldsymbol{\mu}_{is})\sum_{l=1}^{n}(\mathbf{Y}_{l,(j,s)}-\boldsymbol{\mu}_{js})/(\widehat{\sigma}_{is}\widehat{\sigma}_{js}).
\end{aligned}$$

Note that

$$E[\sum_{l=1}^{n}(\mathbf{Y}_{l,(i,s)}-\boldsymbol{\mu}_{is})(\mathbf{Y}_{l,(j,s)}-\boldsymbol{\mu}_{js})/(\sigma_{is}\sigma_{js})]=n\boldsymbol{\Sigma}_{1,(i,j)},$$

and

$$\begin{aligned}
&\frac{1}{n}E[\sum_{l=1}^{n}(\mathbf{Y}_{l,(i,s)}-\boldsymbol{\mu}_{is})/\sigma_{is}\sum_{l=1}^{n}(\mathbf{Y}_{l,(j,s)}-\boldsymbol{\mu}_{js})/\sigma_{js}]\\
= &\frac{1}{n}\sum_{l=1}^{n}E[(\mathbf{Y}_{l,(i,s)}-\boldsymbol{\mu}_{is})(\mathbf{Y}_{l,(j,s)}-\boldsymbol{\mu}_{js})/(\sigma_{is}\sigma_{js})]=\boldsymbol{\Sigma}_{1,(i,j)}.
\end{aligned}$$

Therefore,

$$\frac{1}{n-1}E(\sum_{l=1}^{n}(\mathbf{Y}_{l,(i,s)}-\frac{1}{n}\sum_{l=1}^{n}\mathbf{Y}_{l,(i,s)})(\mathbf{Y}_{l,(j,s)}-\frac{1}{n}\sum_{l=1}^{n}\mathbf{Y}_{l,(j,s)})/(\sigma_{is}\sigma_{js}))=\boldsymbol{\Sigma}_{1,(i,j)}.$$

On the other hand,

$$Var(\frac{1}{n-1}\sum_{l=1}^{n}(\mathbf{Y}_{l,(i,s)}-\frac{1}{n}\sum_{l=1}^{n}\mathbf{Y}_{l,(i,s)})(\mathbf{Y}_{l,(j,s)}-\frac{1}{n}\sum_{l=1}^{n}\mathbf{Y}_{l,(j,s)})/(\sigma_{is}\sigma_{js}))=O(n^{-1}).$$

Therefore,

$$\begin{aligned}
& \frac{1}{n-1}\sum_{l=1}^{n}(\mathbf{Y}_{l,(i,s)} - \frac{1}{n}\sum_{l=1}^{n}\mathbf{Y}_{l,(i,s)})(\mathbf{Y}_{l,(j,s)} - \frac{1}{n}\sum_{l=1}^{n}\mathbf{Y}_{l,(j,s)})/(\sigma_{is}\sigma_{js}) \\
= \quad & \mathbf{\Sigma}_{1,(i,j)} + O_p(n^{-1/2}).
\end{aligned}$$

Similarly, we can show that

$$\begin{aligned}
& \frac{1}{m-1}\sum_{k=1}^{m}(\mathbf{Z}_{k,(i,s)} - \frac{1}{m}\sum_{k=1}^{m}\mathbf{Z}_{k,(i,s)})(\mathbf{Z}_{k,(j,s)} - \frac{1}{m}\sum_{k=1}^{m}\mathbf{Z}_{k,(j,s)})/(\sigma_{is}\sigma_{js}) \\
= \quad & \mathbf{\Sigma}_{1,(i,j)} + O_p(m^{-1/2}).
\end{aligned}$$

Note that $\widehat{\sigma}_{ij}^2 \sim (n+m-2)^{-1}\sigma_{ij}^2\chi_{n+m-2}^2$, Therefore, $\sigma_{ij}^2/\widehat{\sigma}_{ij}^2$ follows an inverse-chi-squared distribution with degrees of freedom as $n+m-2$ multiplied by a scalar $(n+m-2)$. Based on standard mean-variance analysis, we have

$$\frac{\sigma_{ij}^2}{\widehat{\sigma}_{ij}^2} = \frac{n+m-2}{n+m-4} + O_p(\frac{n+m-2}{n+m-4}(n+m-6)^{-1/2}).$$

Combining the two parts, for all i and j, we have

$$\begin{aligned}
& \widehat{\mathbf{\Sigma}}_{1,(i,j)} \\
= \quad & \frac{1}{q}\sum_{s=1}^{q}\frac{1}{n+m-2}\{(n-1)(\mathbf{\Sigma}_{1,(i,j)} + O_p(n^{-1/2}))(1 + O_p((n+m)^{-1/2}) \\
& \quad +(m-1)(\mathbf{\Sigma}_{1,(i,j)} + O_p(m^{-1/2}))(1 + O_p((n+m)^{-1/2})\} \\
= \quad & \mathbf{\Sigma}_{1,(i,j)} + O_p((n+m)^{-1/2}).
\end{aligned}$$

By the property of matrix norms,

$$\|\widehat{\mathbf{\Sigma}}_1 - \mathbf{\Sigma}_1\| \leq \|\widehat{\mathbf{\Sigma}}_1 - \mathbf{\Sigma}_1\|_1 = \max_{1\leq i\leq p} \sum_{j=1}^{p} |\widehat{\mathbf{\Sigma}}_{1,(ij)} - \mathbf{\Sigma}_{1,(ij)}| = O_p(p(n+m)^{-1/2}).$$

Similarly, we have

$$\|\widehat{\mathbf{\Sigma}}_2 - \mathbf{\Sigma}_2\| = O_p(q(n+m)^{-1/2}).$$

By triangular inequality, we have

$$\|\widehat{\mathbf{\Sigma}}_2\| \leq \|\mathbf{\Sigma}_2\| + O_p(q(n+m)^{-1/2}) \leq q + O_p(q(n+m)^{-1/2}).$$

Therefore,

$$\|\widehat{\mathbf{\Sigma}}_2 \otimes \widehat{\mathbf{\Sigma}}_1 - \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1\| = O_p(pq(n+m)^{-1/2}).$$

By the triangular inequality, $|\theta_k - \widehat{\theta}_{k+1}| \geq ||\theta_k - \theta_{k+1}| - |\theta_{k+1} - \widehat{\theta}_{k+1}||$. By Weyl's Theorem in Lemma 1, $|\theta_{k+1} - \widehat{\theta}_{k+1}| \leq \|\widehat{\mathbf{\Sigma}}_2 \otimes \widehat{\mathbf{\Sigma}}_1 - \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1\|$. Therefore, on the event $\{\|\widehat{\mathbf{\Sigma}}_2 \otimes \widehat{\mathbf{\Sigma}}_1 - \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1\| = O(pq(n+m)^{-1/2}\}$, $|\theta_k - \widehat{\theta}_{k+1}\| \geq d_{pq} - \|\widehat{\mathbf{\Sigma}}_2 \otimes \widehat{\mathbf{\Sigma}}_1 - \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1\| \geq d_{pq}/2$ for sufficiently large pq. Similarly, we have $|\widehat{\theta}_{k-1} - \theta_k| \geq d_{pq}/2$.

By $sin(\theta)$ Theorem in Lemma 1, we have

$$\|\widehat{\boldsymbol{\rho}}_k - \boldsymbol{\rho}_k\| = O_p((n+m)^{-1/2})$$

for $k = 1, \cdots, h$.

By Weyl's Theorem in Lemma 1,

$$\sum_{k=1}^{h} |\widehat{\theta}_k - \theta_k| = O_p(hpq(n+m)^{-1/2}).$$

Back to Δ_1 in expression (S1.3), we have

$$|\Delta_1| = O_p(hpq(n+m)^{-1/2}) + O_p(h(n+m)^{-1/2}(pq)^{1/2}\|vec(\boldsymbol{\mu}^\star)\|).$$

Under the assumption that $R(t)^{-1} = O_p((pq)^{-1+\zeta})$, we can show that

$$|\widehat{\mathrm{FDP}}_1(t) - \mathrm{FDP}_{A,1}(t)| = O_p((pq)^{\zeta}(h(n+m)^{-1/2} + (pq)^{-1/2}(\|\boldsymbol{\mu}^\star\|))).$$

The proof is now complete.

Proof of Proposition 2:

By the proof of Theorem 2 (i) in Fan and Han (2017), if we can show that

$$(pq)^{-1}\|cov(vec(\boldsymbol{\epsilon}))\|_F = O((pq)^{-\delta})$$

for some $\delta > 0$, then the conclusion is correct. Recall that

$$cov(vec(\boldsymbol{\epsilon})) = (\sum_{j=k_2+1}^{q} \xi_j \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^T) \otimes (\sum_{i=k_1+1}^{p} \lambda_i \boldsymbol{\nu}_i \boldsymbol{\nu}_i^T).$$

Note that the eigenvalues of the Kronecker product are the product of the eigenvalues from the original two matrices, and the eigenvectors of the Kronecker product are the Kronecker product of the eigenvectors from the original two matrices. Therefore, we have

$$(\sum_{j=k_2+1}^{q} \xi_j \boldsymbol{\gamma}_j \boldsymbol{\gamma}_j^T) \otimes (\sum_{i=k_1+1}^{p} \lambda_i \boldsymbol{\nu}_i \boldsymbol{\nu}_i^T) = \sum_{j=k_2+1}^{q} \sum_{i=k_1+1}^{p} (\lambda_i \xi_j)(\boldsymbol{\nu}_i \otimes \boldsymbol{\gamma}_j)(\boldsymbol{\nu}_i \otimes \boldsymbol{\gamma}_j)^T.$$

The Frobenius norm is invariant under orthonormal transformation. Therefore,

$$\begin{aligned} & \| \sum_{j=k_2+1}^{q} \sum_{i=k_1+1}^{p} (\lambda_i \xi_j)(\boldsymbol{\nu}_i \otimes \boldsymbol{\gamma}_j)(\boldsymbol{\nu}_i \otimes \boldsymbol{\gamma}_j)^T \|_F^2 \\ = & \sum_{j=k_2+1}^{q} \sum_{i=k_1+1}^{p} (\lambda_i \xi_j)^2 = (\sum_{i=k_1+1}^{p} \lambda_i^2)(\sum_{j=k_2+1}^{q} \xi_j^2). \end{aligned}$$

By the conditions in Proposition 2,

$$(pq)^{-1} \| cov(vec(\boldsymbol{\epsilon}) \|_F = (pq)^{-1} (\sum_{i=k_1+1}^{p} \lambda_i^2)^{1/2} (\sum_{j=k_2+1}^{q} \xi_j^2)^{1/2} = O(p^{-\delta_1}) O(q^{-\delta_2}).$$

The conclusion is correct. Following the proof of Proposition 1, we can show the conclusion.

