Supplementary Material for "CARS: Covariate Assisted Ranking and Screening for Large-Scale Two-Sample Inference"

This supplement contains the proofs of other theoretical results (Section A) and additional numerical results (Section B).

A. Additional Theory and Proofs of Other Results

A.1. Proof of Proposition 1

PROOF. Let $A_x = \{x : x \neq 0\}$. According to the definition of θ_{ji} and the conditional independence assumption (2.5), we have

$$\begin{aligned} f(t_{1i}, t_{2i}|\theta_{1i} = 0, \theta_{2i} = 0) &= f(t_{1i}, t_{2i}|\mu_{xi} = 0, \mu_{yi} = 0) \\ &= f(t_{1i}|\mu_{xi} = 0, \mu_{yi} = 0) f(t_{2i}|\mu_{xi} = 0, \mu_{yi} = 0) \\ &= f(t_{1i}|\theta_{1i} = 0, \mu_{xi} = \mu_{yi} = 0) f(t_{2i}|\theta_{2i} = 0, \mu_{xi} = \mu_{yi} = 0) \\ &= f(t_{1i}|\theta_{1i} = 0) f(t_{2i}|\theta_{2i} = 0). \end{aligned}$$

The last equality holds because given $\theta_{ji} = 0$, t_{ji} are just linear combinations of random errors ϵ_{xi} and ϵ_{yi}, which are independent of μ_{xi} and μ_{yi}, $j = 1, 2$. Denote $G_{\mu_x}(\cdot)$ the distribution function of μ_{xi}. Then

$$\begin{aligned} f(t_{1i}, t_{2i}|\theta_{1i} = 0, \theta_{2i} = 1) &= \int_{A_x} f(t_{1i}, t_{2i}|\mu_{xi} = \mu_{yi} = x) dG_{\mu_x}(x) \\ &= \int_{A_x} f(t_{1i}|\theta_{1i} = 0, \mu_{xi} = \mu_{yi} = x) f(t_{2i}|\mu_{xi} = \mu_{yi} = x) dG_{\mu_x}(x) \\ &= f(t_{1i}|\theta_{1i} = 0) \int_{A_x} f(t_{2i}|\theta_{1i} = 0, \mu_{xi} = \mu_{yi} = x) dG_{\mu_x}(x) \\ &= f(t_{1i}|\theta_{1i} = 0) f(t_{2i}|\theta_{1i} = 0, \theta_{2i} = 1). \end{aligned}$$

For the third equality, note that t_{1i} are independent of both μ_{xi} and μ_{yi} given $\theta_{1i} = 0$, whereas t_{2i} are correlated with μ_{xi} and μ_{yi} given that $\theta_{1i} = 0$. Finally,

$$\begin{aligned} f(t_{1i}, t_{2i}|\theta_{1i} = 0) &= f(t_{1i}, t_{2i}|\theta_{1i} = 0, \theta_{2i} = 0)\mathbb{P}(\theta_{2i} = 0|\theta_{1i} = 0) \\ &\quad + f(t_{1i}, t_{2i}|\theta_{1i} = 0, \theta_{2i} = 1)\mathbb{P}(\theta_{2i} = 1|\theta_{1i} = 0) \\ &= f(t_{1i}|\theta_{1i} = 0)\{f(t_{2i}|\theta_{1i} = 0, \theta_{2i} = 0)\mathbb{P}(\theta_{2i} = 0|\theta_{1i} = 0) \\ &\quad + f(t_{2i}|\theta_{1i} = 0, \theta_{2i} = 1)\mathbb{P}(\theta_{2i} = 1|\theta_{1i} = 0)\} \\ &= f(t_{1i}|\theta_{1i} = 0) f(t_{2i}|\theta_{1i} = 0). \end{aligned}$$

Hence T_{1i} and T_{2i} are independent under the null hypothesis $\theta_{1i} = 0$. □

A.2. Proof of Proposition 2

PROOF. **Part (a).** We only need to show that $q^\tau(t_2) \geq q^*(t_2)$ for all τ. We first split the joint density into two parts and then use the independence between T_{1i} and T_{2i} under

the null:

$$\begin{aligned} q^\tau(t_2) &= (1-\tau)^{-1}\int_{t_1\in\mathcal{A}_\tau}\{f(t_1,t_2|\theta_{1i}=0)\mathbb{P}(\theta_{1i}=0)+f(t_1,t_2|\theta_{1i}=1)\mathbb{P}(\theta_{1i}=1)\}\,dt_1 \\ &\geq (1-\tau)^{-1}\int_{t_1\in\mathcal{A}_\tau} q^*(t_2)f_{10}(t_1)dt_1 = q^*(t_2). \end{aligned}$$

The last equality holds because $\int_{t_1\in\mathcal{A}_\tau} f_{10}(t_1)dt_1 = 1-\tau$, which cancels with $(1-\tau)^{-1}$.

Part (b). Let $\mathcal{R}$ be the index set of hypotheses rejected by $\boldsymbol{\delta}^\tau_{OR}$. The FDR of $\boldsymbol{\delta}^\tau_{OR}$ is

$$\begin{aligned} \mathrm{FDR}(\boldsymbol{\delta}^\tau_{OR}) &= \mathbb{E}\left\{\frac{\sum_{i\in\mathcal{R}}(1-\theta_{1i})}{|\mathcal{R}|\vee 1}\right\} \\ &= \mathbb{E}_{\boldsymbol{T}_1,\boldsymbol{T}_2}\left[\mathbb{E}\left\{\frac{\sum_{i\in\mathcal{R}}(1-\theta_{1i})}{|\mathcal{R}|\vee 1}\mid \boldsymbol{T}_1,\boldsymbol{T}_2\right\}\right] \\ &= \mathbb{E}_{\boldsymbol{T}_1,\boldsymbol{T}_2}\left(\frac{1}{|\mathcal{R}|\vee 1}\sum_{i\in\mathcal{R}} T^i_{OR}\right), \end{aligned}$$

where the last equality follows from the definition of the oracle statistic $T^i_{OR} = \mathbb{E}(1-\theta_{1i}|T_{1i},T_{2i})$. For all $i\in\mathcal{R}\subset I$, we have $T^i_{OR}\leq T^{\tau,i}_{OR}$. It follows that

$$\mathrm{FDR}(\boldsymbol{\delta}^\tau_{OR}) \leq \mathbb{E}_{\boldsymbol{T}_1,\boldsymbol{T}_2}\left(\frac{1}{|\mathcal{R}|\vee 1}\sum_{i\in\mathcal{R}} T^{\tau,i}_{OR}\right)\leq\alpha.$$

The last inequality is due to the operation of $\boldsymbol{\delta}^\tau_{OR}$, which guarantees that

$$\{|\mathcal{R}|\vee 1\}^{-1}\sum_{i\in\mathcal{R}} T^{\tau,i}_{OR}\leq\alpha$$

for all realizations of $(\boldsymbol{T}_1,\boldsymbol{T}_2)$. The result on mFDR holds since

$$\mathbb{E}\left\{\sum_{i\in\mathcal{R}}(1-\theta_{1i})\right\}\leq\mathbb{E}_{\boldsymbol{T}_1,\boldsymbol{T}_2}\left(\sum_{i\in\mathcal{R}} T^{\tau,i}_{OR}\right)\leq\alpha\mathbb{E}_{\boldsymbol{T}_1,\boldsymbol{T}_2}(|\mathcal{R}|\vee 1).$$

A.3. Proof of Proposition 3

Let $\tilde{\mathcal{N}} = m\tilde{G}(\tau)$ be the expected size of the screening set. Define

$$\tilde{f}^\tau_{2\cdot}(t_2) = (1/\tilde{\mathcal{N}})\sum_{T_{2i}\in\mathcal{T}_2(\tau)} K_{h_2}(t_2-T_{2i}).$$

Then $\hat{q}^\tau(t_2) = \frac{\tilde{G}(\tau)}{1-\tau}\tilde{f}^\tau_{2\cdot}(t_2)$. Define the conditional density of T_{2i} given that $P_i>\tau$

$$f^\tau_{2\cdot}(t_2) = f(t_2|P_i>\tau) = \tilde{G}(\tau)^{-1}\int_{\mathcal{A}_\tau} f(t_1,t_2)dt_1.$$

Then $q^\tau(t_2) = \frac{\tilde{G}(\tau)}{1-\tau}f^\tau_{2\cdot}(t_2)$. It follows that

$$\mathbb{E}\left\|\hat{q}^\tau - q^\tau\right\|^2 = \left\{\frac{\tilde{G}(\tau)}{1-\tau}\right\}^2\mathbb{E}\left\|\tilde{f}^\tau_{2\cdot}-f^\tau_{2\cdot}\right\|^2,$$

For a fixed τ, we only need to show that

$$\mathbb{E}\left\|\tilde{f}^{\tau}_{2\cdot} - f^{\tau}_{2\cdot}\right\|^2 = \mathbb{E}\int\left\{\tilde{f}^{\tau}_{2\cdot}(t_2) - f^{\tau}_{2\cdot}(t_2)\right\}^2 dt_2 \to 0.$$

Let $\mathbb{E}_{T_2|T_1}$ denote the conditional expectation that is taken over T_{2i}'s while holding T_{1i}'s fixed. The conditional density of T_2 given T_1 is denoted $f_{2\cdot}(t_2|T_{1i})$. We will work on the bias and variance in turn. First,

$$\mathbb{E}_{T_2|T_1}\{K_{h_2}(t_2 - T_{2i})\} = f_{2\cdot}(t_2|T_{1i}) + \frac{h_2^2}{2}f^{(2)}_{2\cdot}(t_2|T_{1i})\int z^2K(z)dz + o(h_2^2).$$

Now consider $\mathbb{E}\{\tilde{f}^{\tau}_{2\cdot}(t_2)\} = \mathbb{E}_{T_1}[\mathbb{E}_{T_2|T_1}\{\tilde{f}^{\tau}_{2\cdot}(t_2)\}]$. It follows that

$$\begin{aligned}\mathbb{E}\{\tilde{f}^{\tau}_{2\cdot}(t_2)\} &= (1/\tilde{\mathcal{N}})\mathbb{E}_{T_1}\left[\textstyle\sum_{i=1}^m\left\{f_{2\cdot}(t_2|T_{1i}) + \frac{h_2^2}{2}f^{(2)}_{2\cdot}(t_2|T_{1i})\int z^2K(z)dz\right.\right.\\ &\quad \left.\left.+o(h_2^2)\right\}\mathbb{I}\left(T_{1i}\in\mathcal{A}_\tau\right)\right]\\ &= f^{\tau}_{2\cdot}(t_2) + \frac{h_2^2}{2}f^{(2)}_{2\cdot}(t_2|\tau)\int z^2K(z)dz + o(h_2^2),\end{aligned}$$

where $f^{(2)}_{2\cdot}(t_2|\tau)$ is defined in the proposition. The second equality holds because

$$\mathbb{E}_{T_1}\{f(t_2|T_{1i})I(T_{1i}\in\mathcal{A}_\tau)\} = \int_{\mathcal{A}_\tau} f_{2\cdot}(t_2|t_1)f(t_1)dt_1 = \tilde{G}(\tau)f^{\tau}_{2\cdot}(t_2).$$

We have assumed the square integrability of $f^{(2)}_{2\cdot}(t_2|\tau)$, according to which define

$$R\left\{f^{(2)}_{2|\tau}\right\} = \int\{f^{(2)}_{2\cdot}(t_2|\tau)\}^2dt_2.$$

It follows that the integrated squared bias is

$$\int[\mathbb{E}\{\tilde{f}^{\tau}_{2\cdot}(t_2)\} - \tilde{f}^{\tau}_{2\cdot}(t_2)]^2dt_2 = (h_2^4/4)R\left\{f^{(2)}_{2\cdot}(\tau)\right\}\{\mu_2(K)\}^2(1+o(1)),$$

where we use the notation $\mu_2(K) = \int z^2K(z)dz$.

