Supplement to "Order-Constrained ROC Regression with Application to Facial Recognition"

This supplement is organized as follows. Sections A and B contain proofs of the two statements presented in the body of the paper. Some supplementary statements and their proofs are contained in Sections C and D. Section E is dedicated to additional simulation results.

A Proof of Proposition 1

We start by showing that

$$\widehat{\boldsymbol{\beta}}^{\mathrm{WLS}} \stackrel{\mathcal{D}}{=} \boldsymbol{\beta}^* + \frac{\sigma}{\sqrt{N}}\boldsymbol{\zeta}, \qquad \boldsymbol{\zeta} \sim N_d(\mathbf{0}, \boldsymbol{\Omega}), \qquad \boldsymbol{\Omega} = \begin{bmatrix} \frac{N}{n}\mathbf{C}_0^{-1} & -\frac{N}{n}\mathbf{C}_0^{-1} \\ -\frac{N}{n}\mathbf{C}_0^{-1} & \frac{N}{n}\mathbf{C}_0^{-1} + \frac{N}{m}\cdot\frac{\tau^2}{\sigma^2}\mathbf{C}_1^{-1} \end{bmatrix},$$

$$\mathbf{C}_0 = \tfrac{1}{n}[\mathbf{1}_n \; \mathbf{Z}_0]^\top[\mathbf{1}_n \; \mathbf{Z}_0], \; \mathbf{Z}_0 = [\mathbf{X}_{1\bullet}; \ldots; \mathbf{X}_{n\bullet}], \quad \mathbf{C}_1 = \tfrac{1}{m}[\mathbf{1}_m \; \mathbf{Z}_1]^\top[\mathbf{1}_m \; \mathbf{Z}_1], \; \mathbf{Z}_1 = [\mathbf{X}_{n+1\bullet}; \ldots; \mathbf{X}_{N\bullet}],$$

where ";" here denotes row-wise concatenation. Under model (16), $\widehat{\boldsymbol{\beta}}^{\mathrm{WLS}} \sim N_d(\boldsymbol{\beta}^*, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = (\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}$, and it remains to show that $\frac{N}{\sigma^2}\boldsymbol{\Sigma} = \boldsymbol{\Omega}$. Invoking relation (17), we obtain

$$\boldsymbol{\Sigma} = \mathbf{L}^{-1}\begin{bmatrix} \sigma^2\frac{1}{n}\mathbf{C}_0^{-1} & \mathbf{0} \\ \mathbf{0} & \tau^2\frac{1}{m}\mathbf{C}_1^{-1} \end{bmatrix}(\mathbf{L}^\top)^{-1}, \qquad \mathbf{L}^{-1} = \begin{pmatrix} \mathbf{I}_{p+1} & \mathbf{0} \\ -\mathbf{I}_{p+1} & \mathbf{I}_{p+1} \end{pmatrix}, \quad (\mathbf{L}^\top)^{-1} = (\mathbf{L}^{-1})^\top.$$

Performing the above matrix multiplication confirms that $\frac{N}{\sigma^2}\cdot\boldsymbol{\Sigma} = \boldsymbol{\Omega}$. To conclude the proof of the proposition, we make use of the primal-dual relation (12). According to the previous display, we use that $\frac{N}{\sigma^2}(\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1} = \boldsymbol{\Omega}$, and then re-write the minimizer of the dual problem:

$$
\begin{aligned}
\widehat{\boldsymbol{\lambda}} &= \underset{\boldsymbol{\lambda}\in\mathbb{R}^q_+}{\operatorname{argmin}} \left\{ \frac{1}{2}\boldsymbol{\lambda}^\top \mathbf{A}(\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}\mathbf{A}^\top\boldsymbol{\lambda} + \boldsymbol{\lambda}^\top\mathbf{A}\widehat{\boldsymbol{\beta}}^{\mathrm{WLS}} \right\} \\
&= \underset{\boldsymbol{\lambda}\in\mathbb{R}^q_+}{\operatorname{argmin}} \left\{ \frac{1}{2}\boldsymbol{\lambda}^\top \frac{\mathbf{A}(N^{-1}\sigma^2\mathbf{X}^\top\mathbf{W}\mathbf{X})^{-1}\mathbf{A}^\top}{N/\sigma}\boldsymbol{\lambda} + \boldsymbol{\lambda}^\top\left(\frac{\boldsymbol{\Delta}^*}{\sigma} + \frac{\boldsymbol{\xi}}{\sqrt{N}}\right) \right\}, \ \ \boldsymbol{\xi} \sim N(\mathbf{0}, \mathbf{A}\boldsymbol{\Omega}\mathbf{A}^\top) \\
&= \underset{\boldsymbol{\lambda}\in\mathbb{R}^q_+}{\operatorname{argmin}} \left\{ \frac{1}{2}\frac{\boldsymbol{\lambda}^\top\mathbf{A}\boldsymbol{\Omega}\mathbf{A}^\top}{N/\sigma}\boldsymbol{\lambda} + \boldsymbol{\lambda}^\top\left(\frac{\boldsymbol{\Delta}^*}{\sigma} + \frac{\boldsymbol{\xi}}{\sqrt{N}}\right) \right\}.
\end{aligned}
$$