Proof of Theorem 2:

We consider a least squares estimator for $vec(\mathbf{W})$:

$$\begin{aligned} \widehat{vec(\mathbf{W})} &= [(\mathbf{D}^T \otimes \mathbf{C})^T (\mathbf{D}^T \otimes \mathbf{C})]^{-1} (\mathbf{D}^T \otimes \mathbf{C})^T vec(\mathbf{X}) \\ &= [(\mathbf{D}\mathbf{D}^T) \otimes (\mathbf{C}^T \mathbf{C})]^{-1} (\mathbf{D}^T \otimes \mathbf{C})^T vec(\mathbf{X}) \\ &= [diag(\xi_1^{-1}, \cdots, \xi_{k_2}^{-1}) \otimes diag(\lambda_1^{-1}, \cdots, \lambda_{k_1}^{-1})] (\mathbf{D}^T \otimes \mathbf{C})^T vec(\mathbf{X}). \end{aligned}$$

Correspondingly,

$$\begin{aligned} (\mathbf{D}^T \otimes \mathbf{C}) \widehat{vec(\mathbf{W})} &= [(\xi_1^{-1/2} \boldsymbol{\gamma}_1, \cdots, \xi_{k_2}^{-1/2} \boldsymbol{\gamma}_{k_2}) \otimes (\lambda_1^{-1/2} \boldsymbol{\nu}_1, \cdots, \lambda_{k_1}^{-1/2} \boldsymbol{\nu}_{k_1})] \\ & \quad \times [(\xi_1^{1/2} \boldsymbol{\gamma}_1, \cdots, \xi_{k_2}^{1/2} \boldsymbol{\gamma}_{k_2})^T \otimes (\lambda_1^{1/2} \boldsymbol{\nu}_1, \cdots, \lambda_{k_1}^{1/2} \boldsymbol{\nu}_{k_1})^T] vec(\mathbf{X}) \\ &= [(\sum_{i=1}^{k_2} \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^T) \otimes (\sum_{j=1}^{k_1} \boldsymbol{\nu}_j \boldsymbol{\nu}_j^T)] vec(\mathbf{X}). \end{aligned}$$

In $\mathbf{D}^T \otimes \mathbf{C}$, each column is $\sqrt{\xi_l \lambda_k}\boldsymbol{\gamma}_l \otimes \boldsymbol{\nu}_k$, for $l = 1, \cdots, k_2$ and $k = 1, \cdots, k_1$. Similar to the proof of Theorem 1, we can show that there exist positive constants C_1, C_2, C_3, C_4 such that

$$\begin{aligned} & |\widehat{\mathrm{FDP}}_2(t) - \mathrm{FDP}_{A,2}(t)| \\ \leq \; & \frac{1}{R(t)}\Big[C_1 \sum_{l=1}^{pq} |\widehat{\mathbf{b}}_l^T \widehat{\mathbf{W}} - \mathbf{b}_l^T \widetilde{\mathbf{W}}_2| + C_2 \sum_{l=1}^{pq} (\|\widehat{\mathbf{b}}_l\|^2 - \|\mathbf{b}_l\|^2) \\ & + C_3 \sum_{l=1}^{pq} |\mathbf{b}_l^T (\widetilde{\mathbf{W}}_2 - \widetilde{\mathbf{W}}_1)| + C_4 |\mathbf{1}^T (\mathbf{D}^T \otimes \mathbf{C})(\widetilde{\mathbf{W}}_1 - \mathbf{W})|\Big]. \end{aligned} \tag{S1.4}$$

We will analyze each term in the last line. For the second term in (S1.4), we have

$$\begin{aligned} & \sum_{l=1}^{pq} [\|\widehat{\mathbf{b}}_l\|^2 - \|\mathbf{b}_l\|^2] \\ = \; & \sum_{j=1}^{q} \sum_{i=1}^{p} |\sum_{k=1}^{k_1} \sum_{l=1}^{k_2} (\sqrt{\widehat{\xi}_l \widehat{\lambda}_k} \widehat{\gamma}_{li} \widehat{\boldsymbol{\nu}}_{kj})^2 - \sum_{k=1}^{k_1} \sum_{l=1}^{k_2} (\sqrt{\xi_l \lambda_k} \gamma_{li} \boldsymbol{\nu}_{kj})^2| \\ = \; & \sum_{j=1}^{q} \sum_{i=1}^{p} |\sum_{k=1}^{k_1} \sum_{l=1}^{k_2} (\widehat{\xi}_l \widehat{\lambda}_k - \xi_l \lambda_k) \widehat{\boldsymbol{\gamma}}_{li}^2 \widehat{\boldsymbol{\nu}}_{kj}^2 \\ & \quad + \sum_{k=1}^{k_1} \sum_{l=1}^{k_2} \xi_l \lambda_k (\widehat{\boldsymbol{\gamma}}_{li}^2 \widehat{\boldsymbol{\nu}}_{kj}^2 - \boldsymbol{\gamma}_{li}^2 \boldsymbol{\nu}_{kj}^2)| \\ \leq \; & \sum_{k=1}^{k_1} \sum_{l=1}^{k_2} |\widehat{\xi}_l \widehat{\lambda}_k - \xi_l \lambda_k| (\sum_{j=1}^{q} \sum_{i=1}^{p} \widehat{\boldsymbol{\gamma}}_{li}^2 \widehat{\boldsymbol{\nu}}_{kj}^2) \\ & \quad + \sum_{k=1}^{k_1} \sum_{l=1}^{k_2} \xi_l \lambda_k \sum_{j=1}^{q} \sum_{i=1}^{p} |\widehat{\boldsymbol{\gamma}}_{li}^2 \widehat{\boldsymbol{\nu}}_{kj}^2 - \boldsymbol{\gamma}_{li}^2 \boldsymbol{\nu}_{kj}^2|. \end{aligned} \tag{S1.5}$$

For the second term in (S1.5),

$$\begin{aligned}
& \sum_{j=1}^{q}\sum_{i=1}^{p}|\widehat{\gamma}_{li}^2\widehat{\boldsymbol{\nu}}_{kj}^2-\boldsymbol{\gamma}_{li}^2\boldsymbol{\nu}_{kj}^2| \\
= \; & \sum_{j=1}^{q}\sum_{i=1}^{p}|\widehat{\gamma}_{li}\widehat{\boldsymbol{\nu}}_{kj}+\boldsymbol{\gamma}_{li}\boldsymbol{\nu}_{kj}|\times|\widehat{\gamma}_{li}\widehat{\boldsymbol{\nu}}_{kj}-\boldsymbol{\gamma}_{li}\boldsymbol{\nu}_{kj}| \\
\leq \; & (\sum_{j=1}^{q}\sum_{i=1}^{p}|\widehat{\gamma}_{li}\widehat{\boldsymbol{\nu}}_{kj}+\boldsymbol{\gamma}_{li}\boldsymbol{\nu}_{kj}|^2)^{1/2}(\sum_{j=1}^{q}\sum_{i=1}^{p}(\widehat{\gamma}_{li}\widehat{\boldsymbol{\nu}}_{kj}-\boldsymbol{\gamma}_{li}\boldsymbol{\nu}_{kj})^2)^{1/2} \\
\leq \; & [2\sum_{j=1}^{q}\sum_{i=1}^{p}(\widehat{\gamma}_{li}^2\widehat{\boldsymbol{\nu}}_{kj}^2+\boldsymbol{\gamma}_{li}^2\boldsymbol{\nu}_{kj}^2)]^{1/2}\times\|\widehat{\boldsymbol{\gamma}}_l\otimes\widehat{\boldsymbol{\nu}}_k-\boldsymbol{\gamma}_l\otimes\boldsymbol{\nu}_k\| \\
= \; & 2\|\widehat{\boldsymbol{\gamma}}_l\otimes\widehat{\boldsymbol{\nu}}_k-\boldsymbol{\gamma}_l\otimes\boldsymbol{\nu}_k\|.
\end{aligned}$$

Therefore, for some positive constant,

$$\sum_{l=1}^{pq}\left[\|\widehat{\mathbf{b}}_l^2-\|\mathbf{b}_l\|^2\right]\leq C(\sum_{k=1}^{k_1}\sum_{l=1}^{k_2}|\widehat{\xi_l}\widehat{\lambda}_k-\xi_l\lambda_k|+\xi_l\lambda_k\|\widehat{\boldsymbol{\gamma}}_l\otimes\widehat{\boldsymbol{\nu}}_k-\boldsymbol{\gamma}_l\otimes\boldsymbol{\nu}_k\|).$$

For the first term in (S1.5),

$$\begin{aligned}
& \sum_{k=1}^{k_1}\sum_{l=1}^{k_2}|\widehat{\xi_l}\widehat{\lambda}_k-\xi_l\lambda_k| \\
= \; & \sum_{k=1}^{k_1}\sum_{l=1}^{k_2}|\widehat{\xi_l}\widehat{\lambda}_k-\xi_l\widehat{\lambda}_k+\xi_l\widehat{\lambda}_k-\xi_l\lambda_k| \\
\leq \; & \sum_{k=1}^{k_1}\sum_{l=1}^{k_2}(|\widehat{\xi_l}-\xi_l|\widehat{\lambda}_k+\xi_l|\widehat{\lambda}_k-\lambda_k|) \\
= \; & \sum_{l=1}^{k_2}|\widehat{\xi_l}-\xi_l|\sum_{k=1}^{k_1}\widehat{\lambda}_k+\sum_{l=1}^{k_2}\xi_l\sum_{k=1}^{k_1}|\widehat{\lambda}_k-\lambda_k|.
\end{aligned}$$

In the proof of Theorem 1, we have shown $\|\widehat{\boldsymbol{\Sigma}}_1-\boldsymbol{\Sigma}_1\|=O_p(p(n+m)^{-1/2})$ and $\|\widehat{\boldsymbol{\Sigma}}_2-\boldsymbol{\Sigma}_2\|=O_p(q(n+m)^{-1/2})$. Note that in the last line, $\sum_{l=1}^{k_2}|\widehat{\xi_l}-\xi_l|=$