Next, we compute the variance term. Consider the following decomposition

$$\mathsf{Var}\{\tilde{f}^{\tau}_{2\cdot}(t_2)\} = \mathsf{Var}_{T_1}[\mathbb{E}_{T_2|T_1}\{\tilde{f}^{\tau}_{2\cdot}(t_2)\}] + \mathbb{E}_{T_1}[\mathsf{Var}_{T_2|T_1}\{\tilde{f}^{\tau}_{2\cdot}(t_2)\}],$$

where the first and second terms can be respectively computed as

$$\begin{aligned}\mathsf{Var}_{T_1}[\mathbb{E}_{T_2|T_1}\{\tilde{f}^{\tau}_{2\cdot}(t_2)\}] &= \left\{m\tilde{G}^2(\tau)\right\}^{-1}\left[\int\{f_{2|1}(t_2|t_1)\}^2f_{1\cdot}(t_1)dt_1 - \{f_{2\cdot}(t_2)\}^2\right]\{1+o(1)\},\\ \mathbb{E}_{T_1}[\mathsf{Var}_{T_2|T_1}\{\tilde{f}^{\tau}_{2\cdot}(t_2)\}] &= \frac{1}{\tilde{\mathcal{N}}h_2}\int_{A_\tau} f_{2\cdot}(t_2|t_1)f_{1\cdot}(t_1)dt_1R(K)\{1+o(1)\}\\ &= (mh_2)^{-1}f^{\tau}_{2\cdot}(t_2)R(K)\{1+o(1)\}.\end{aligned}$$

It is easy to see that the second term is the leading term (note that we fixed τ_k and let $m\to\infty$). Hence

$$\int\mathsf{Var}\{\tilde{f}^{\tau}_{2\cdot}(t_2)\}dt_2 = (mh_2)^{-1}R(K)\{1+o(1)\}.$$

The common choice of bandwidth $h\sim m^{-1/5}$ makes $\mathbb{E}\left\|\tilde{f}^{\tau}_{2\cdot} - f^{\tau}_{2\cdot}\right\|^2\to 0$, and the desired result follows. □

A.4. Proof of Proposition 4

Part (a). The total approximation error can be computed as

$$B_q(\tau) = \int |q^\tau(t_2) - q^*(t_2)|dt_2 = \frac{1-G(\tau)}{1-\tau} - (1-\pi_1) = \frac{\pi_1\{1-G_1(\tau)\}}{1-\tau}.$$

Noting that $G_1(1) = 1$, and applying the mean value theorem, we have

$$\frac{d}{d\tau}B_q(\tau) = \frac{\{1-G_1(\tau)\} - g_1(\tau)(1-\tau)}{(1-\tau)^2} = \frac{g_1(\xi) - g_1(\tau)}{1-\tau}$$

for some $\xi \in (\tau, 1)$. If G_1 is concave, then $g_1(\xi) - g_1(\tau) < 0$ and the desired result follows.

Part (b). Let $\eta(x) = \{1-G_1(x)\}/(1-x)$. If g_1 satisfies $\lim_{x\uparrow 1} g_1(x) = 0$, then using L'Hospital's Rule, we claim that $\lim_{x\uparrow 1}\eta(x)$ exists and $\lim_{x\uparrow 1}\eta(x) = \lim_{x\uparrow 1} g_1(x) = 0$. Consider the following decomposition of the joint density

$$\int_{\mathcal{A}_\tau} f(t_1,t_2)dt_1 = \int_{A_\tau}\{q^*(t_2)f_{10}(t_1) + \pi_1 f(t_2|t_1,\theta_{1i}=1)f_{1\cdot}(t_1)\}dt_1.$$

Noting that both T_{1i} and T_{2i} are standardized, the conditional density $f(t_2|t_1,\theta_1=1)$ is bounded by some constant C_0. Using the definition of $\eta(x)$, we have

$$q^\tau(t_2) = (1-\tau)^{-1}\int_{\mathcal{A}_\tau} f(t_1,t_2)dt_1 \le q^*(t_2) + C_0\pi_1\eta(\tau)$$

Applying L'Hospital's Rule again, we have $\lim_{\tau\uparrow 1} q^\tau(s) \le q^*(s)$. In Proposition 2.(a), we have shown that $q^\tau(s) \ge q^*(s)$ for all τ. Combining the two inequalities, we conclude that $\lim_{\tau\uparrow 1} q^\tau(s) = q^*(s)$. Then the desired result follows. □

A.5. Proof of Proposition 5

Proof. Proposition 4 defines $\eta(x) = \{1-G_1(x)\}/(1-x)$ and shows that

$$\lim_{\tau_k\to 1} q^{\tau_k}(t_2) = q^*(t_2), \quad |q^\tau(t_2) - q^*(t_2)| \le C_0\pi_1\eta(\tau).$$

Note that $G_1(0) = 0$ and $G_1(1) = 1$. It follows from the concavity of G_1 that $\eta(\tau) \le 1$. Moreover, $\int\{q^{\tau_k}(t_2) - q^*(t_2)\}dt_2 = \pi_1\eta(\tau_k)$. Hence

$$\int\{q^{\tau_k}(t_2) - q^*(t_2)\}^2 dt_2 \le C_0\pi_1\int\{q^{\tau_k}(t_2) - q^*(t_2)\}dt_2 = C_0\pi_1^2\eta(\tau_k) \to 0 \tag{A.1}$$

when $\tau_k \to 1$, according to L'Hospital's Rule and our assumption that $\lim_{x\uparrow 1} g_1(x) = 0$. Therefore, we only need to show that for a given τ^k,

$$\mathbb{E}\|\hat{q}^* - q^{\tau_k}\|^2 = \mathbb{E}\left[\int\{\hat{q}^*(t_2) - q^{\tau_k}(t_2)\}^2 dt_2\right] \xrightarrow{m\to 0} 0.$$

Consider $\hat{q}^*(t_2)$ defined in (3.13). Let

$$a_j = k^{-1}(\hat{s}_2\hat{s}_0 - \hat{s}_1^2)^{-1}\{\hat{s}_2 - \hat{s}_1(\tau_j - 1)\}K_{h_\tau}(\tau_j - 1).$$

Then $\hat{q}^*(t_2) = \sum_{j=1}^k a_j \hat{q}^{\tau_j}$, $\sum_{j=1}^k a_j = 1$ and $\sum_{j=1}^k a_j \leq 1$. Following similar steps in the proof of Proposition 3,

$$\begin{aligned}
\mathbb{E}\{\hat{q}^*(t_2)\} &= \mathbb{E}_{T_1}[\mathbb{E}_{T_2|T_1}\{\hat{q}^*(t_2)\}] \\
&= \sum_{j=1}^k a_j \frac{1-G(\tau_j)}{1-\tau_j}\left[f_{2\cdot}^{\tau_j}(t_2) + \frac{h_2^2}{2}u_2(K)\int_{t_1\in\mathcal{A}_{\tau_j}} f_{2\cdot}^{(2)}(t_2|t_1)f_{1\cdot}(t_1)dt_1 + o(h_2^2)\right] \\
&= q^{\tau_k}(t_2) + \frac{1}{2}h_\tau^2 q^{\tau_k,(2)}(t_2)C^*(K) + o(h_\tau^2) \\
&\quad + \frac{h_2^2}{2}u_2(K)\sum_{j=1}^k a_j \frac{1-G(\tau_j)}{1-\tau_j}\int_{t_1\in\mathcal{A}_{\tau_j}} f_{2\cdot}^{(2)}(t_2|t_1)f_{1\cdot}(t_1)dt_1 + o(h_2^2),
\end{aligned}$$

where $C^*(K)$ is a kernel dependent constant. The last equality follows from Wand and Jones (1995) (pp. 128) for kernel smoothing estimator at boundary points [applying the theory to $\sum_{j=1}^k a_j q^{\tau_j}(t_2)$]. Note that $\frac{1-G(\tau_j)}{1-\tau_j} \leq 1$ and $\sum_j a_j = 1$, we have

$$\sum_{j=1}^k a_j \frac{1-G(\tau_j)}{1-\tau_j}\int_{t_1\in\mathcal{A}_{\tau_j}} f_{2\cdot}^{(2)}(t_2|t_1)f_{1\cdot}(t_1)dt_1 \leq \int f_{2\cdot}^{(2)}(t_2|t_1)f_{1\cdot}(t_1)dt_1.$$

According to the square integrability of $q^{\tau_k,(2)}(t_2)$ and $f_{2\cdot}^{(2)}(t_2|\tau)$, we can define

$$\begin{aligned}
R\left\{q^{\tau_k,(2)}\right\} &= \int\left\{q^{\tau_k,(2)}(t_2)\right\}^2 dt_2, \\
R\left\{f_{t_2|\tau}^{(2)}\right\} &= \int\left\{f_{2\cdot}^{(2)}(t_2|\tau)\right\}^2 dt_2.
\end{aligned}$$

Then the leading term of $\int[\mathbb{E}\{q^*(t_2)\} - q^{\tau_k}(t_2)]^2 dt_2$ is bounded above by

$$\frac{1}{2}h_\tau^4 R\left\{q^{\tau_k,(2)}\right\}\{C^*(K)\}^2 + \frac{1}{2}h_2^4\{u_2(K)\}^2 R\left\{f_{t_2|\tau}^{(2)}\right\},$$

which converges to 0 when $(m,k) \to \infty$.