Using Lemma 3 below yields that the "argmin" in (18) results as

$$
\underset{\boldsymbol{\lambda}\in\mathbb{R}^q_+}{\operatorname{argmin}} \left\{ \frac{1}{2}\boldsymbol{\lambda}^\top\mathbf{A}\boldsymbol{\Omega}\mathbf{A}^\top\boldsymbol{\lambda} + \boldsymbol{\lambda}^\top\left(\frac{\boldsymbol{\Delta}^*}{\sigma} + \frac{\boldsymbol{\xi}}{\sqrt{N}}\right) \right\} = \frac{\widehat{\boldsymbol{\lambda}}}{N/\sigma}.
$$

Putting together the pieces and simplifying terms yields the assertion.

B Proof of Corollary 1

In view of (11) and (12), it is clear that $\{\mathbf{A}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} \geq \mathbf{0}\} = \{\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{\mathrm{LS}}\}$. We have $\{\mathbf{A}\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} \geq \mathbf{0}\} = \bigcap_{k=1}^q \{\mathbf{a}_k^\top\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} \geq 0\}$. Using the Gaussian tail bound $P(Z < \mu - t) \leq \exp\left(-\frac{t^2}{2\sigma^2}\right)$, $t > 0$ for $Z \sim N(\mu, \sigma^2)$, the fact that $N^{1/2}\mathbf{a}_k^\top\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} \sim N(\Delta_k^*, \sigma^2\mathbf{a}_k^\top\boldsymbol{\Omega}\mathbf{a}_k)$, $k = 1, \ldots, q$, and a union bound, we obtain that

$$
P\left(\bigcup_{k=1}^q \left\{\mathbf{a}_k^\top\widehat{\boldsymbol{\beta}}^{\mathrm{LS}} < 0\right\}\right) \leq \sum_{k=1}^q \exp\left(-N \cdot \frac{\Delta_k^{*2}}{2\sigma^2 \cdot \mathbf{a}_k^\top\boldsymbol{\Omega}\mathbf{a}_k}\right).
$$

This concludes the proof of the corollary.

C Additional Corollary

Corollary 2. *Under the conditions of Proposition 1, suppose additionally that $p = 1$ and $\mathcal{X} = [0,1]$, i.e., X is a single continuous covariate with range $[0,1]$. We then have*

$$\widehat{\boldsymbol{\beta}} \overset{\mathcal{D}}{=} \boldsymbol{\beta}^* + \sigma\boldsymbol{\zeta} + \sigma\boldsymbol{\Omega}\mathbf{A}^\top \Bigg\{ \mathbf{G}^{-1}\boldsymbol{\nu} \cdot I\left(\mathbf{G}^{-1}\boldsymbol{\nu} \geq \mathbf{0}\right) + \begin{pmatrix} \frac{\nu_1}{G_{11}} \\ 0 \end{pmatrix} I\left(\nu_1 > 0, \frac{G_{12}\nu_1}{G_{11}} - \nu_2 > 0\right) + \begin{pmatrix} 0 \\ \frac{\nu_2}{G_{22}} \end{pmatrix} I\left(\nu_2 > 0, \frac{G_{12}\nu_2}{G_{22}} - \nu_1 > 0\right) \Bigg\},$$

with $\boldsymbol{\zeta}, \boldsymbol{\Omega}$ as in Proposition 1, $\boldsymbol{\nu} = -\left(\frac{\boldsymbol{\Delta}^*}{\sigma} + \frac{1}{\sqrt{N}}\boldsymbol{\xi}\right)$, and $\boldsymbol{\xi} = (\zeta_3, \zeta_3 + \zeta_4)^\top \sim N_2(\mathbf{0}, \mathbf{G})$, where the entries of the 2-by-2 matrix $\mathbf{G}$ are given by

$$G_{11} = \frac{N}{n}\left(\frac{\bar{x}^2}{s_x^2} + 1\right) + \frac{N}{m} \cdot \frac{\tau^2}{\sigma^2}\left(\frac{\bar{z}^2}{s_z^2} + 1\right), \quad G_{12} = G_{11} - \frac{N}{n}\frac{\bar{x}}{s_x^2} - \frac{N}{m} \cdot \frac{\tau^2}{\sigma^2}\frac{\bar{z}}{s_z^2},$$
$$G_{22} = 2G_{12} - G_{11} + \frac{N}{n}\frac{1}{s_x^2} + \frac{N}{m} \cdot \frac{\tau^2}{\sigma^2}\frac{1}{s_z^2}.$$