$O_p(k_2 q(n+m)^{-1/2})$, and $\sum_{k=1}^{k_1} |\widehat{\lambda}_k - \lambda_k| = O_p(k_1 p(n+m)^{-1/2})$. Furthermore, we have $\sum_{l=1}^{k_2} \xi_l \le q$ and

$$\sum_{k=1}^{k_1} \widehat{\lambda}_k = \sum_{k=1}^{k_1} (\widehat{\lambda}_k - \lambda_k + \lambda_k) = O_p(k_1 p(n+m)^{-1/2}) + p.$$

Next, for the first term in (S1.5), we can show that

$$\begin{aligned}
&\sum_{l=1}^{pq} |\widehat{\mathbf{b}}_l^T \widehat{\mathbf{W}} - \mathbf{b}_l^T \widetilde{\mathbf{W}}_2| \\
\le\ & \sqrt{pq} \|(\sum_{i=1}^{k_2} \widehat{\boldsymbol{\gamma}}_i \widehat{\boldsymbol{\gamma}}_i^T) \otimes (\sum_{j=1}^{k_1} \widehat{\boldsymbol{\nu}}_j \widehat{\boldsymbol{\nu}}_j^T) - (\sum_{i=1}^{k_1} \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^T) \otimes (\sum_{j=1}^{k_1} \boldsymbol{\nu}_j \boldsymbol{\nu}_j^T)\| \|vec(\mathbf{X})\|.
\end{aligned}$$

In the last line, by the triangular inequality and the property of Kronecker product on the eigenvalues, we have

$$\begin{aligned}
&\|(\sum_{i=1}^{k_2} \widehat{\boldsymbol{\gamma}}_i \widehat{\boldsymbol{\gamma}}_i^T) \otimes (\sum_{j=1}^{k_1} \widehat{\boldsymbol{\nu}}_j \widehat{\boldsymbol{\nu}}_j^T) - (\sum_{i=1}^{k_1} \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^T) \otimes (\sum_{j=1}^{k_1} \boldsymbol{\nu}_j \boldsymbol{\nu}_j^T)\| \\
\le\ & \|(\sum_{i=1}^{k_2} \widehat{\boldsymbol{\gamma}}_i \widehat{\boldsymbol{\gamma}}_i^T) \otimes (\sum_{j=1}^{k_1} \widehat{\boldsymbol{\nu}}_j \widehat{\boldsymbol{\nu}}_j^T - \sum_{j=1}^{k_1} \boldsymbol{\nu}_j \boldsymbol{\nu}_j^T)\| + \|(\sum_{i=1}^{k_2} \widehat{\boldsymbol{\gamma}}_i \widehat{\boldsymbol{\gamma}}_i^T - \sum_{i=1}^{k_2} \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^T) \otimes (\sum_{j=1}^{k_1} \boldsymbol{\nu}_j \boldsymbol{\nu}_j^T)\| \\
=\ & \|(\sum_{i=1}^{k_2} \widehat{\boldsymbol{\gamma}}_i \widehat{\boldsymbol{\gamma}}_i^T)\| \times \|\sum_{j=1}^{k_1} \widehat{\boldsymbol{\nu}}_j \widehat{\boldsymbol{\nu}}_j^T - \sum_{j=1}^{k_1} \boldsymbol{\nu}_j \boldsymbol{\nu}_j^T\| + \|\sum_{i=1}^{k_2} \widehat{\boldsymbol{\gamma}}_i \widehat{\boldsymbol{\gamma}}_i^T - \sum_{i=1}^{k_2} \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^T\| \times \|\sum_{j=1}^{k_1} \boldsymbol{\nu}_j \boldsymbol{\nu}_j^T\| \\
=\ & \|\sum_{j=1}^{k_1} \widehat{\boldsymbol{\nu}}_j \widehat{\boldsymbol{\nu}}_j^T - \sum_{j=1}^{k_1} \boldsymbol{\nu}_j \boldsymbol{\nu}_j^T\| + \|\sum_{i=1}^{k_2} \widehat{\boldsymbol{\gamma}}_i \widehat{\boldsymbol{\gamma}}_i^T - \sum_{i=1}^{k_2} \boldsymbol{\gamma}_i \boldsymbol{\gamma}_i^T\| \\
\le\ & 2\sum_{j=1}^{k_1} \|\widehat{\boldsymbol{\nu}}_j - \boldsymbol{\nu}_j\| + 2\sum_{i=1}^{k_2} \|\widehat{\boldsymbol{\gamma}}_i - \boldsymbol{\gamma}_i\|.
\end{aligned}$$

By $\sin(\theta)$ Theorem, we can show that the last line is $O_p(k_1(n+m)^{-1/2}) + O_p(k_2(n+m)^{-1/2})$.

For the third term in (S1.4), we have

$$\begin{aligned}\sum_{l=1}^{pq}|\mathbf{b}_l^T(\widetilde{\mathbf{W}}_2-\widetilde{\mathbf{W}}_1)| &\le \mathbf{1}^T|(\sum_{i=1}^{k_2}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T)\otimes(\sum_{j=1}^{k_1}\boldsymbol{\nu}_j\boldsymbol{\nu}_j^T)(vec(\mathbf{X})-vec(\widetilde{\mathbf{X}}))| \\ &\le \sqrt{pq}\|(\sum_{i=1}^{k_2}\boldsymbol{\gamma}_i\boldsymbol{\gamma}_i^T)\otimes(\sum_{j=1}^{k_1}\boldsymbol{\nu}_j\boldsymbol{\nu}_j^T)\|\times\|vec(\mathbf{X})-vec(\widetilde{\mathbf{X}})\| \\ &= \sqrt{pq}\|vec(\mathbf{X})-vec(\widetilde{\mathbf{X}})\|.\end{aligned}$$

In the proof of Theorem 1, we have shown that $\|vec(\mathbf{X})-vec(\widetilde{\mathbf{X}})\| = O_p((n+m-2)^{-1/2}(\|vec(\boldsymbol{\mu}^\star)\|+(pq)^{1/2}))$.

For the last term in (S1.4), similar to the proof of Theorem 1, we can show that $|\mathbf{1}^T(\mathbf{D}^T\otimes\mathbf{C})(\widetilde{\mathbf{W}}_1-\mathbf{W}))| = (pq)^{1/2}\|vec(\boldsymbol{\mu}^\star)\|$. Combining the four terms in (S1.4), we can show that

$$\begin{aligned}&|\widehat{\mathrm{FDP}}_2(t)-\mathrm{FDP}_{A,2}(t)| \\ =\ & O_p((pq)^{\zeta}(k_1k_2(n+m)^{-1}+(k_1+k_2)(n+m)^{-1/2}+(pq)^{-1/2}\|vec(\boldsymbol{\mu}^\star)\|).\end{aligned}$$

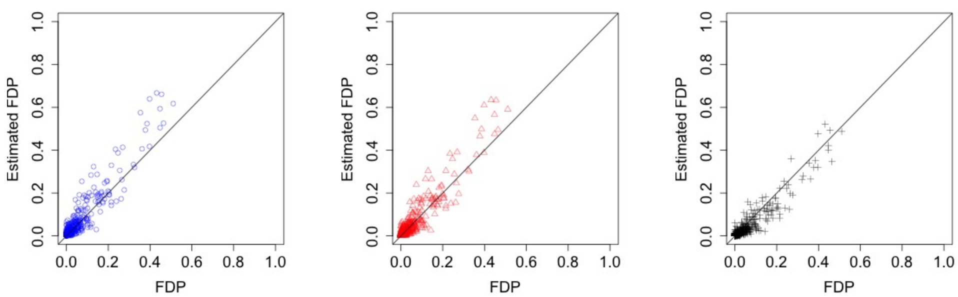
The proof is now complete.

S2 Numerial Results

In this section, we provide additional figures for comparing estimated FDP with the true value of FDP under Models 1-3 with more settings. Also, we include the comparison of proposed methods using MLE-type covariance estimators, we use the R package called *MixMatrix* to realize the MLE algorithm described in Dutilleuel (1999). The simulation results are shown in Figure S14, Table S2, and Table S3. For model 1 (strict factor model), both proposed methods using the MLE estimator perform slightly better than using the (pooled-)sample correlation estimator. However, the situation has now reversed regarding Model 2 (approximate factor model). When the matrix normality assumption is violated (Model 3), those two estimators play roughly the same for the two methods. The results hold the same for $p = q = 500$, except that the sandwich method using the sample correlation estimator performs better than the MLE estimator for Model 3. As described in Dutilleuel (1999), it's worth noting that for the MLE algorithm of matrix normal distribution, different initial solutions of the MLE algorithm will provide different final solutions $\widehat{\mathbf{U}}$ and $\widehat{\mathbf{V}}$, but the same direct product $\widehat{\mathbf{V}} \otimes \widehat{\mathbf{U}}$. Thus the noodle method tends to perform stably when adopting the MLE-type estimators.

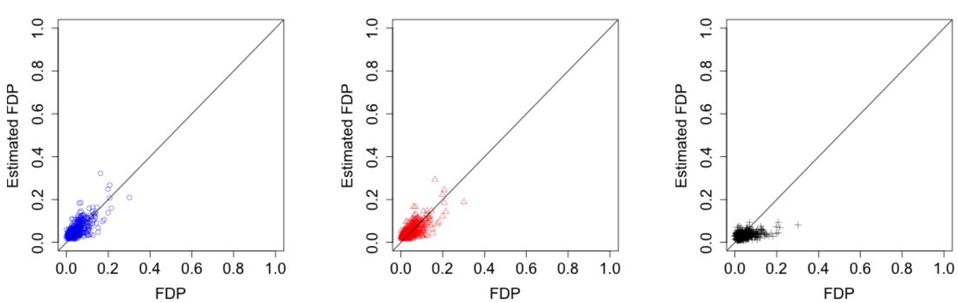
S3 Data Analysis

For the current paper, we simplify the model setting via assuming $\mathbf{U}, \mathbf{V}$ are the same across samples to avoid some technical challenges. If we consider $\mathbf{U}, \mathbf{V}$ are not the same across the two samples, let $\mathbf{Y} \sim \mathcal{MN}(\boldsymbol{\mu}_y, \mathbf{U}_1, \mathbf{V}_1)$ and $\mathbf{Z} \sim \mathcal{MN}(\boldsymbol{\mu}_z, \mathbf{U}_2, \mathbf{V}_2)$, to apply noodle method, we could test the



(a) $f(3,3), \mathbf{B} \sim N(0,1)$ (b) $f(3,3), \mathbf{B} \sim N(0,1)$ (c) $f(3,3), \mathbf{B} \sim N(0,1)$

Figure S1: Model 1, the estimated values of FDP obtained by noodle method (blue circle), sandwich method (red triangle) and PFA (black crossover) are compared with the true value of FDP. Here, $n = m = 50$, $p = q = 100$, and $t = 0.001$.