The argument below regarding the variance part does not pursue the study of the optimal rate of convergence. In fact, the analysis is intractable because q^{τ_j} are dependent quantities. We have show in Proposition 3 that

$$\int \mathsf{Var}\{\hat{q}^{\tau_j}(t_2)\}dt_2 \leq (mh_2)^{-1}R(K)\{1+o(1)\}$$

for all τ_j. It follows that

$$\begin{aligned}
\mathsf{Var}\{\hat{q}^*(t_2)\} &\leq (mh_2)^{-1}R(K)(\sum_{j=1}^k a_k^2)\{1+o(1)\} \\
&\leq (mh_2)^{-1}R(K)\{1+o(1)\} \to 0
\end{aligned}$$

when $m \to \infty$. We note that the actual convergence rate would be better because the variance will be greatly reduced by averaging. Combining the results on the bias and

variance terms, we have

$$\begin{aligned} & \mathbb{E}\int\{\hat{q}^*(t_2)-q^{\tau_k}(t_2)\}^2dt_2 \qquad (A.2)\\ \leq\ & \left[\frac{1}{2}h_\tau^4R(q^{\tau_k,(2)})\{C^*(K)\}^2+\frac{1}{2}h_2^4\{u_2(K)\}^2R(f^{(2)}_{t_2|t_1})+\frac{R(K)}{mh_2}\right]\{1+o(1)\}. \end{aligned}$$

Combining (A.1) and (A.2), we conclude that $\mathbb{E}\left\|\hat{q}^*-q^*\right\|^2\to 0$ when $(m,k)\to 0$, completing the proof. □

A.6. Proof of Proposition 6

Proof. We first point out that the arguments in this proof should be understood as the conditional version (given that μ_{xi} and μ_{yi} are known). Define the following central sample moments that are used in later calculations:

$$m_{xi}=\frac{1}{n_x}\sum_{j=1}^{n_x}(X_{ij}-\mu_{xi}),\ m_{yi}=\frac{1}{n_y}\sum_{j=1}^{n_y}(Y_{ij}-\mu_{yi}),$$
$$m_{xxi}=\frac{1}{n_x}\sum_{j=1}^{n_x}(X_{ij}-\mu_{xi})^2,\ m_{yyi}=\frac{1}{n_y}\sum_{j=1}^{n_y}(Y_{ij}-\mu_{yi})^2.$$

It follows that $S_{xi}^2=m_{xxi}-m_{xi}^2,\ S_{yi}^2=m_{yyi}-m_{yi}^2$. Next define central moments,

$$\mu_{xi3}=\mathbb{E}(X_{ij}-\mu_{xi})^3,\ \mu_{yi3}=\mathbb{E}(Y_{ij}-\mu_{yi})^3,$$
$$\mu_{xi4}=\mathbb{E}(X_{ij}-\mu_{xi})^4,\mu_{yi4}=\mathbb{E}(Y_{ij}-\mu_{yi})^4.$$

According to the CLT, we have

$$\sqrt{\frac{n_xn_y}{n}}\left\{\begin{pmatrix}m_{xi}\\m_{yi}\\m_{xxi}\\m_{yyi}\end{pmatrix}-\begin{pmatrix}0\\0\\\sigma_{xi}^2\\\sigma_{yi}^2\end{pmatrix}\right\}\xrightarrow{L}N(0,\Sigma_i),\ \text{where}$$

$$\Sigma_i=\begin{pmatrix}\gamma_y\sigma_{xi}^2 & 0 & \gamma_y\mu_{xi3} & 0\\ 0 & \gamma_x\sigma_{yi}^2 & 0 & \gamma_x\mu_{yi3}\\ \gamma_y\mu_{xi3} & 0 & \gamma_y(\mu_{xi4}-\sigma_{xi}^4) & 0\\ 0 & \gamma_x\mu_{yi3} & 0 & \gamma_x(\mu_{yi4}-\sigma_{yi}^4)\end{pmatrix}.$$

Let $g_i(s,t,u,v)=\begin{pmatrix}(s-t+\mu_{xi}-\mu_{yi})(\gamma_xv+\gamma_yu)^{-\frac{1}{2}}\\ \left(s+\frac{\gamma_yu}{\gamma_xv}t+\mu_{xi}+\frac{\gamma_y}{\gamma_x}\mu_{yi}\right)\left\{(\gamma_xv+\gamma_yu)\frac{\gamma_yu}{\gamma_xv}\right\}^{-\frac{1}{2}}\end{pmatrix}$. It follows that

$$\dot{g}_i(0,0,\sigma_{xi}^2,\sigma_{yi}^2)=$$
$$\begin{pmatrix}\sigma_{pi}^{-1} & -\sigma_{pi}^{-1} & -\frac{1}{2}\gamma_y(\mu_{xi}-\mu_{yi})\sigma_{pi}^{-3} & -\frac{1}{2}\gamma_x(\mu_{xi}-\mu_{yi})\sigma_{pi}^{-3}\\ \sigma_{pi}^{-1}\kappa_i^{*-\frac{1}{2}} & \sigma_{pi}^{-1}\kappa_i^{*\frac{1}{2}} & -\frac{1}{2}\gamma_y\left(\mu_{xi}+\mu_{yi}\frac{\gamma_y}{\gamma_x}\right)\sigma_{pi}^{-3}\left(1+2\kappa_i^*\right)\kappa_i^{*-\frac{3}{2}} & \frac{1}{2}\gamma_x\left(\mu_{xi}+\mu_{yi}\frac{\gamma_y}{\gamma_x}\right)\sigma_{pi}^{-3}\kappa_i^{*\frac{1}{2}}\end{pmatrix}.$$

The asymptotic variance-covariance matrix of $\sqrt{\gamma_x\gamma_y n}(T_{1i}, T_{2i})^\intercal$ is

$$\Sigma^i_{T_1,T_2} = \begin{pmatrix} \sigma_i^2(t_1,t_1) & \sigma_i^2(t_1,t_2) \\ \sigma_i^2(t_1,t_2) & \sigma_i^2(t_2,t_2) \end{pmatrix},$$

where the entries can be computed by apply the Delta method:

$$\begin{aligned}
\sigma_i^2(t_1,t_1) &= 1 + \sigma_{pi}^{-4}(\mu_{xi}-\mu_{yi})\left(\gamma_x^2\mu_{yi3} - \gamma_y^2\mu_{xi3}\right) \\
&\quad + \frac{\sigma_{pi}^{-6}}{4}(\mu_{xi}-\mu_{yi})^2\left\{\gamma_x^3\left(\mu_{yi4}-\sigma_{yi}^4\right) + \gamma_y^3\left(\mu_{xi4}-\sigma_{xi}^4\right)\right\}, \\
\sigma_i^2(t_1,t_2) &= -\frac{\sigma_{pi}^{-4}}{2}(\mu_{xi}-\mu_{yi})\left(\gamma_x^2\mu_{yi3}\kappa_i^{*\frac{1}{2}} + \gamma_y^2\mu_{xi3}\kappa_i^{*-\frac{1}{2}}\right) \\
&\quad - \frac{\sigma_{pi}^{-4}}{2}\left(\mu_{xi}+\mu_{yi}\frac{\gamma_y}{\gamma_x}\right)\left\{\gamma_x^2\mu_{yi3}\kappa_i^{*\frac{1}{2}} + \gamma_y^2\mu_{xi3}\left(1+2\kappa_i^*\right)\kappa_i^{*-\frac{3}{2}}\right\}, \\
&\quad + \frac{\sigma_{pi}^{-6}}{4}\left(\mu_{xi}+\mu_{yi}\frac{\gamma_y}{\gamma_x}\right)(\mu_{xi}-\mu_{yi})\left\{\gamma_y^3\left(1+2\kappa_i^*\right)\kappa_i^{*-\frac{3}{2}}\left(\mu_{xi4}-\sigma_{xi}^4\right)\right. \\
&\quad \left. - \gamma_x^3\kappa_i^{*\frac{1}{2}}\left(\mu_{yi4}-\mu_{yi}^4\right)\right\}, \\
\sigma_i^2(t_2,t_2) &= 1 + \sigma_{pi}^{-4}\left(\mu_{xi}+\mu_{yi}\frac{\gamma_y}{\gamma_x}\right)\left\{\kappa_i^*\gamma_x^2\mu_{yi3} - \left(1+2\kappa_i^*\right)\kappa_i^{*-2}\gamma_y^2\mu_{xi3}\right\} \\
&\quad + \frac{\sigma_{pi}^{-6}}{4}\left(\mu_{xi}+\mu_{yi}\frac{\gamma_y}{\gamma_x}\right)^2\left\{\gamma_x^3\kappa_i^*\left(\mu_{yi4}-\sigma_{yi}^4\right) + \gamma_y^3\left(1+2\kappa_i^*\right)^2\kappa_i^{*-3}\left(\mu_{xi4}-\sigma_{xi}^4\right)\right\}.
\end{aligned}$$

The off-diagonal terms degenerate to zero when (i) the null hypothesis $\mu_{xi} = \mu_{yi}$ is true and (ii) the distributions are symmetric, i.e. $\mu_{y3i} = \mu_{x3i} = 0$, completing the proof. □

A.7. Proof of Proposition 7

The proposition can be considered as a special case of the theory in Section D of Basu et al. (2017) and can be proved similarly. As the proofs in Basu et al. (2017) are done in different settings with the weighted FDR definition and random weights, we provide the proof here for completeness.

Proof. **Part (a).** Let $\mathcal{X} = m^{-1}\sum_{i=1}^m(1-\theta_i)\delta_i$. Note that when $\mathcal{Y} = 0$ we must have $\mathcal{X} = 0$. The asymptotic equivalence follows if we can show the following

$$\mathrm{mFDR}(\boldsymbol{\delta}) - \mathrm{FDR}(\boldsymbol{\delta}) \le \mathbb{E}\left\{\left|\frac{\mathcal{X}}{\mathcal{Y}} - \frac{\mathcal{X}}{\mathbb{E}\mathcal{Y}}\right|\mathbb{I}(\mathcal{Y}>0)\right\} = \mathrm{o}(1). \tag{A.3}$$

Since $\mathcal{X} \le \mathcal{Y}$ and both are non-negative expressions, using Cauchy-Schwarz

$$\begin{aligned}
\mathbb{E}\left\{\left|\frac{\mathcal{X}}{\mathcal{Y}} - \frac{\mathcal{X}}{\mathbb{E}\mathcal{Y}}\right|\mathbb{I}(\mathcal{Y}>0)\right\} &= \mathbb{E}\left\{\frac{\mathcal{X}}{\mathcal{Y}}\mathbb{I}(\mathcal{Y}>0)\frac{|\mathcal{Y}-\mathbb{E}\mathcal{Y}|}{\mathbb{E}\mathcal{Y}}\right\} \\
&\le \frac{\left(\mathbb{E}\left|\mathcal{Y}-\mathbb{E}\mathcal{Y}\right|^2\right)^{1/2}}{\mathbb{E}\mathcal{Y}} \\
&= \frac{\mathrm{Var}\left(\mathcal{Y}\right)^{1/2}}{\mathbb{E}\mathcal{Y}},
\end{aligned}$$

The desired result follows from Conditions (i) and (ii).