The quantities $\bar{x}, s_x^2$ and $\bar{z}, s_z^2$ are in turn given by

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad s_x^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2, \qquad \bar{z} = \frac{1}{m}\sum_{i=n+1}^{N} x_i, \qquad s_z^2 = \frac{1}{m}\sum_{i=n+1}^{N}(x_i - \bar{z})^2.$$

Proof. The above corollary follows from Proposition 1 by evaluating the "argmin" in (18), which reduces to a simple expression. Let $\mathbf{G} = \mathbf{A}\boldsymbol{\Omega}\mathbf{A}^\top$ and $\boldsymbol{\nu} = -\left(\frac{\boldsymbol{\Delta}^*}{\sigma} + \frac{1}{\sqrt{N}}\boldsymbol{\xi}\right)$ with $\boldsymbol{\xi} \sim N_2(\mathbf{0}, \mathbf{G})$. The optimization problem inside the curly brackets in (18) then becomes

$$\min_{\boldsymbol{\lambda} \in \mathbb{R}_+^2} \frac{1}{2}\boldsymbol{\lambda}^\top \mathbf{G}\boldsymbol{\lambda} - \boldsymbol{\nu}^\top \boldsymbol{\lambda} \tag{S.1}$$

Note that we can have (i) $\{\widehat{\lambda}_1 > 0,\, \widehat{\lambda}_2 > 0\}$, (ii) $\{\widehat{\lambda}_1 > 0,\, \widehat{\lambda}_2 = 0\}$, (iii) $\{\widehat{\lambda}_1 = 0,\, \widehat{\lambda}_2 > 0\}$ and (iv) $\{\widehat{\lambda}_1 = 0,\, \widehat{\lambda}_2 = 0\}$. If case (iv) occurs, the result immediately follows. The optimality conditions of (S.1) imply that case (i) requires that $\mathbf{G}^{-1}\boldsymbol{\nu} \geq \mathbf{0}$ in which case $\widehat{\boldsymbol{\lambda}} = \mathbf{G}^{-1}\boldsymbol{\nu}$. Similarly, case (ii) requires $G_{11}\widehat{\lambda}_1 = \nu_1$ and $G_{12}\widehat{\lambda}_1 \geq \nu_2 \;\Leftrightarrow\; (G_{12}/G_{11})\nu_1 - \nu_2 \geq 0$. Case (iii) is analogous to case (ii). It remains to calculate the entries of the matrix $\mathbf{G} = \mathbf{A}\boldsymbol{\Omega}\mathbf{A}^\top$ with $\boldsymbol{\Omega}$ as in Corollary 2 and

$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}.$$

It suffices to compute the bottom 2-by-2 diagonal block of $\boldsymbol{\Omega}$. Direct calculations show that

$$\mathbf{C}_0^{-1} = \begin{pmatrix} \frac{\bar{x}^2}{s_x^2} + 1 & -\frac{\bar{x}}{s_x^2} \\ -\frac{\bar{x}}{s_x^2} & \frac{1}{s_x^2} \end{pmatrix}, \qquad \mathbf{C}_1^{-1} = \begin{pmatrix} \frac{\bar{z}^2}{s_z^2} + 1 & -\frac{\bar{z}}{s_z^2} \\ -\frac{\bar{z}}{s_z^2} & \frac{1}{s_z^2} \end{pmatrix},$$

and the given expressions for G_{11}, G_{12}, and G_{22} are then obtained by straightforward computations. $\square$

D Additional Lemmas

The following lemma states that in the case of constant variances within the two groups defined by D, i.e., $\sigma_0(X; \boldsymbol{\alpha}_0^*) \equiv \sigma$ and $\sigma_1(X; \boldsymbol{\alpha}_1^*) \equiv \tau$, least squares and weighted least squares (9) yield identical solutions.