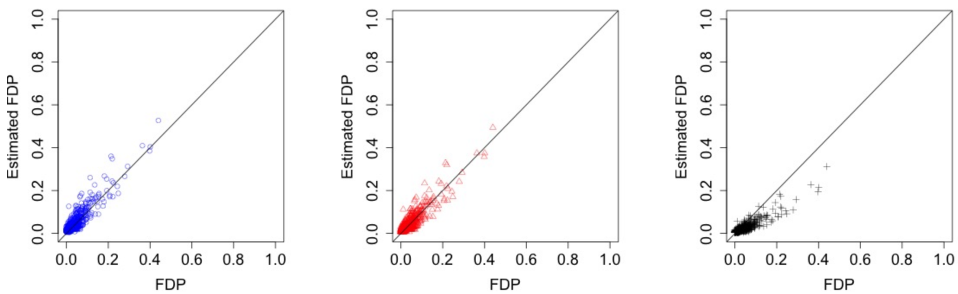


(a) $(\rho_1, \rho_2) = (0.5, 0.8)$ (b) $(\rho_1, \rho_2) = (0.5, 0.8)$ (c) $(\rho_1, \rho_2) = (0.5, 0.8)$

Figure S2: Model 2, the estimated values of FDP obtained by noodle method (blue circle), sandwich method (red triangle) and PFA (black crossover) are compared with the true value of FDP. Here, $n = m = 50$, $p = q = 100$, and $t = 0.001$.

following hypothesis:

$$H_0 : \mathbf{R}_{V_1} \otimes \mathbf{R}_{U_1} = \mathbf{R}_{V_2} \otimes \mathbf{R}_{U_2} \quad \text{versus} \quad H_a : \mathbf{R}_{V_1} \otimes \mathbf{R}_{U_1} \neq \mathbf{R}_{V_2} \otimes \mathbf{R}_{U_2},$$



(a) $f(3,3), \mathbf{W} \sim \sqrt{\frac{2}{3}} t_6$ (b) $f(3,3), \mathbf{W} \sim \sqrt{\frac{2}{3}} t_6$ (c) $f(3,3), \mathbf{W} \sim \sqrt{\frac{2}{3}} t_6$

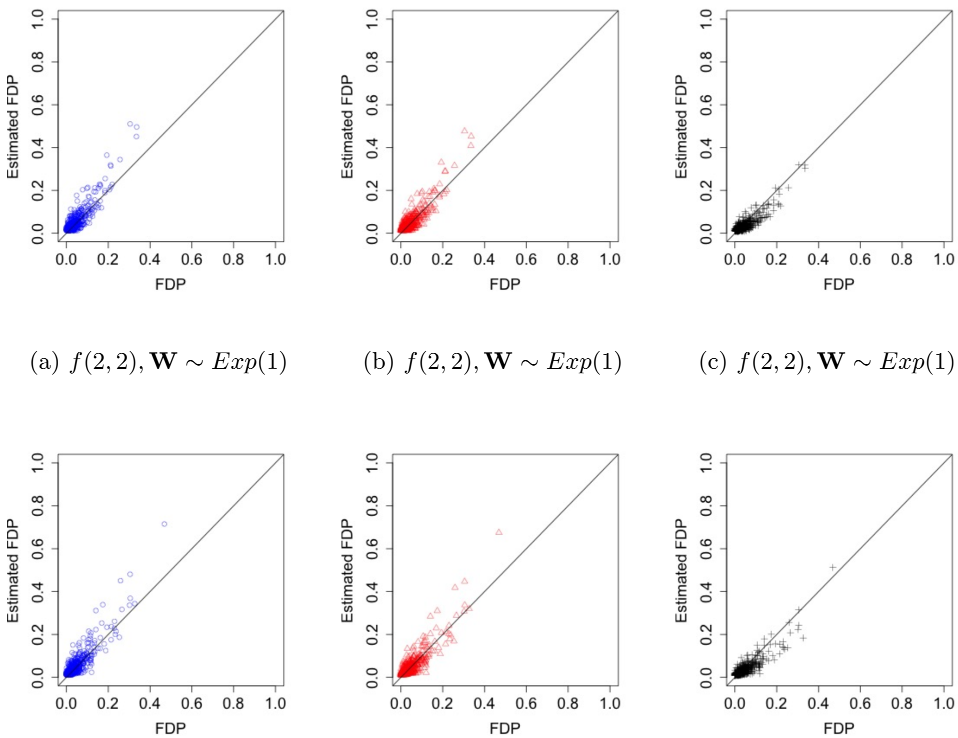
Figure S3: Model 3: factor (3,3) case, the estimated values of FDP obtained by noodle method (blue circle), sandwich method (red triangle) and PFA (black crossover) are compared with the true value of FDP. Each element in $\mathbf{W}$ follows $\sqrt{\frac{2}{3}} t_6$ distribution. Here, $n = m = 50$, $p = q = 100$, and $t = 0.001$.

where $\mathbf{R}_\Sigma$ is the correlation matrix of $\mathbf{\Sigma}$.

To test the above hypothesis for the real data analysis, we adopt the vector-based testing procedure proposed by Cai, Liu and Xia (2013) after standardization. Given the significance level $\alpha = 0.05$, they define

$$\Phi_\alpha = I(M_n \geq q_\alpha + 4 \log d - \log \log d)$$

where M_n is the test statistics they introduced to measure the discrepancy of two covariance matrices, $q_\alpha = -\log(8\pi) - 2 \log \log(1-\alpha)^{-1}$. If $\Phi_\alpha = 1$, the null hypothesis H_0 is rejected. For our real data, the null hypothesis $H_0 : \mathbf{R}_{V_1} \otimes \mathbf{R}_{U_1} = \mathbf{R}_{V_2} \otimes \mathbf{R}_{U_2}$ is not rejected with $M_n = 35.055 < q_{0.05} + 4 \log pq - \log \log pq = 39.260$. The acceptance of the above hypothesis is the key condition under which our proposed methods work.



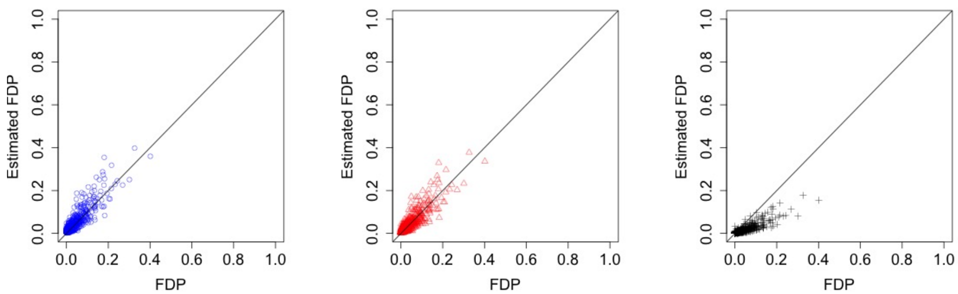
(a) $f(2,2), \mathbf{W} \sim Exp(1)$ (b) $f(2,2), \mathbf{W} \sim Exp(1)$ (c) $f(2,2), \mathbf{W} \sim Exp(1)$

(d) $f(2,2), \mathbf{W} \sim \sqrt{\frac{2}{3}}t_6$ (e) $f(2,2), \mathbf{W} \sim \sqrt{\frac{2}{3}}t_6$ (f) $f(2,2), \mathbf{W} \sim \sqrt{\frac{2}{3}}t_6$

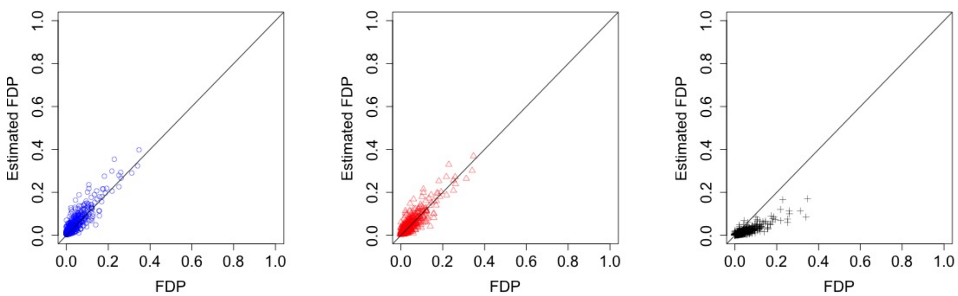
Figure S4: Model 3: factor (2,2) case, the estimated values of FDP obtained by noodle method (blue circle), sandwich method (red triangle) and PFA (black crossover) are compared with the true value of FDP. The first row represents the case where $\mathbf{W}$ with each element following Exp(1) distribution, and the second row shows the case where $\mathbf{W}$ with each element following $\sqrt{\frac{2}{3}}t_6$ distribution. Here, $n = m = 50$, $p = q = 100$, and $t = 0.001$.