Part (b). According to the proof of Theorem 2, Part (a), $\mathbb{P}\left(T_{OR}^{\tau,i} < Q_C^{\tau,-1}(\alpha)\right)$ is bounded away from zero, then there exists $\widetilde{p_\alpha} > 0$, such that $\mathbb{P}\left(T_{OR}^{\tau,i} < Q_C^{\tau,-1}(\alpha)\right) \geq \widetilde{p_\alpha}$. Therefore,

$$\mathbb{P}\left(\hat{T}_{OR}^{\tau,i} < \hat{Q}_C^{\tau,-1}(\alpha)\right) = \mathbb{P}\left(T_{OR}^{\tau,i} < Q_C^{\tau,-1}(\alpha)\right) + \mathrm{o}(1),$$
$$m^{-1}\mathbb{E}\left\{\sum_{i=1}^{m} \mathbb{I}\left(T_{OR}^{\tau,i} < Q_C^{\tau,-1}(\alpha)\right)\right\} = \mathbb{P}\left(T_{OR}^{\tau,i} < Q_C^{\tau,-1}(\alpha)\right) \geq \widetilde{p_\alpha} > 0.$$

Let $\mathcal{Y}' = m^{-1}\sum_{i=1}^{m} \mathbb{I}(T_{OR}^{\tau,i} \leq t_\infty^\tau)$. Note that

$$\mathrm{Var}(\mathcal{Y}') = m^{-1}\mathrm{Var}\left\{\mathbb{I}\left(T_{OR}^{\tau,i} \leq t_\infty\right)\right\} \leq m^{-1} = \mathrm{o}(1).$$

To show that $\mathrm{Var}(\mathcal{Y}) = \mathrm{o}(1)$, use the decomposition

$$\mathrm{Var}(\mathcal{Y}) = \mathrm{Var}(\mathcal{Y}') + \mathrm{Var}(\mathcal{Y} - \mathcal{Y}') + 2\mathrm{Cov}(\mathcal{Y} - \mathcal{Y}', \mathcal{Y}').$$

We only need to show that $\mathrm{Var}(\mathcal{Y} - \mathcal{Y}') = \mathrm{o}(1)$. Then by Cauchy-Schwarz inequality and using $\mathrm{Var}(\mathcal{Y}') = \mathrm{o}(1)$, it follows that

$$\mathrm{Cov}(\mathcal{Y} - \mathcal{Y}', \mathcal{Y}') = \mathrm{o}(1).$$

Recall that $\hat{T}_{OR}^\tau - T_{OR}^\tau = \mathrm{o}_P(1)$ and $\hat{Q}_C^{\tau,-1}(\alpha) - t_\infty^\tau = \mathrm{o}_P(1)$. We have

$$\begin{aligned}
\mathrm{Var}(\mathcal{Y} - \mathcal{Y}') =& m^{-2}\mathrm{Var}\left[\sum_{i=1}^{m}\left\{\mathbb{I}\left(\hat{T}_{OR}^{\tau,i} < \hat{Q}_C^{\tau,-1}(\alpha)\right) - \mathbb{I}\left(T_{OR}^{\tau,i} < t_\infty^\tau\right)\right\}\right] \\
\leq& m^{-2}\mathbb{E}\left[\sum_{i=1}^{m}\left\{\mathbb{I}\left(\hat{T}_{OR}^{\tau,i} < \hat{Q}_C^{\tau,-1}(\alpha)\right) - \mathbb{I}\left(T_{OR}^{\tau,i} < t_\infty^\tau\right)\right\}\right]^2 \\
=& \left(1 - \frac{1}{m}\right)\mathbb{E}\left[\prod_{k=i,j;i\neq j}\left\{\mathbb{I}\left(\hat{T}_{OR}^{\tau,k} < \hat{Q}_C^{\tau,-1}(\alpha)\right) - \mathbb{I}\left(T_{OR}^{\tau,k} < t_\infty^\tau\right)\right\}\right] \\
&+ \frac{1}{m}\mathbb{E}\left\{\mathbb{I}\left(\hat{T}_{OR}^\tau < \hat{Q}_C^{\tau,-1}\right) - \mathbb{I}\left(T_{OR}^\tau < t_\infty^\tau\right)\right\}^2 = \mathrm{o}(1).
\end{aligned}$$

The last equality follows since

$$\begin{aligned}
&\mathbb{E}\left[\prod_{k=i,j}\left\{\mathbb{I}\left(\hat{T}_{OR}^{\tau,k} < \hat{Q}_C^{\tau,-1}(\alpha)\right) - \mathbb{I}\left(T_{OR}^{\tau,k} < t_\infty^\tau\right)\right\}\right] \\
\leq& \mathbb{E}\left[\left\{\mathbb{I}\left(\hat{T}_{OR}^{\tau,i} < \hat{Q}_C^{\tau,-1}(\alpha)\right) - \mathbb{I}\left(T_{OR}^{\tau,i} < t_\infty^\tau\right)\right\}\right] = o(1).
\end{aligned}$$

A.8. Effects of baseline removal (Remark 1)

This section explains that removing the baseline effects would not violate our assumption on conditional independence. To fix ideas, we consider the application context of microarray time-course experiments.

Let $\boldsymbol{\xi}$ denote the true baseline expression levels of m genes. Consider two time points x and y. The (unknown) true effect sizes at these two time points are given by

$$\boldsymbol{\mu}_x = \boldsymbol{\xi} + \boldsymbol{\eta}_x, \quad \boldsymbol{\mu}_y = \boldsymbol{\xi} + \boldsymbol{\eta}_y.$$

Note that we do not require that $\boldsymbol{\xi}$ is sparse but some baseline measurements should be available. By contrast, we assume that $\boldsymbol{\eta}_x = (\eta_{xi} : i = 1, \cdots, m)$ and $\boldsymbol{\eta}_y = (\eta_{yi} : i = 1, \cdots, m)$ are sparse vectors representing random perturbation effects from based line levels $\boldsymbol{\xi}$. The goal is to identify nonzero effects $\eta_{xi} - \eta_{yi} \neq 0$.

We observe $X_i = \mu_{xi} + \epsilon_{xi}$ and $Y_i = \mu_{yi} + \epsilon_{yi}$, with $\epsilon_{xi} \sim N(0, 1)$ and $\epsilon_{yi} \sim N(0, 1)$. Moreover, we do not observe the true latent variable $\boldsymbol{\xi}$. The baseline effects would be measured (at time 0) with some errors $Z_i = \xi_i + \epsilon_{zi}$. We assume that ϵ_{xi}, ϵ_{yi} and ϵ_{zi} are independent with each other. After removing the baseline effects, we have new observations:

$$\begin{aligned} \tilde{X}_i &= \xi_i + \eta_{xi} + \epsilon_{xi} - (\xi_i + \epsilon_{zi}); \\ \tilde{Y}_i &= \xi_i + \eta_{yi} + \epsilon_{yi} - (\xi_i + \epsilon_{zi}). \end{aligned}$$

The pairs of differences and sums are given by

$$\begin{aligned} T_{1i} &= \tilde{X}_i - \tilde{Y}_i = (\eta_{xi} - \eta_{yi}) + (\epsilon_{xi} - \epsilon_{yi}); \\ T_{2i} &= \tilde{X}_i + \tilde{Y}_i = (\eta_{xi} + \eta_{yi}) + (\epsilon_{xi} + \epsilon_{yi}) - 2(\xi_i + \epsilon_{zi}). \end{aligned}$$

Under the null hypothesis $\eta_{xi} = \eta_{yi}$, the error terms $\epsilon_{xi} - \epsilon_{yi}$ and $\epsilon_{xi} + \epsilon_{yi}$ are independent, and $\epsilon_{xi} - \epsilon_{yi}$ is independent of $\xi_i + \epsilon_{zi}$. Hence T_{1i} and T_{2i} satisfy the conditional independence assumption (2.5), which further leads to Equation (2.6) that is needed in the CARS methodology.

A.9. Proof of the claim in Remark 3

Proof. We study the special case where (i) the observations have equal sample size and (known) equal variance $\sigma_{xi}^2 = \sigma_{yi}^2 = \sigma_i^2$, and (ii) the two nonzero means have the same magnitude with opposite signs $\mu_{xi} = -\mu_{yi}$. Then

$$(T_{1i}, T_{2i}) = \frac{\sqrt{n}}{2\sigma_i}(\bar{X}_i - \bar{Y}_i, \bar{X}_i + \bar{Y}_i).$$

Next, note that $\bar{X}_i = \mu_{xi} + \bar{\epsilon}_{xi}, \bar{Y}_i = -\mu_{xi} + \bar{\epsilon}_{yi}$, where $\bar{\epsilon}_{xi} \sim N(0, (2/n)\sigma_i^2)$ and $\bar{\epsilon}_{yi} \sim N(0, (2/n)\sigma_i^2)$. We have

$$\begin{aligned} f(t_1, t_2) &= \int f(t_1, t_2 | \mu_x = \mu) dG_{\mu_x}(\mu) \\ &= \int f(t_1 | \mu_x = \mu) f_{2\cdot}(t_2) dG_{\mu_x}(\mu) \\ &= f_{1\cdot}(t_1) f_{2\cdot}(t_2); \\ f(t_2 | \theta_{1i} = 0) &= f(t_2 | \mu_{xi} = \mu_{yi} = 0) = f_{2\cdot}(t_2). \end{aligned}$$

It follows that the oracle statistic

$$T_{OR}(t_1, t_2) = \frac{(1 - \pi_1) f(t_2 | \theta_{1i} = 0) f_{10}(t_1)}{f(t_1, t_2)} = \frac{(1 - \pi_1) f_{10}(t_1)}{f_{1\cdot}(t_1)},$$

which gives the Lfdr defined in (3.1).

A.10. Proof of insufficiency and further explanation of Remark 7

We prove the claim in Remark 7 that the primary statistic is not a sufficient statistic. The proof uses the notations in Section 4. We only need to provide a counter example. Consider the following special case where the grouping variable S_i takes two possible values 1 or 2 with equal probability:

$$P(S_i = 1) = 0.5, \quad P(S_i = 2) = 0.5.$$

If $S_i = j$, $j = 1, 2$, then T_i follows a mixture distribution

$$T_i \sim (1 - \pi_j)N(0, 1) + \pi_j N(1, 1),$$

where π_j is the conditional proportion of non-null cases for group j. An equivalent way to conceptualize the above model is to introduce an unknown location parameter μ_i: $T_i|\mu_i \sim N(\mu_i, 1)$, where μ_i follows a distribution with two point masses 0 and 1:

$$\mu_i \sim (1 - \pi_j)\delta_0 + \pi_j\delta_1.$$

Here δ_0 and δ_1 are the Dirac delta functions centered at 0 and 1, respectively.

Now we prove T_i's insufficiency. By definition, if T_i is sufficient, then $P(S_i = 1|T_i = t, \mu_i)$ should not depend on μ_i. However, some simple algebras give the following:

$$P(S_i = 1|T_i = t, \mu = 0) = \frac{1 - \pi_1}{2 - \pi_1 - \pi_2}, \quad P(S_i = 1|T_i = t, \mu = 1) = \frac{\pi_1}{\pi_1 + \pi_2},$$

and the desired result follows.

REMARK 8. The ancilarity paradox means that S_i seems to be "useless" for inference of μ_i since it is an external variable and the location information should have been captured by T_i. However, making inference using T_i alone would lead to substantial loss of information. In this example, S_i is called an *ancillary component* because (i) T_i alone is insufficient; (ii) S_i is ancillary; and (iii) (T_i, S_i) is sufficient.

A.11. The conservative bivariate density estimator

We first present a formula that gives an equivalent expression for the bivariate density function:

$$\begin{aligned} f(t_{1i}, t_{2i}) &= (1 - \pi_2)f(t_{1i}, t_{2i}|\theta_{2i} = 0) + \pi_2 f(t_{1i}, t_{2i}|\theta_{2i} = 1) \\ &= (1 - \pi_2)f_{10}(t_{1i})f_{20}(t_{2i}) + f(t_{1i}, t_{2i}, \theta_{2i} = 1), \end{aligned}$$

where the second equality follows from the logical relationship (2.2) and the conditional independence between T_{1i} and T_{2i} under the null (Equation 2.6). The estimation of $f(t_{1i}, t_{2i}, \theta_{2i} = 1)$ requires the knowledge of θ_{2i}, which is unknown in practice. We propose to estimate θ_{2i} via a screening procedure $\theta_{2i}^{\upsilon} = \mathbb{I}(R_i < \upsilon)$, where R_i is a screening statistic such as the Lfdr statistic or the p-value and υ is a tuning parameter. The R package uses $\text{Lfdr}_2(t_{2i})$ as the screening statistic with the default screening level $\upsilon = 0.1$.