References

Cai, T., Liu, W. and Xia, Y. (2013). Two-sample covariance matrix testing and support recovery in high-dimensional and sparse settings. *Journal of the American Statistical Association*



(a) $f(4,4), \mathbf{W} \sim Exp(1)$ (b) $f(4,4), \mathbf{W} \sim Exp(1)$ (c) $f(4,4), \mathbf{W} \sim Exp(1)$



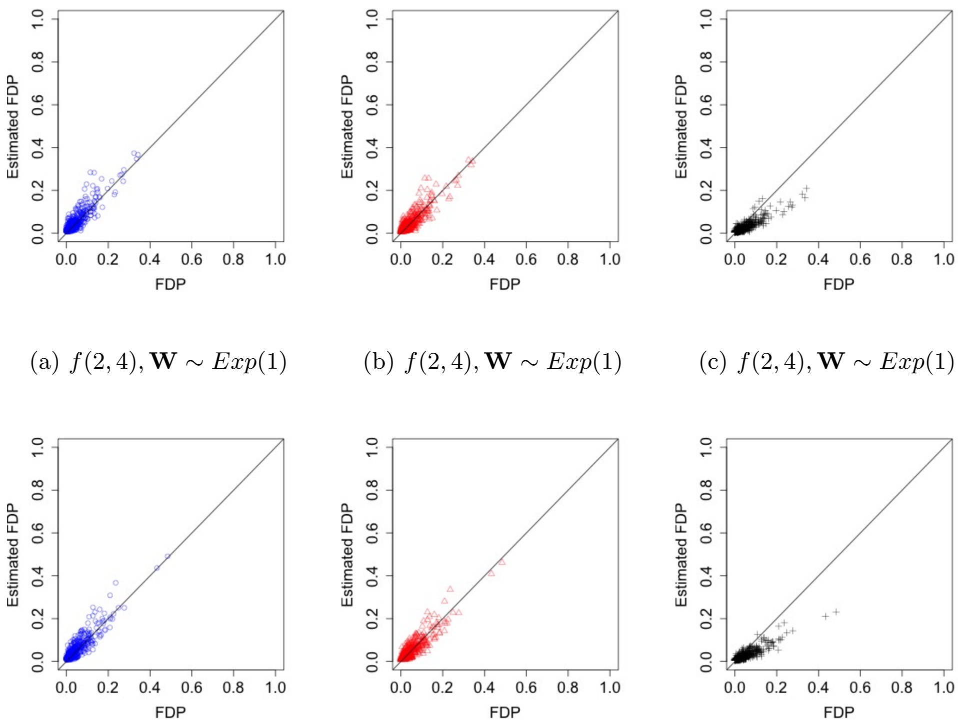
(d) $f(4,4), \mathbf{W} \sim \sqrt{\frac{2}{3}}t_6$ (e) $f(4,4), \mathbf{W} \sim \sqrt{\frac{2}{3}}t_6$ (f) $f(4,4), \mathbf{W} \sim \sqrt{\frac{2}{3}}t_6$

Figure S5: Model 3: factor (4,4) case, the estimated values of FDP obtained by noodle method (blue circle), sandwich method (red triangle) and PFA (black crossover) are compared with the true value of FDP. The first row represents the case where $\mathbf{W}$ with each element following Exp(1) distribution, and the second row shows the case where $\mathbf{W}$ with each element following $\sqrt{\frac{2}{3}}t_6$ distribution. Here, $n = m = 50$, $p = q = 100$, and $t = 0.001$.

108, 265-277.

Davis, C. and Kahan, W. (1970). The rotation of eigenvectors by a perturbation III. *SIAM Journal on Numerical Analysis* **7**, 1-46.

Dutilleuel, P. (1999). The MLE algorithm for the matrix normal distribution. *Journal of Sta-*



(a) $f(2,4), \mathbf{W} \sim Exp(1)$ (b) $f(2,4), \mathbf{W} \sim Exp(1)$ (c) $f(2,4), \mathbf{W} \sim Exp(1)$

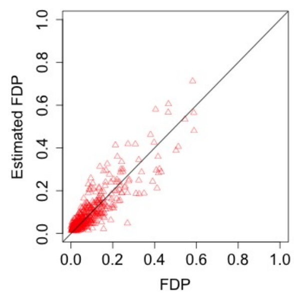
(d) $f(2,4), \mathbf{W} \sim \sqrt{\frac{2}{3}}t_6$ (e) $f(2,4), \mathbf{W} \sim \sqrt{\frac{2}{3}}t_6$ (f) $f(2,4), \mathbf{W} \sim \sqrt{\frac{2}{3}}t_6$

Figure S6: Model 3: factor (2,4) case, the estimated values of FDP obtained by the noodle method (blue circle), sandwich method (red triangle), and PFA (black crossover) are compared with the true value of FDP. The first row represents the case where $\mathbf{W}$ with each element following Exp(1) distribution, and the second row shows the case where $\mathbf{W}$ with each element following $\sqrt{\frac{2}{3}}t_6$ distribution. Here, $n = m = 50$, $p = q = 100$, and $t = 0.001$.

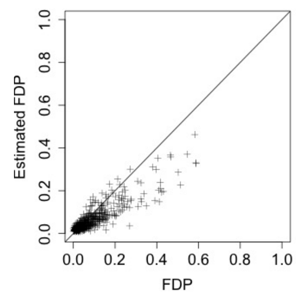
tistical Computation and Simulation **64**, 105-123.

Horn, R. and Johnson, C. (1990). Matrix analysis. *Cambridge Univesity Press.*

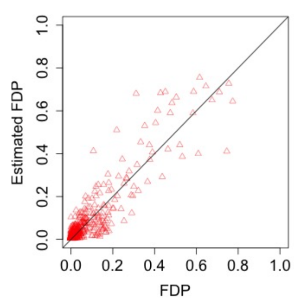
Fan, J. and Han, X. (2017). Estimation of the false discovery proportion with unknown dependence. *Journal of the Royal Statistical Society, Series B* **79**, 1143-1164.



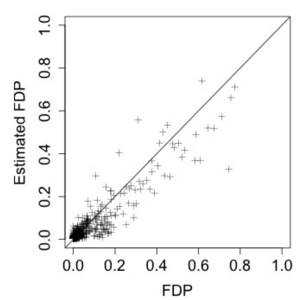
(a) $f(2,4), \mathbf{B} \sim U(-1,1)$



(b) $f(2,4), \mathbf{B} \sim U(-1,1)$

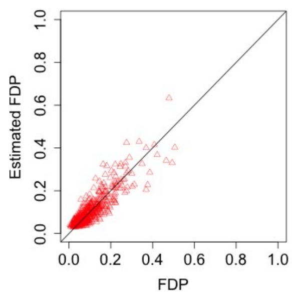


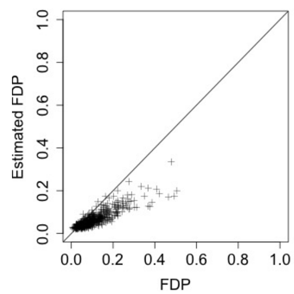
(c)$f(3,3), \mathbf{B} \sim N(0,1)$



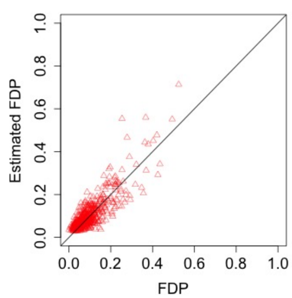
(d)$f(3,3), \mathbf{B} \sim N(0,1)$

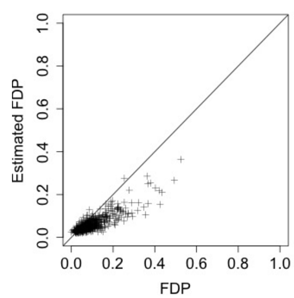
Figure S7: Model 1, the estimated values of FDP obtained by sandwich method (red triangle) and PFA (black crossover) are compared with the true value of FDP. Here, $n = m = 100$, $p = q = 500$, and $t = 0.0001$.





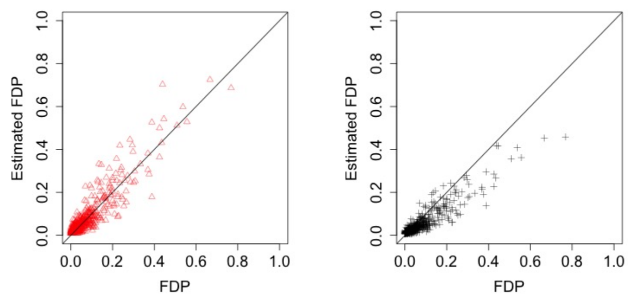
(a) $(\rho_1, \rho_2) = (0.5, 0.3)$ (b) $(\rho_1, \rho_2) = (0.5, 0.3)$



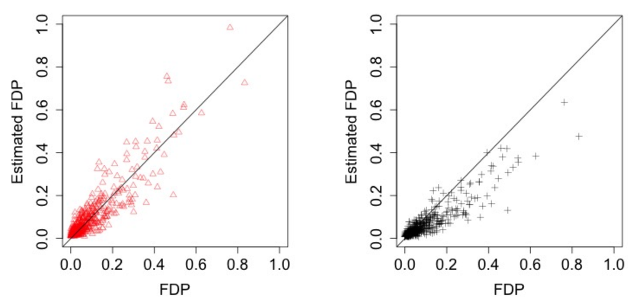


(c)$(\rho_1, \rho_2) = (0.5, 0.8)$ (d)$(\rho_1, \rho_2) = (0.5, 0.8)$

Figure S8: Model 2, the estimated values of FDP obtained by sandwich method (red triangle) and PFA (black crossover) are compared with the true value of FDP. Here, $n = m = 100$, $p = q = 500$, and $t = 0.0001$.

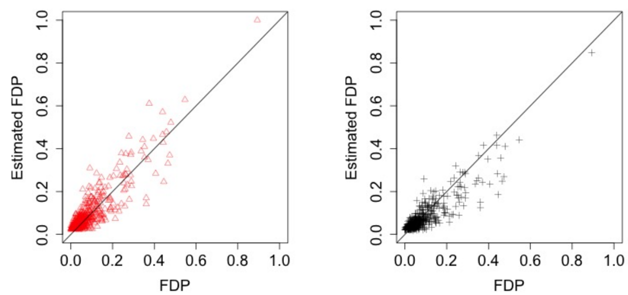


(a) $f(3,3), \mathbf{W} \sim Exp(1)$ (b) $f(3,3), \mathbf{W} \sim Exp(1)$

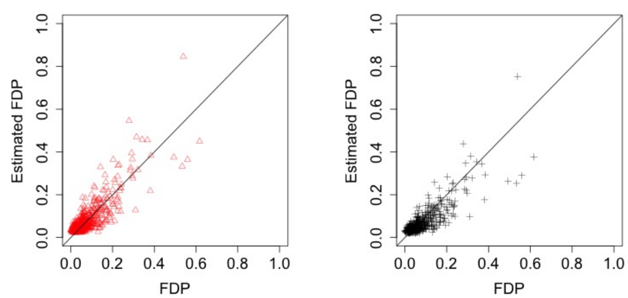


(c)$f(3,3), \mathbf{W} \sim \sqrt{\frac{2}{3}} t_6$ (d)$f(3,3), \mathbf{W} \sim \sqrt{\frac{2}{3}} t_6$

Figure S9: Model 3: factor (3,3) case, the estimated values of FDP obtained by sandwich method (red triangle) and PFA (black crossover) are compared with the true value of FDP. Each element in $\mathbf{W}$ follows $\sqrt{\frac{2}{3}} t_6$ distribution. Here, $n = m = 100$, $p = q = 500$, and $t = 0.0001$.

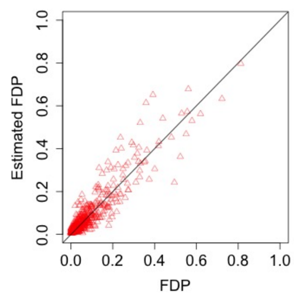


(a) $f(2,2), \mathbf{W} \sim Exp(1)$ (b) $f(2,2), \mathbf{W} \sim Exp(1)$

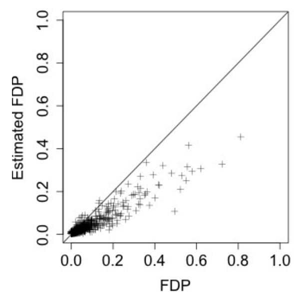


(c)$f(2,2), \mathbf{W} \sim \sqrt{\frac{2}{3}}t_6$ (d)$f(2,2), \mathbf{W} \sim \sqrt{\frac{2}{3}}t_6$

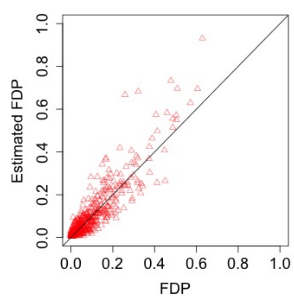
Figure S10: Model 3: factor (2,2) case, the estimated values of FDP obtained by sandwich method (red triangle) and PFA (black crossover) are compared with the true value of FDP. The first row represents the case where $\mathbf{W}$ with each element following Exp(1) distribution, and the second row shows the case where $\mathbf{W}$ with each element following $\sqrt{\frac{2}{3}}t_6$ distribution. Here, $n = m = 100$, $p = q = 500$, and $t = 0.0001$.



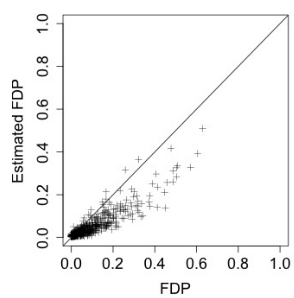
(a) $f(4,4), \mathbf{W} \sim Exp(1)$



(b) $f(4,4), \mathbf{W} \sim Exp(1)$

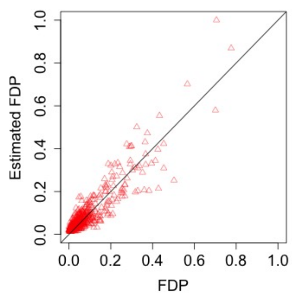


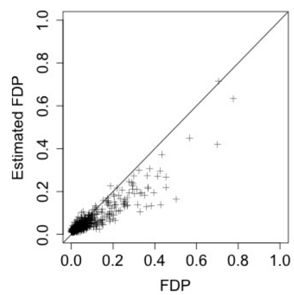
(c)$f(4,4), \mathbf{W} \sim \sqrt{\frac{2}{3}}t_6$



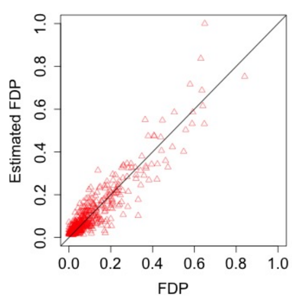
(d)$f(4,4), \mathbf{W} \sim \sqrt{\frac{2}{3}}t_6$

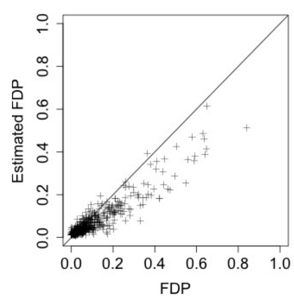
Figure S11: Model 3: factor (4,4) case, the estimated values of FDP obtained by sandwich method (red triangle) and PFA (black crossover) are compared with the true value of FDP. The first row represents the case where $\mathbf{W}$ with each element following Exp(1) distribution, and the second row shows the case where $\mathbf{W}$ with each element following $\sqrt{\frac{2}{3}}t_6$ distribution. Here, $n = m = 100$, $p = q = 500$, and $t = 0.0001$.





(a) $f(2,4), \mathbf{W} \sim Exp(1)$ (b) $f(2,4), \mathbf{W} \sim Exp(1)$





(c)$f(2,4), \mathbf{W} \sim \sqrt{\frac{2}{3}}t_6$ (d)$f(2,4), \mathbf{W} \sim \sqrt{\frac{2}{3}}t_6$

Figure S12: Model 3: factor (2,4) case, the estimated values of FDP obtained by sandwich method (red triangle) and PFA (black crossover) are compared with the true value of FDP. The first row represents the case where $\mathbf{W}$ with each element following Exp(1) distribution, and the second row shows the case where $\mathbf{W}$ with each element following $\sqrt{\frac{2}{3}}t_6$ distribution. Here, $n = m = 100$, $p = q = 500$, and $t = 0.0001$.

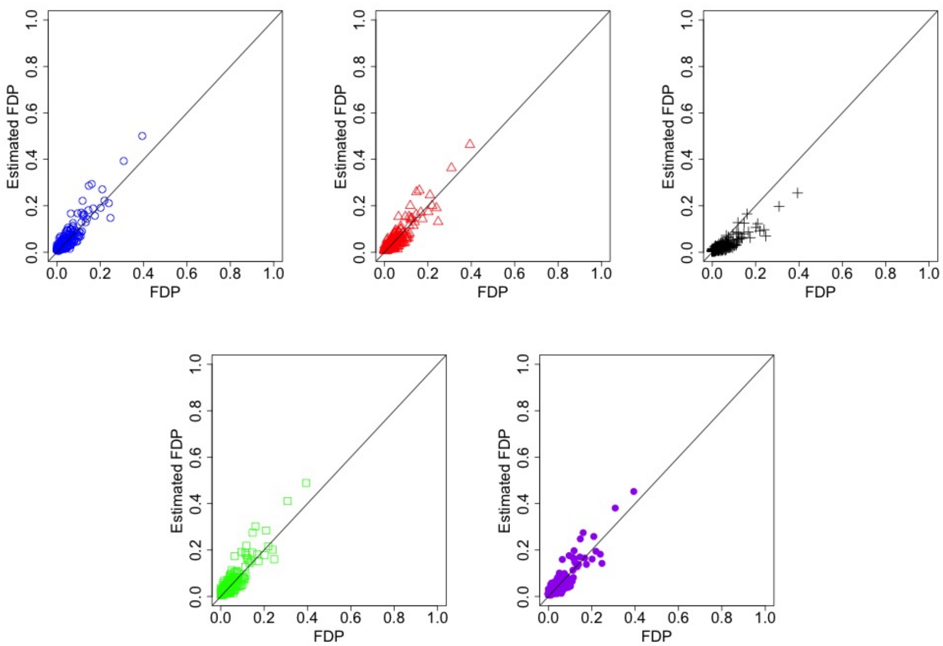
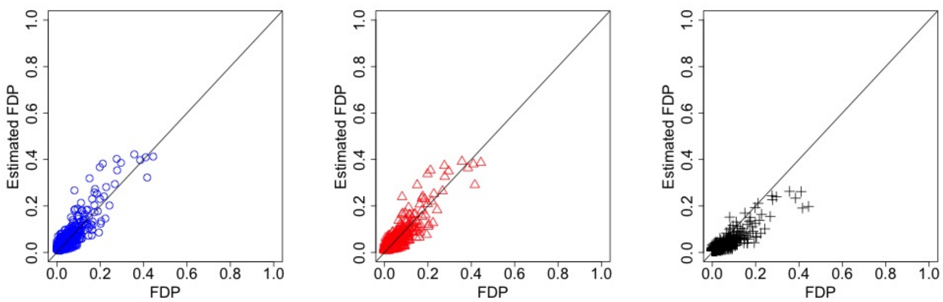
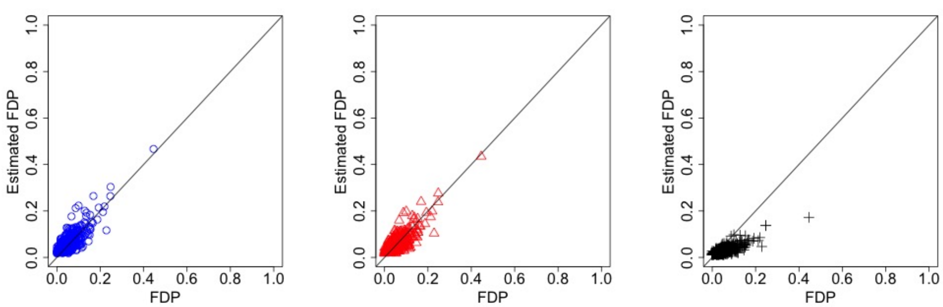


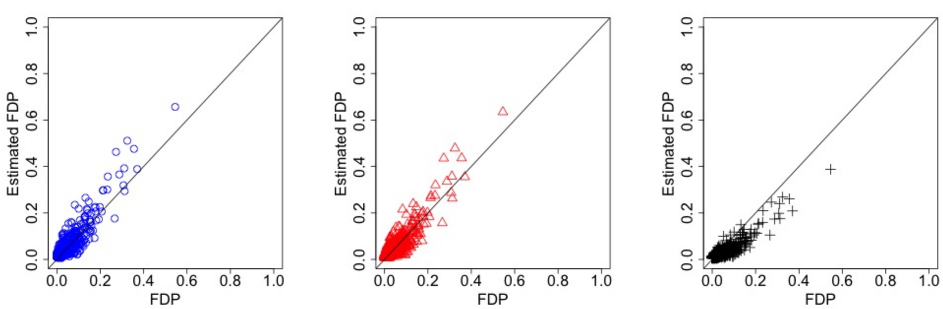
Figure S13: Model 1: factor(3,3) case, the estimated values of FDP obtained by noodle method with $\widehat{\mathbf{\Sigma}}_S$ (blue circle) and $\widehat{\mathbf{\Sigma}}_{MLE}$ (green square), sandwich method with $\widehat{\mathbf{\Sigma}}_S$ (red triangle) and $\widehat{\mathbf{\Sigma}}_{MLE}$ (purple solid dot), and PFA (black crossover) are compared with the true value of FDP. Here, $n = m = 50, p = q = 100$, and $t = 0.001$.