Let $\mathcal{A}_\upsilon = \{(t_1, t_2) : R_i(t_2) < \upsilon\}$. Define the following test statistic

$$T_{OR}^{\upsilon,i}(t_1, t_2) = q^*(t_2)f_{10}(t_1)/f^{\upsilon}(t_1, t_2), \tag{A.4}$$

where $f^{\upsilon}(t_1, t_2) = (1-\pi_2)f_{10}(t_1)f_{20}(t_2)+\{G(\upsilon)-(1-\pi_2)G_0(\upsilon)\}f(t_1, t_2|\theta_2^{\upsilon} = 1)$. Consider the following index set $\mathcal{W} = \{i : T_{OR}^i < 1 - \pi_2\}$. This set can be viewed as the collection

of "interesting" hypotheses that are worthy of further investigation. We shall study the property of $T_{OR}^{i,\upsilon}$ on $\mathcal{W}$ because hypotheses not in $\mathcal{W}$ would not have any impact in the multiple testing problem – observe that π_2 is small and large T_{OR}^{i} would not play a role in inference according to the operation of Procedure 3.6.

The next proposition implies that the modified density estimator would lead to a conservative FDR control.

PROPOSITION 8. *Consider* $T_{OR}^{\upsilon,i}$ *defined in* (A.4). *If* $i \in \mathcal{W}$, *then* $T_{OR}^{i} \leq T_{OR}^{\upsilon,i}$.

Proof. We first show that if $i \in \mathcal{W}$, then $f(t_{1i}, t_{2i}) \leq f(t_{1i}, t_{2i}|\theta_{2i} = 1)$. Note that

$$\mathbb{P}(\theta_{2i} = 0|T_{1i} = t_{1i}, T_{2i} = t_{2i}) = \frac{(1-\pi_2)f(t_{1i}, t_{2i}|\theta_{2i} = 0)}{(1-\pi_2)f(t_{1i}, t_{2i}|\theta_{2i} = 0) + \pi_1 f(t_{1i}, t_{2i}|\theta_{2i} = 1)}.$$

If $i \in \mathcal{W}$, it follows from the logical relationship (2.2) that

$$\mathbb{P}(\theta_{2i} = 0|T_{1i} = t_{1i}, T_{2i} = t_{2i}) = \mathbb{P}(\theta_{1i} = 0, \theta_{2i} = 0|T_{1i} = t_{1i}, T_{2i} = t_{2i}) \leq T_{OR}^{i} < 1 - \pi_2.$$

Therefore $f(t_{1i}, t_{2i}|\theta_{2i} = 0) \leq f(t_{1i}, t_{2i}|\theta_{2i} = 1)$ for $i \in \mathcal{W}$. Without causing ambiguity, we focus on a generic hypothesis $i \in \mathcal{W}$ and drop the index i. It follows that $f(t_1, t_2) = (1-\pi_2)f(t_1, t_2|\theta_2 = 0) + \pi_2 f(t_1, t_2|\theta_2 = 1) \leq f(t_1, t_2|\theta_2 = 1)$. Hence

$$\begin{aligned} f(t_1, t_2, \theta_2 = 1) &\geq f(t_1, t_2, \theta_2 = 1, \theta_2^{\upsilon} = 1) \\ &= P(\theta_2 = 1, \theta_2^{\upsilon} = 1) f(t_1, t_2|\theta_2 = 1, \theta_2^{\upsilon} = 1) \\ &\geq C_{\upsilon} f(t_1, t_2|\theta_2^{\upsilon} = 1), \end{aligned}$$

where $C_{\upsilon} = G(\upsilon) - (1-\pi_2)G_0(\upsilon)$, $G(\upsilon) = \mathbb{P}(R_i < \upsilon)$ and $G_0(\upsilon) = \mathbb{P}(R_i < \upsilon|\theta_{2i} = 0)$. The last equality follows from the following two facts

$$\begin{aligned} \mathbb{P}(\theta_2^{\upsilon} = 1, \theta_2 = 1) &= P(\theta_2^{\upsilon} = 1) - P(\theta_2^{\upsilon} = 1, \theta_2 = 0), \quad \text{and} \\ f(t_1, t_2|\theta_2 = 1, \theta_2^{\upsilon} = 1) &= f(t_1, t_2|\theta_2 = 1)\mathbb{I}_{\mathcal{A}_{\upsilon}}/P(\mathcal{A}_{\upsilon}) \\ &\geq f(t_1, t_2)\mathbb{I}_{\mathcal{A}_{\upsilon}}/P(\mathcal{A}_{\upsilon}) = f(t_1, t_2|\theta_2^{\upsilon} = 1). \end{aligned}$$

Therefore, we conclude that $f^{\upsilon}(t_1, t_2) \leq f(t_1, t_2)$ and the desired result follows. □

Finally, we discuss how to estimate $f^{\upsilon}(t_1, t_2)$. The first term $(1-\pi_2)f_{10}(t_1)f_{20}(t_2)$ can be estimated by plugging in $\hat{\pi}_2$; the null densities f_{10} and f_{20} are known. The second term can be rewritten as $G(\upsilon)\{1 - Q(\upsilon)\} f(t_1, t_2|\theta_2^{\upsilon} = 1)$, where $Q(\upsilon) = (1-\pi_2)G_0(\upsilon)/G(\upsilon)$ can be viewed as the mFDR level of $I(R_i < \upsilon)$ for testing hypothesis $\theta_{2i} = 0$. In practice, $G(\cdot)$ can be estimated by the empirical CDF $\mathbb{G}(\upsilon) = m^{-1}\sum_{i=1}^{m} \mathbb{I}(R_i < \upsilon)$. To estimate $f(t_1, t_2|\theta_{2i}^{\upsilon} = 1)$, we apply the R package `ash` to the sample $\mathcal{T} = \{i : R_i < \upsilon\}$. [The `ash` package uses the average shifted histogram, which seems to, according to our numerical studies, provide better estimation than the more standard package `np`.] The `ash` package only calculates the density on grids, we further use the function *interp_surface* in the R package `fields` to interpolate the estimated density. When the Lfdr is used for screening, $Q(\upsilon)$ can be estimated as [cf. Sun and Cai (2007)]

$$\hat{Q}(\upsilon) = \frac{\sum_{i=1}^{m} \mathbb{I}\{\widehat{\text{Lfdr}}_2(t_{2i}) < \upsilon\}\widehat{\text{Lfdr}}_2(t_{2i})}{\sum_{i=1}^{m} \mathbb{I}(\widehat{\text{Lfdr}}_2(t_{2i}) < \upsilon)}.$$

When the p-value is used for screening, we use the well known estimate $\hat{Q}(\upsilon) = \upsilon/\mathbb{G}(\upsilon)$ [cf. Benjamini and Hochberg (2000); Genovese and Wasserman (2002)].

A.12. Screening via Lfdr: formulae and implementation

Consider a screening procedure based on $\zeta_i < \tau$, where $\zeta_i \coloneqq \zeta_i(t_{1i})$ is a *screening* statistic and $\tau > 0$ is a tuning parameter. Denote $G(\tau)$ the CDF of ζ_i. Let $\mathcal{A}_\tau = \{x : \zeta(x) > \tau\}$ and $S(\tau) = P(\mathcal{A}_\tau)$ the survival function. Then

$$q^\tau(t_2) = \lim_{h\to 0} \frac{\mathbb{E}\{Q^\tau(t_2, h)\}}{h} = S(\tau)^{-1} \int_{\mathcal{A}_\tau} f(t_1, t_2) dt_1, \tag{A.5}$$

which can be estimated by

$$\hat{q}^\tau(t_2) = \frac{\sum_{i\in\mathcal{T}(\tau)} K_{h_2}(t_2 - t_{2i})}{m\hat{S}(\tau)^{-1}}. \tag{A.6}$$

In the p-value approach, we have an explicit correction factor of $1 - \tau$. The factor is now replaced by $\hat{S}(\tau)$. If the Lfdr is used for screening, then a new correction factor would be needed. We can simulate B data points (e.g. $B = 10^5$) from the null distribution $N(0, 1)$ and calculate their Lfdrs using the method in Sun and Cai (2007), where the parameters are estimated using the primary statistics $\{t_{1i} : 1 \le i \le m\}$. The correction factor is then obtained as the proportion of Lfdrs exceeding τ.

B. Supplementary Numerical Results

B.1. The case with unknown and equal variances

Let $\epsilon_{xij} \sim N(0,1)$ and $\epsilon_{yik} \sim N(0,1)$. Set $n_x = 50, n_y = 60$. Therefore our $\kappa = \frac{n_y}{n_x} = 1.2$. We compute T_{1i} and T_{2i} using estimated variances. The following settings are considered in our simulation study.

Setting 1: we set $\boldsymbol{\mu}_{x,1:k} = 5/\sqrt{30}$, $\boldsymbol{\mu}_{x,(k+1):(2k)} = 4/\sqrt{30}$, $\boldsymbol{\mu}_{x,(2k+1):m} = 0$, $\boldsymbol{\mu}_{y,1:k} = 2/\sqrt{30}$, $\boldsymbol{\mu}_{y,(k+1):(2k)} = 4/\sqrt{30}$ and $\boldsymbol{\mu}_{y,(2k+1):m} = 0$. We vary k from 100 to 1000.

Setting 2: we use k_1 and k_2 to denote the number of nonzero locations and the number of locations with differential effects, respectively. Let $\boldsymbol{\mu}_{x,1:k_2} = 5/\sqrt{30}$, $\boldsymbol{\mu}_{x,(k_2+1):k_1} = 4/\sqrt{30}$, $\boldsymbol{\mu}_{x,(k_1+1):m} = 0$, $\boldsymbol{\mu}_{y,1:k_2} = 2/\sqrt{30}$, $\boldsymbol{\mu}_{y,(k_2+1):k_1} = 4/\sqrt{30}$ and $\boldsymbol{\mu}_{y,(k_1+1):m} = 0$. We fix $k_1 = 2000$ and vary k_2 from 200 to k_1.

Setting 3: the sparsity level is fixed at $k = 750$. We set $\boldsymbol{\mu}_{x,1:(k)} = \mu_0/\sqrt{30}$, $\boldsymbol{\mu}_{x,(k+1):(2k)} = 3/\sqrt{30}$, $\boldsymbol{\mu}_{x,(2k+1):m} = 0$, $\boldsymbol{\mu}_{y,1:k} = 2/\sqrt{30}$, and $\boldsymbol{\mu}_{y,(k+1):(2k)} = 3/\sqrt{30}$, $\boldsymbol{\mu}_{y,(2k+1):m} = 0$. We vary μ_0 from 3.5 to 5.

In Settings 1-3, we apply different methods to the simulated data and obtain results by averaging over 500 replications. We plot the FDR and Average Power levels as functions of varied parameter values. The results are displayed in Figure 6. We can similarly see that both CARS and US are more powerful than conventional univariate methods such BH and AZ, and CARS is superior than US.