(a) $f(2,4), \mathbf{B} \sim U(-1,1)$

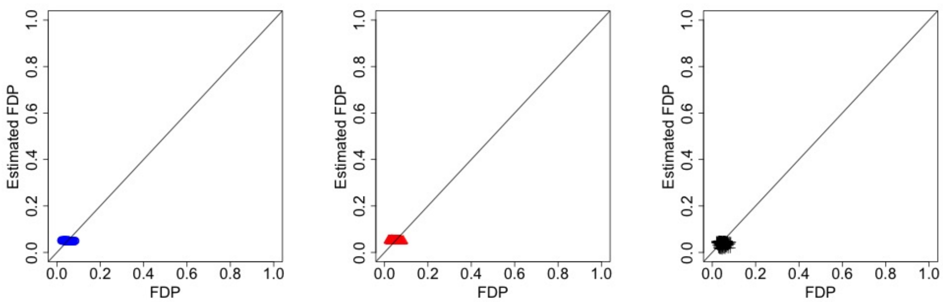


(b) $(\rho_1, \rho_2) = (0.5, 0.3)$

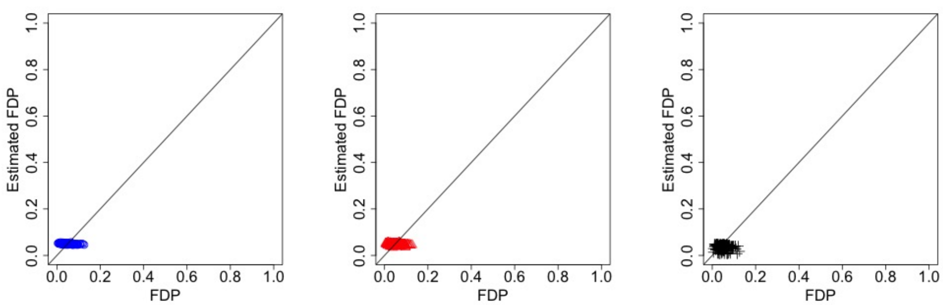


(c) $f(3,3), \mathbf{W} \sim Exp(1)$

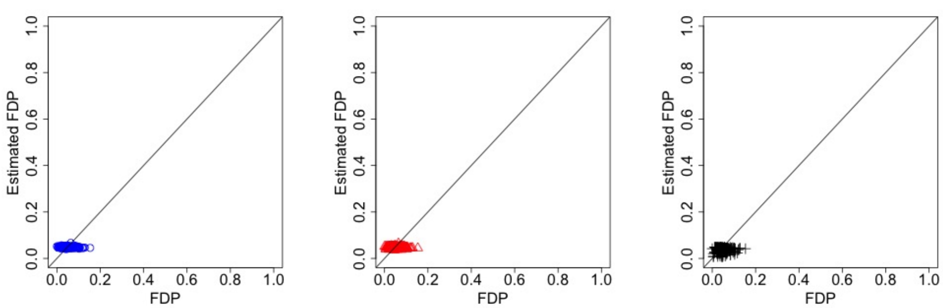
Figure S14: The estimated values of FDP obtained by noodle method (blue circle), sandwich method (red triangle) and PFA (black crossover) are compared with the true value of FDP. From top to bottom, each row corresponds to one scenario from Model 1, Model 2 and Model 3. Here, $n = 25, m = 100, p = q = 100$, and $t = 0.001$.



(a) both identical matrices: $\mathbf{\Sigma}_1 = \mathbf{\Sigma}_2 = \mathbf{I}_{100}$

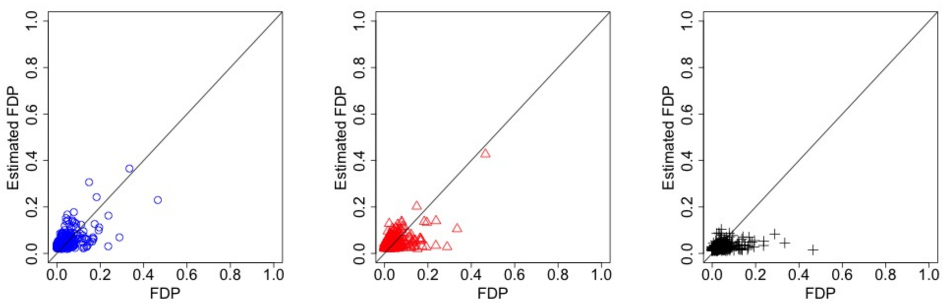


(b) $\mathbf{\Sigma}_1 = \mathbf{I}_{100}$ and $\mathbf{\Sigma}_2$ is factor 4 matrix

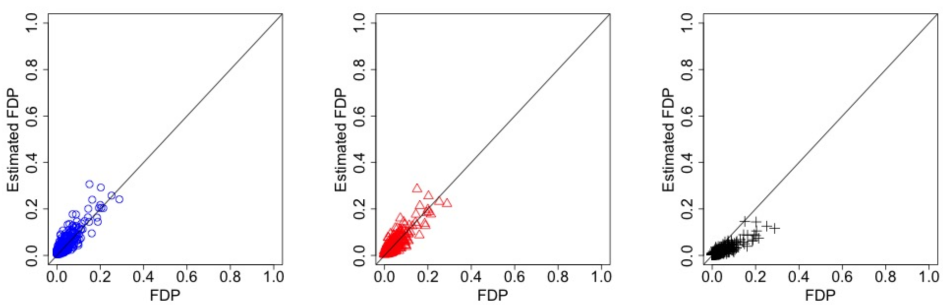


(c) $\mathbf{\Sigma}_2 = \mathbf{I}_{100}$ and $\mathbf{\Sigma}_1$ is factor 2 matrix

Figure S15: The estimated values of FDP obtained by the noodle method (blue circle), sandwich method (red triangle), and PFA (black crossover) are compared with the true value of FDP. From top to bottom, each row corresponds to the Model 1 setting with different structures of the correlation matrix. Here, $n = m = 50, p = q = 100$, and $t = 0.001$.



(a) $k_1 = k_2 = 3, h = 9, \hat{k}_1 = \hat{k}_2 = 2, \hat{h} = 4$



(b) $k_1 = k_2 = 3, h = 9, \hat{k}_1 = \hat{k}_2 = 5, \hat{h} = 12$

Figure S16: The estimated values of FDP obtained by the noodle method (blue circle), sandwich method (red triangle), and PFA (black crossover) are compared with the true value of FDP. (a) underestimate the factor numbers (b) overestimate the factor numbers. Here, $n = m = 50, p = q = 100$ and $t = 0.001$.

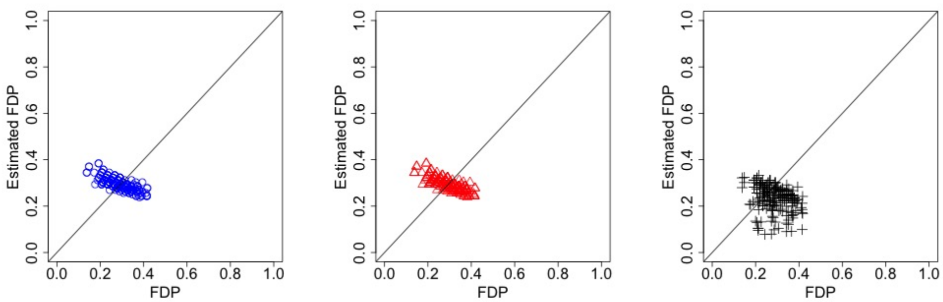
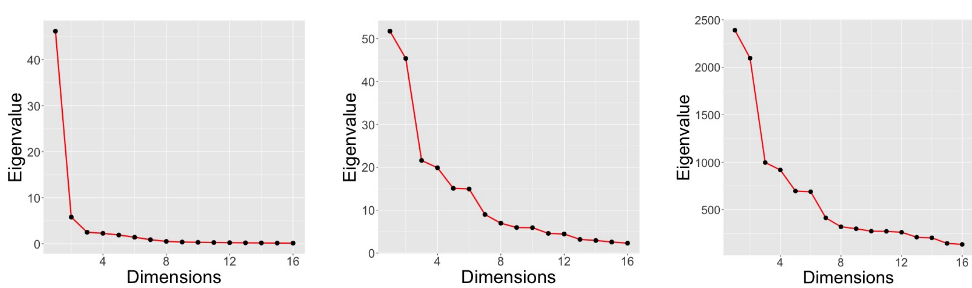


Figure S17: For the case when each element of a matrix follows a normal distribution but the whole matrix does not: the estimated values of FDP obtained by noodle method (blue circle), sandwich method (red triangle) and PFA (black crossover) are compared with the true value of FDP. Here, $n = m = 50, p = q = 100$ and $t = 0.001$.



(a) λ from $corr(\hat{\Sigma}_{64\times 64})$ (b) ξ from $corr(\hat{\Sigma}_{256\times 256})$ (c) $\lambda \otimes \xi$

Figure S18: Plot of eigenvalues for two estimated correlation matrix and their sorted product from EEG data, $\lambda = (\lambda_1, ..., \lambda_p)$, $\xi = (\xi_1, ..., \xi_q)$.

Table S1: Mean and standard deviation of $\widehat{FDP}(t) - FDP(t)$ are presented in percent, with $n = m = 100$, $p = q = 500$ and $t = 0.0001$.