B.2. Comparison with sample splitting when $\theta_{1i} = \theta_{2i}$

Similar to the general setting, we set $n_x = 50, n_y = 60$ and construct the primary and auxiliary statistics accordingly. We consider two simulation settings.

Setting 1: $\boldsymbol{\mu}_{x,1:k} = 5/\sqrt{30}$, $\boldsymbol{\mu}_{x,(k+1):m} = 0$, $\boldsymbol{\mu}_{y,1:k} = 2/\sqrt{30}$ and $\boldsymbol{\mu}_{y,(k+1):m} = 0$. Vary k from 50 to 1000 to investigate the effect of sparsity on testing results.

Setting 2: the sparsity level is fixed at $k = 500$, $\boldsymbol{\mu}_{y,1:m} = 0$, $\boldsymbol{\mu}_{z,1:k} = \mu_0/\sqrt{30}$, and $\boldsymbol{\mu}_{z,(k+1):m} = 0$. Vary μ_0 from 1.5 to 3.5 to investigate the impact of effect sizes.

The sample splitting (SS) method has been added to the comparison. Specifically, SS splits both n_x and n_y into two equal parts. The first half of data are used to compute the initial p-values. We then use the p-value cut-off 0.05 to select locations and use these locations for the second stage analysis. Next the second stage p-values are computed for the selected locations using the second half of the data. Finally we apply the BH procedure to the second stage p-values and record the decisions. The results are displayed in Figure 7. The following observations can be made based on the simulation.

(a). All methods can control the FDR at the nominal level 0.05, with BH slightly conservative and US very conservative.

(b). DD can be very conservative in many settings but still enjoys substantial power gain over competing methods.

(c). Univariate methods (BH and AZ) can be greatly improved by US, which is further improved by CARS.

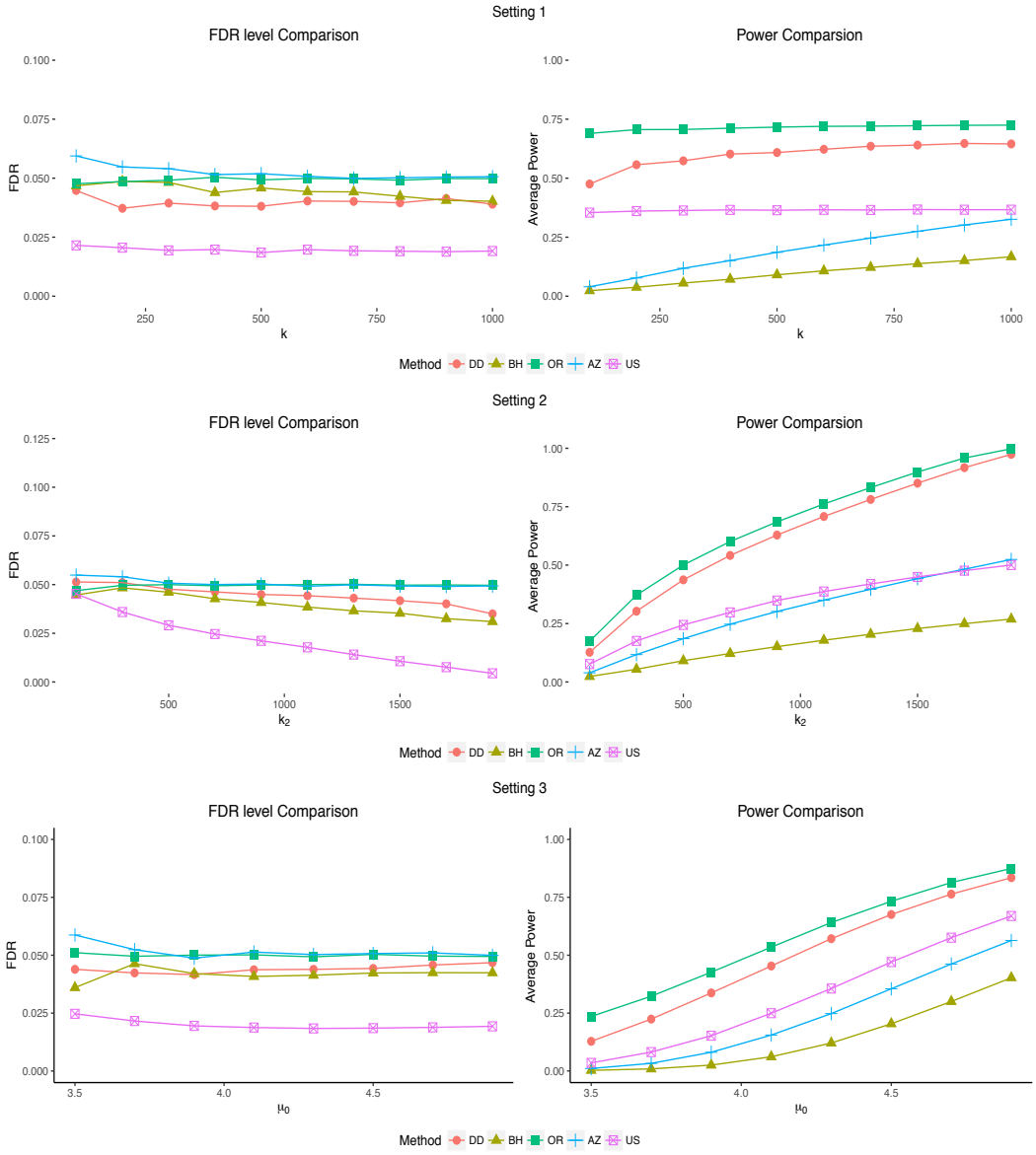

Fig. 6. General Case (Unknown Variance): the FDR and MDR varied by non-null proportion (top row) and conditional proportion (bottom row). The displayed procedures are DD (○), BH (△), OR (■), AZ (+) and US (□).

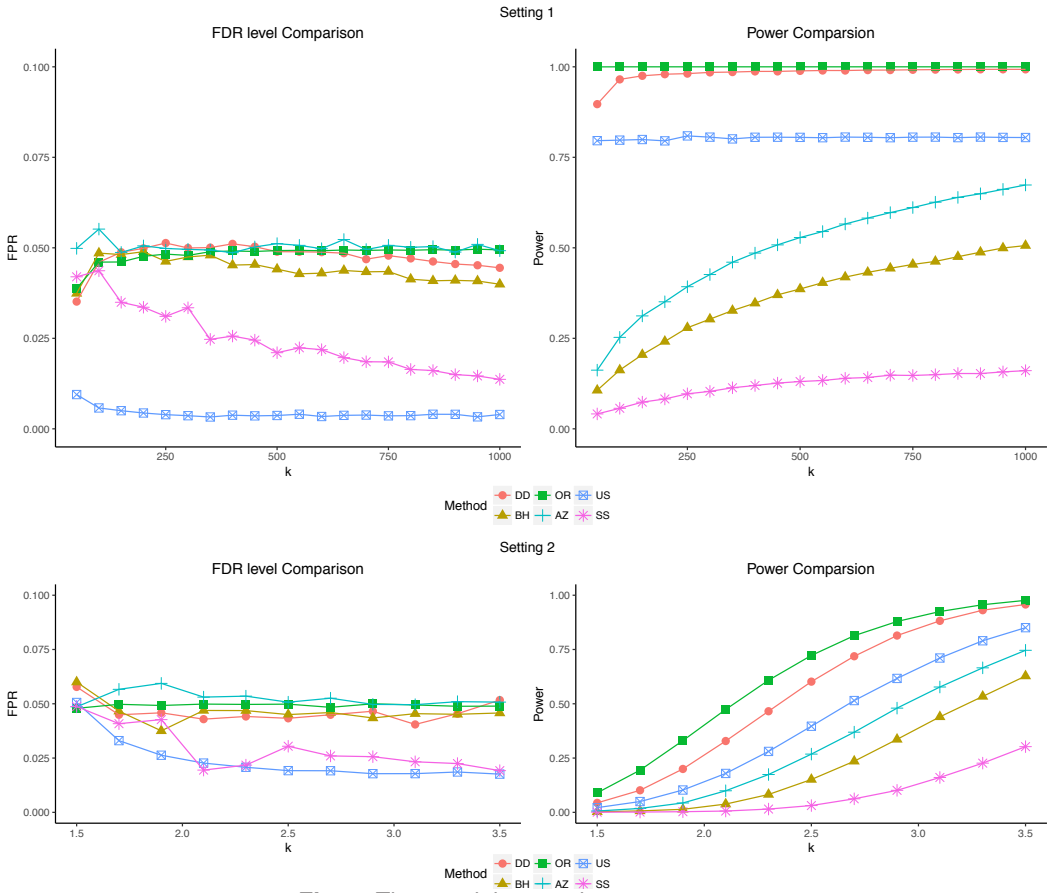

Fig. 7. The special case when $\theta_{1i} = \theta_{2i}$.

(d). Setting 1 shows that the gain in efficiency decreases as k increases.

(e). Setting 2 shows that the efficiency gain of CARS increases as k_2 increases.

(f). The sample splitting method is inefficient.

B.3. Stability of tuning

This section investigates the robustness of tuning. Let $\mathcal{T}(\tau) = \{i : \text{Lfdr}_1(t_{1i}) > \tau\}$. We first examine the performance of CARS wrt the choice of τ by focusing on the case when θ_{1i} and θ_{2i} are perfectly correlated. The following settings are considered:

Setting 1: same as Setting 1 in Section 5.2, except that $k = 500$. Vary τ from 0.6 to 0.9.

Setting 2: same as Setting 1 above, except that $\boldsymbol{\mu}_{x,1:500} = 4/\sqrt{50}$. Vary τ from 0.6 to 0.9.

Setting 3: same as Setting 1, except that $k = 1,000$. Vary τ from 0.6 to 0.9.

Setting 4: same as the Setting 3, except that $\boldsymbol{\mu}_{x,1:1000} = 4/\sqrt{50}$. Vary τ from 0.6 to 0.9.

We apply the oracle procedure (OR) and the data-driven CARS procedure (DD) to the simulated data and record the FDR, power and size of $\hat{\mathcal{S}} = \{\widehat{\text{Lfdr}}_1 > \tau\}$. Note that for OR, we always have $\text{Card}(\mathcal{S}) = 4500$ or 4000. The results are summarized in Figure 9. We can see that the value of τ affects the size of $\hat{\mathcal{S}}$ and in general provides better error control when it is large. This is consistent with our theory in Proposition 4. However, to avoid inflated variability in estimated quantities, we do not recommend larger τ's. As a rule of thumb, our practical recommendation is to take $\tau = 0.9$ to ensure both precise error control and stability of the methodology.

Next we investigate the stability of CARS wrt the choice of υ in (A.4). We fix $\tau = 0.9$ and vary υ from 0.1 to 0.3. The following settings are considered:

Setting 1: same as Setting 1 in Section 5.2, except that $k = 50$.

Setting 2: same as Setting 1 above, except that $\boldsymbol{\mu}_{x,1:50} = 4/\sqrt{50}$.

Setting 3: same as Setting 1, except that $k = 100$.

Setting 4: same as the Setting 3, except that $\boldsymbol{\mu}_{x,1:100} = 4/\sqrt{50}$.

The results are summarized in Figure 9. We can see that the value of υ affects the size of $\hat{\mathcal{T}}$ but in general does not affect the FDR or power. As a rule of thumb, our practical recommendation is to take $\upsilon = 0.1$.

B.4. Comparison of ranking

A fundamental aspect of the multiple comparison issue is *ranking*. For example, if a biologist has a limited budget and asks us to provide a list of top k genes (say, $k = 10$). Then on top of FDR control, s/he may care more about how many true findings are actually in the list that we provide. We carry out a small simulation study to compare different methods when $k_2 = 20$. The goal is to show that, by exploiting the auxiliary information, CARS yields a much better ranking (in the sense that for a pre-specified budget, namely a fixed number of rejections, CARS identifies more true positives than competing methods).

Consider the case with the unknown and equal variances. Let $\epsilon_{xij} \sim N(0,1)$ and $\epsilon_{yik} \sim N(0,1)$. Set $n_x = 50, n_y = 60$. Therefore $\kappa = \frac{n_y}{n_x} = 1.2$. Number of repetitions is 100. The number of genes $m = 5000$. Setting: $\boldsymbol{\mu}_{x,1:20} = 5/\sqrt{50}$, $\boldsymbol{\mu}_{x,21:40} = 4/\sqrt{50}$, $\boldsymbol{\mu}_{x,41:m} = 0$, $\boldsymbol{\mu}_{y,1:20} = 2/\sqrt{50}$, $\boldsymbol{\mu}_{y,21:40} = 4/\sqrt{50}$, $\boldsymbol{\mu}_{y,41:m} = 0$. We vary number of rejections k from 5 to 20, and calculate for even given k, how many true positives are in the top k hypotheses. The rankings are based on p-values (BH), Lfdr statistics (Sun and Cai 2007), oracle statistic T_{OR} (OR, proposed with known parameters), and data-driven statistic $\hat{T}^{\tau}_{OR}$ (DD, proposed with estimated parameters), respectively. The expected number of true positives (ETP) is calculated based on the average of 100 simulated data sets. We can see from Figure 10 that the ranking of CARS is superior compared to other methods.

B.5. Dependent tests

This section presents a small simulation study to investigate the performance of CARS under dependence.

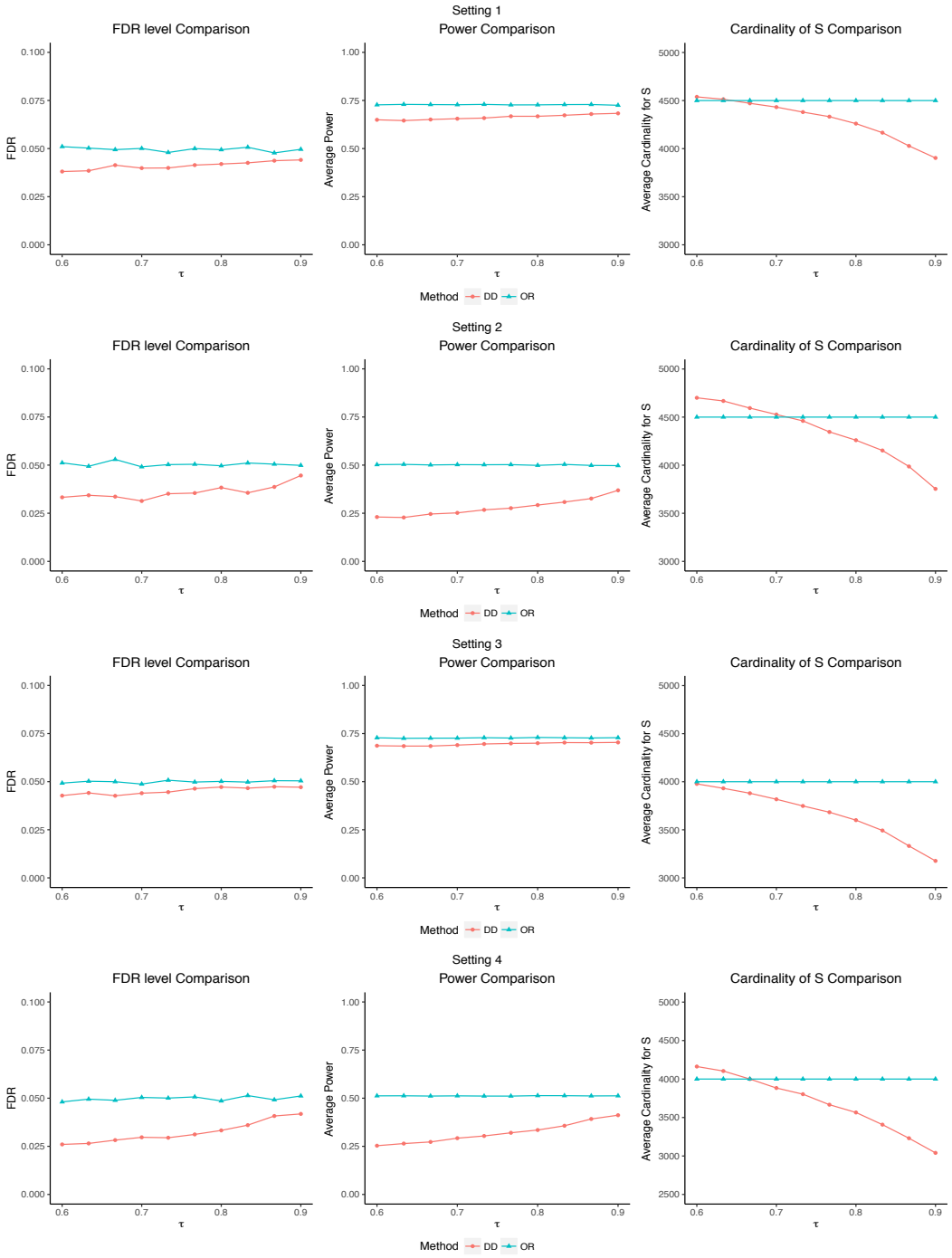

Fig. 8. Effect of τ. The displayed procedures are DD ($\circ$), OR ($\triangle$).

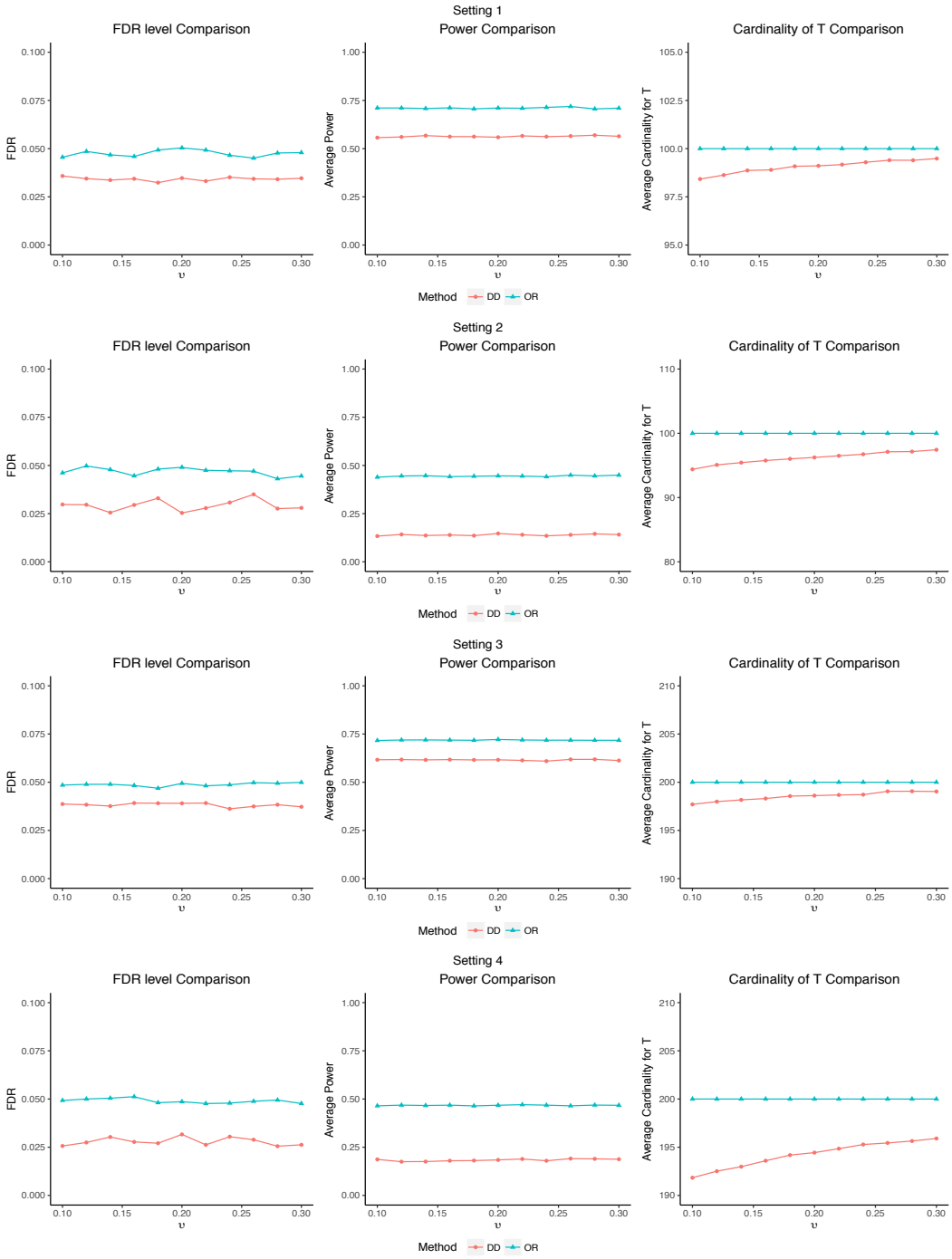

Fig. 9. Effect of υ. The displayed procedures are DD ($\circ$), OR ($\triangle$).

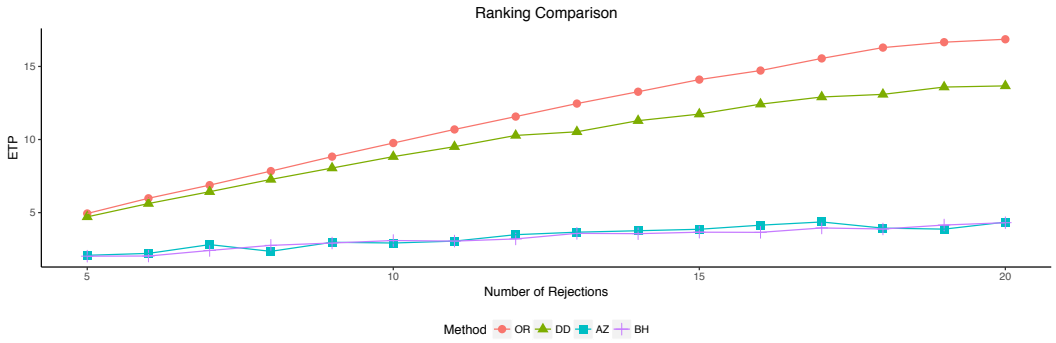

Fig. 10. Comparison of ranking

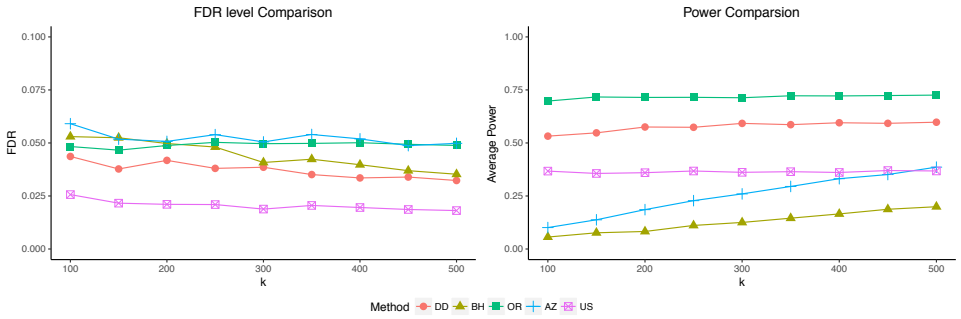

Fig. 11. Comparison of methods under weak dependence.