Results (%) for the following methods:				
$n = m = 100$	Sandwich Method		PFA	
$p = q = 500$	Bias	Sd	Bias	Sd
Model 1				
$f(2,4), B \sim U(-1,1)$	0.457	4.499	-2.38	4.673
$f(3,3), B \sim N(0,1)$	0.359	6.043	-1.185	5.568
Model 2				
$(\rho_1, \rho_2) = (0.5, 0.3)$	0.205	3.466	-4.114	4.542
$(\rho_1, \rho_2) = (0.5, 0.8)$	0.309	4.259	-4.064	4.823
Model 3				
$f(2,2), W \sim Exp(1)$	1.256	4.648	-0.639	4.232
$f(2,2), W \sim \sqrt{\frac{2}{3}}t_6$	0.887	4.701	-0.971	4.43
$f(2,4), W \sim Exp(1)$	0.972	4.832	-2.249	4.805
$f(2,4), W \sim \sqrt{\frac{2}{3}}t_6$	0.53	4.312	-2.484	4.65
$f(3,3), W \sim Exp(1)$	1.136	5.041	-2.286	5.288
$f(3,3), W \sim \sqrt{\frac{2}{3}}t_6$	0.505	4.76	-2.501	4.705
$f(4,4), W \sim Exp(1)$	0.809	4.907	-4.098	5.590
$f(4,4), W \sim \sqrt{\frac{2}{3}}t_6$	0.780	4.270	-3.702	5.270

Table S2: Mean and standard deviation of $\widehat{FDP}(t) - FDP(t)$ are presented in percent, with $n = m = 50$, $p = q = 100$ and $t = 0.001$.

Results (%) for the following methods:										
$n = m = 50$	Noodle_S		Noodle_MLE		Sandwich_S		Sandwich_MLE		PFA	
$p = q = 100$	Bias	SD	Bias	SD	Bias	SD	Bias	SD	Bias	SD
Model 1										
$f(2,4), B \sim U(-1,1)$	0.704	2.701	0.686	2.833	0.169	2.576	0.154	2.722	-1.976	3.124
$f(3,3), B \sim N(0,1)$	1.106	3.435	1.046	3.465	0.672	3.119	0.620	3.181	-0.925	2.882
Model 2										
$(\rho_1, \rho_2) = (0.5, 0.3)$	0.821	2.577	0.868	2.640	0.288	2.410	0.330	2.460	-2.617	2.650
$(\rho_1, \rho_2) = (0.5, 0.8)$	1.051	3.388	1.083	3.434	0.498	3.173	0.525	3.212	-1.899	3.558
Model 3										
$f(2,2), W \sim Exp(1)$	1.295	2.866	1.311	2.960	0.820	2.583	0.833	2.646	-0.642	2.496
$f(2,2), W \sim \sqrt{\frac{2}{3}}t_6$	0.755	2.345	0.728	2.333	0.364	2.208	0.337	2.213	-0.735	2.268
$f(2,4), W \sim Exp(1)$	1.055	3.072	1.037	3.320	0.525	2.831	0.510	3.058	-1.697	2.731
$f(2,4), W \sim \sqrt{\frac{2}{3}}t_6$	0.636	2.187	0.666	2.330	0.176	2.067	0.205	2.192	-1.613	2.491
$f(3,3), W \sim Exp(1)$	1.360	3.179	1.357	3.513	0.739	2.838	0.739	3.152	-1.766	2.591
$f(3,3), W \sim \sqrt{\frac{2}{3}}t_6$	0.914	2.244	0.923	2.370	0.427	2.150	0.436	2.271	-1.445	2.442
$f(4,4), W \sim Exp(1)$	1.321	3.162	1.346	3.549	0.684	2.956	0.714	3.323	-2.772	3.446
$f(4,4), W \sim \sqrt{\frac{2}{3}}t_6$	1.243	2.654	1.223	2.771	0.657	2.461	0.641	2.589	-2.359	2.975

Table S3: Mean and standard deviation of $\widehat{FDP}(t) - FDP(t)$ are presented in percent, with $n = m = 100$, $p = q = 500$ and $t = 0.0001$.

Results (%) for the following methods:						
$n = m = 100$	Sandwich_S		Sandwich_MLE		PFA	
$p = q = 500$	Bias	SD	Bias	SD	Bias	SD
Model 1						
$f(2,4), B \sim U(-1,1)$	0.418	5.341	0.386	5.501	-2.560	5.570
$f(3,3), B \sim N(0,1)$	0.810	4.354	0.739	4.628	-2.649	4.978
Model 2						
$(\rho_1, \rho_2) = (0.5, 0.3)$	0.615	3.904	0.607	3.958	-3.930	5.084
$(\rho_1, \rho_2) = (0.5, 0.8)$	0.764	4.165	0.766	4.174	-3.873	4.919
Model 3						
$f(2,2), W \sim Exp(1)$	0.660	4.977	0.720	5.096	-1.145	4.537
$f(2,2), W \sim \sqrt{\frac{2}{3}}t_6$	0.510	5.582	0.625	5.551	-1.425	5.000
$f(2,4), W \sim Exp(1)$	0.398	5.116	0.463	5.441	-2.846	5.512
$f(2,4), W \sim \sqrt{\frac{2}{3}}t_6$	0.631	4.894	0.681	4.896	-2.407	5.099
$f(3,3), W \sim Exp(1)$	1.498	5.170	1.609	5.810	-2.036	5.450
$f(3,3), W \sim \sqrt{\frac{2}{3}}t_6$	0.183	4.861	0.287	4.780	-2.778	5.226
$f(4,4), W \sim Exp(1)$	0.699	4.237	0.637	4.723	-4.189	4.975
$f(4,4), W \sim \sqrt{\frac{2}{3}}t_6$	0.308	4.367	0.370	4.520	-4.275	5.813

Table S4: Comparison of the dependence-adjusted procedure with the fixed threshold procedure under Model 1: $f(2,4), B \sim U(-1,1)$

$p = q = 100$	Fixed threshold			Dependence-adjusted		
	$FDR(\%)$	$FNR(\%)$	Threshold	$FDR(\%)$	$FNR(\%)$	Threshold
			Noodle Method			
$n = m = 20$	3.580	2.915	0.00047	3.576	1.229	0.001
$n = m = 40$	2.339	0.677	0.00087	2.350	0.112	0.001
$n = m = 60$	2.354	0.069	0.00111	2.349	0.006	0.001
			Sandwich Method			
$n = m = 20$	3.665	2.898	0.00049	3.670	1.330	0.001
$n = m = 40$	2.366	0.673	0.00088	2.357	0.112	0.001
$n = m = 60$	2.354	0.069	0.00111	2.360	0.006	0.001
			PFA			
$n = m = 20$	11.180	2.055	0.00281	11.171	1.359	0.001
$n = m = 40$	6.426	0.370	0.00277	6.420	0.091	0.001
$n = m = 60$	4.699	0.039	0.00227	4.707	0.005	0.001

Table S5: Comparison of the dependence-adjusted procedure with the fixed threshold procedure under Model 2: $(\rho_1, \rho_2) = (0.5, 0.8)$

$p = q = 100$	Fixed threshold			Dependence-adjusted		
	$FDR(\%)$	$FNR(\%)$	Threshold	$FDR(\%)$	$FNR(\%)$	Threshold
			Noodle Method			
$n = m = 20$	5.387	2.697	0.00082	5.394	2.304	0.001
$n = m = 40$	2.445	0.679	0.00088	2.447	0.249	0.001
$n = m = 60$	2.275	0.076	0.00104	2.282	0.022	0.001
			Sandwich Method			
$n = m = 20$	4.571	2.824	0.00061	4.571	1.941	0.001
$n = m = 40$	2.445	0.679	0.00088	2.448	0.248	0.001
$n = m = 60$	2.275	0.076	0.00104	2.284	0.022	0.001
			PFA			
$n = m = 20$	8.982	2.284	0.00190	8.974	2.134	0.001
$n = m = 40$	4.956	0.458	0.00194	4.963	0.382	0.001
$n = m = 60$	4.902	0.039	0.00232	4.906	0.022	0.001

Table S6: Comparison of the dependence-adjusted procedure with the fixed threshold procedure under Model 3: $f(2,4), W \sim Exp(1)$

$p = q = 100$	Fixed threshold			Dependence-adjusted		
	$FDR(\%)$	$FNR(\%)$	Threshold	$FDR(\%)$	$FNR(\%)$	Threshold
			Noodle Method			
$n = m = 20$	3.798	2.729	0.00068	3.815	1.849	0.001
$n = m = 40$	2.390	0.680	0.00092	2.394	0.170	0.001
$n = m = 60$	2.350	0.088	0.00105	2.355	0.013	0.001
			Sandwich Method			
$n = m = 20$	3.842	2.723	0.00069	3.821	1.867	0.001
$n = m = 40$	2.390	0.680	0.00092	2.401	0.169	0.001
$n = m = 60$	2.350	0.088	0.00105	2.355	0.013	0.001
			PFA			
$n = m = 20$	9.514	2.089	0.00258	9.508	1.663	0.001
$n = m = 40$	6.481	0.388	0.00282	6.470	0.137	0.001
$n = m = 60$	4.719	0.051	0.00218	4.723	0.012	0.001

Table S7: Mean and standard deviation of $\widehat{\mathrm{FDP}}(t) - \mathrm{FDP}(t)$ are presented in percent, with $n = 25, m = 100$, $p = q = 100$ and $t = 0.001$.

Results (%) for the following methods:						
$n = 25, m = 100$	Noodle Method		Sandwich Method		PFA	
$p = q = 100$	Bias	Sd	Bias	Sd	Bias	Sd
Model 1						
$f(2,4), B \sim U(-1,1)$	0.902	3.198	0.295	2.995	-1.748	3.372
$f(3,3), B \sim N(0,1)$	0.972	3.847	0.537	3.554	-0.610	3.087
Model 2						
$(\rho_1, \rho_2) = (0.5, 0.3)$	0.853	2.614	0.307	2.505	-2.361	3.049
$(\rho_1, \rho_2) = (0.5, 0.8)$	0.806	3.107	0.239	2.961	-2.358	3.512
Model 3						
$f(2,2), W \sim Exp(1)$	0.823	3.240	0.311	3.079	-0.917	3.185
$f(2,2), W \sim \sqrt{\frac{2}{3}}t_6$	1.001	3.424	0.458	3.139	-0.817	3.005
$f(2,4), W \sim Exp(1)$	0.921	2.940	0.329	2.708	-1.686	2.925
$f(2,4), W \sim \sqrt{\frac{2}{3}}t_6$	0.900	3.105	0.287	2.858	-1.814	2.968
$f(3,3), W \sim Exp(1)$	1.049	3.247	0.386	2.961	-1.911	3.029
$f(3,3), W \sim \sqrt{\frac{2}{3}}t_6$	1.116	3.244	0.465	2.952	-1.832	3.060
$f(4,4), W \sim Exp(1)$	1.456	3.565	0.751	3.251	-2.601	3.459
$f(4,4), W \sim \sqrt{\frac{2}{3}}t_6$	1.135	3.050	0.448	2.843	-2.715	3.523