We demonstrate the performance of CARS under the setting with the unknown and equal variances case. Let Σ be the covariance matrix.

$$\begin{aligned}
\Sigma_{i,i} &= 1,\ i = 1, \cdots, m, \\
\Sigma_{i,i+1} &= 0.5,\ i = 1, \cdots, m-1, \\
\Sigma_{i-1,i} &= 0.5,\ i = 2, \cdots, m, \\
\Sigma_{i,i+2} &= 0.4,\ i = 1, \cdots, m-2, \\
\Sigma_{i-2,i} &= 0.4,\ i = 3, \cdots, m.
\end{aligned}$$

All other entries equal to zero. Let $\epsilon_{\boldsymbol{x}_j} \sim N(0, \Sigma)$ and $\epsilon_{\boldsymbol{y}_k} \sim N(0, \Sigma)$. Set $n_x = 50, n_y = 60$. Therefore $\kappa = \frac{n_y}{n_x} = 1.2$. Number of repetitions is 100. $m = 2000$.

Setting: We set $\boldsymbol{\mu}_{x,1:k} = 5/\sqrt{50}$, $\boldsymbol{\mu}_{x,(k+1):2k} = 4/\sqrt{50}$, $\boldsymbol{\mu}_{x,(2k+1):m} = 0$, $\boldsymbol{\mu}_{y,1:k} = 2/\sqrt{50}$, $\boldsymbol{\mu}_{y,(k+1):2k} = 4/\sqrt{50}$, $\boldsymbol{\mu}_{y,(2k+1):m} = 0$. We vary k from 100 to 500. The results are summarized by Figure 11. Both OR and DD seem to work well for FDR control. However, we want to emphasize that the simulation results here are very limited and we do not intend to draw any conclusions based on these results. Meanwhile, we conjecture that CARS would still be asymptotically valid under some weak dependence assumptions.

B.6. Supernova Example Images

Here we provide additional numerical results for the analysis in Section 5.5 in the main text. Figure 12 shows the rejected pixels in the 516×831 layout for each method under different FDR levels.

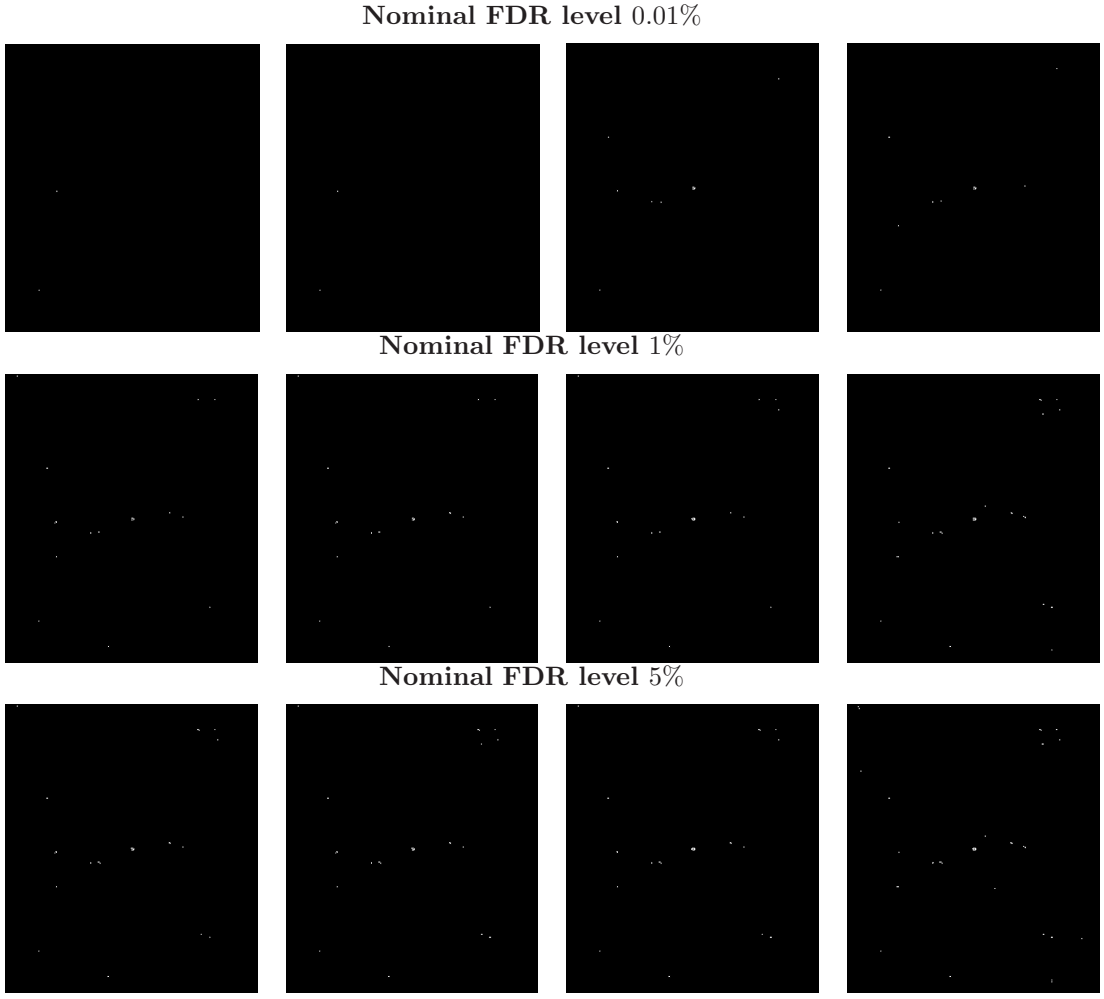

Fig. 12. Left to right: BH procedure, Adaptive Z procedure, Uncorrelated Screening, CARS.

B.7. Analysis of the Microarray Time-Course (MTC) Data

In this section we apply the CARS procedure to the MTC dataset collected by Calvano et al. (2005) for studying systemic inflammation in humans. This dataset contains eight study subjects which are randomly assigned to treatment and control groups and then administered with endotoxin and placebo, respectively. Affimetrix U133A chips were used to profile the expression levels of 22,283 genes in human leukocytes measured before infusion (0 hour) and at 2, 4, 6, 9, and 24 hours afterwards. One of the goals in this experiment

is to identify, in the treatment group, early to middle response genes that are differentially expressed within 4 hours and thus revealing meaningful early activation gene sequence which could potentially govern the immune responses.

We first preprocess the data according to the steps discussed in Sun and Wei (2011). To further identify the genes that regulate this sequence, we take time point 0 as the baseline and time point 4 and 24 as the interval over which differential gene expressions are estimated. Denote $Y_{j,i}$ as the average gene expression value for gene i at time point j. Let $\tilde{Y}_{j,i} = Y_{j,i} - Y_{0,i}$ denotes the baseline adjusted expression level for gene i at time point j. The primary and auxiliary statistics are

$$(T_{1i}, T_{2i}) = \sqrt{\frac{1}{2}} \left(\frac{\tilde{Y}_{4,i} - \tilde{Y}_{24,i}}{S_p}, \frac{\tilde{Y}_{4,i} + \hat{\kappa}^* \tilde{Y}_{24,i}}{\sqrt{\hat{\kappa}^*} S_p} \right)$$

where $m = 22,283$, $S^2_{y,4} = \mathrm{Var}(\tilde{Y}_{4,i})$, $S^2_{y,24} = \mathrm{Var}(\tilde{Y}_{24,i})$, $\hat{\kappa}^* = S^2_{y,4}/S^2_{y,24}$ and $S^2_p = S^2_{y,4} + S^2_{y,24}$. At FDR level 0.05, our CARS procedure discovered 429 differentially expressed genes among the total of $22,283$ genes. By contrast, BH procedure discovered 121 genes.

B.8. Comparison of mFDR and FDR

This section investigates the asymptotic equivalence of mFDR and FDR in different settings. In our simulation, we choose $m = 5000$, $n_x = 50$, $n_y = 60$, $\boldsymbol{\mu}_{x,1:k} = 5/\sqrt{30}$, $\boldsymbol{\mu}_{x,(k+1):(2k)} = 4/\sqrt{30}$, $\boldsymbol{\mu}_{x,(2k+1):m} = 0$ and $\boldsymbol{\mu}_{y,1:k} = 3/\sqrt{30}$, $\boldsymbol{\mu}_{y,(k+1):(2k)} = 4/\sqrt{30}$ and $\boldsymbol{\mu}_{y,(2k+1):m} = 0$. The sparsity level k is varied from 150 to 1000.

We apply the oracle CARS procedure (OR) and data-driven CARS procedure (DD) to the simulated data sets. The mFDR is computed as the ratio of the average number of false positives over the average number of rejections. The FDR is computed as the average of false discovery proportions. The results are obtained based on 200 replications and summarized by Figure 13. We can see that the FDR and mFDR levels are very similar for both the OR and DD methods.

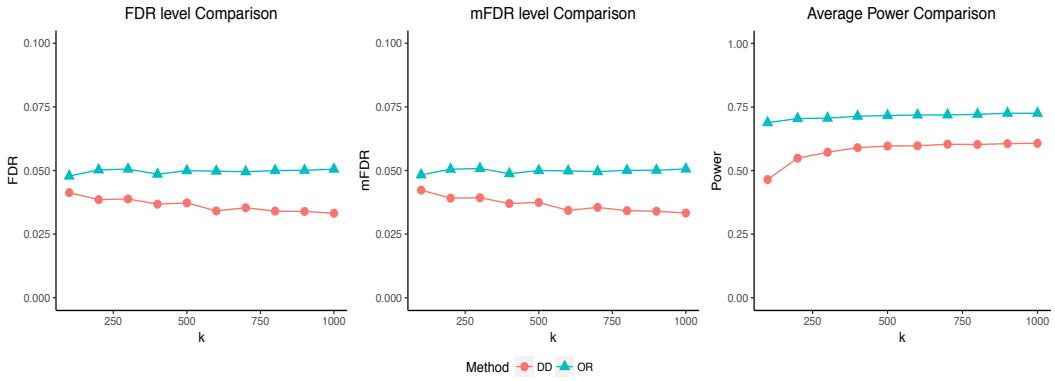

Fig. 13. Asymptotic equivalence of mFDR and FDR for OR and DD