Lemma 1. *Consider the weighted least squares criterion*

$$\min_{\boldsymbol{\beta} \in \mathbb{R}^d} \frac{1}{2} \|\mathbf{W}^{1/2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\|_2^2$$

with $\mathbf{X}$ *as defined below* (9) *and weight matrix*

$$\mathbf{W} = \operatorname{diag}\left(\underbrace{\frac{1}{\sigma^2}, \ldots, \frac{1}{\sigma^2}}_{n \text{ times}}, \underbrace{\frac{1}{\tau^2}, \ldots, \frac{1}{\tau^2}}_{m \text{ times}}\right).$$

We then have $\widehat{\boldsymbol{\beta}}^{\mathrm{WLS}} = \widehat{\boldsymbol{\beta}}^{\mathrm{LS}}$*, where* $\widehat{\boldsymbol{\beta}}^{\mathrm{WLS}}$ *denotes the minimizer of the above weighted least squares problem, and* $\widehat{\boldsymbol{\beta}}^{\mathrm{LS}}$ *denotes the ordinary least squares solution.*

Proof. We have

$$\begin{aligned}
\widehat{\boldsymbol{\beta}}^{\mathrm{WLS}} &= (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{W} \mathbf{y} \\
&= (\mathbf{L}^\top \mathbf{Z}^\top \mathbf{W} \mathbf{Z} \mathbf{L})^{-1} \mathbf{L}^\top \mathbf{Z}^\top \mathbf{W} \mathbf{y} \qquad \text{(cf. (17))} \\
&= \mathbf{L}^{-1} \left(\mathbf{Z}^\top \mathbf{W} \mathbf{Z}\right)^{-1} \mathbf{Z}^\top \mathbf{W} \mathbf{y} \\
&= \mathbf{L}^{-1} \begin{pmatrix} \frac{1}{\sigma^2}[\mathbf{1}_n \ \mathbf{Z}_0]^\top [\mathbf{1}_n \ \mathbf{Z}_0] & \mathbf{0} \\ \mathbf{0} & \frac{1}{\tau^2}[\mathbf{1}_m \ \mathbf{Z}_1]^\top [\mathbf{1}_m \ \mathbf{Z}_1] \end{pmatrix}^{-1} \begin{pmatrix} \frac{1}{\sigma^2}[\mathbf{1}_n \ \mathbf{Z}_0]^\top & \mathbf{0} \\ \mathbf{0} & \frac{1}{\tau^2}[\mathbf{1}_m \ \mathbf{Z}_1]^\top \end{pmatrix} \mathbf{y} \\
&= \mathbf{L}^{-1} \begin{pmatrix} ([\mathbf{1}_n \ \mathbf{Z}_0]^\top [\mathbf{1}_n \ \mathbf{Z}_0])^{-1} [\mathbf{1}_n \ \mathbf{Z}_0]^\top \\ ([\mathbf{1}_m \ \mathbf{Z}_1]^\top [\mathbf{1}_m \ \mathbf{Z}_1])^{-1} [\mathbf{1}_m \ \mathbf{Z}_1]^\top \end{pmatrix} \mathbf{y} \\
&= \mathbf{L}^{-1} (\mathbf{Z}^\top \mathbf{Z})^{-1} \mathbf{Z}^\top \mathbf{y} \\
&= (\mathbf{Z}^\top \mathbf{Z} \mathbf{L})^{-1} \mathbf{Z}^\top \mathbf{y} \\
&= (\mathbf{L}^\top \mathbf{Z}^\top \mathbf{Z} \mathbf{L})^{-1} \mathbf{L}^\top \mathbf{Z}^\top \mathbf{y} \\
&= (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y} = \widehat{\boldsymbol{\beta}}^{\mathrm{LS}}.
\end{aligned}$$

□

The second lemma shows that the constrained estimator of the mean difference $\widehat{\boldsymbol{\Delta}} = \mathbf{A}\widehat{\boldsymbol{\beta}}$ is closer to $\boldsymbol{\Delta}^* = \mathbf{A}\boldsymbol{\beta}^*$ than the unconstrained solution $\widehat{\boldsymbol{\Delta}}^{\mathrm{WLS}} = \mathbf{A}\widehat{\boldsymbol{\beta}}^{\mathrm{WLS}}$ with respect to the norm $\|\cdot\|_{\mathbf{H}^{-1/2}} := \|\mathbf{H}^{-1/2} \cdot\|_2$, with $\mathbf{H} = \mathbf{A}(\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{A}^\top$ as defined below (11).

Lemma 2. *With probability one, we have* $\|\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}^*\|^2_{\mathbf{H}^{-1/2}} \leq \|\widehat{\boldsymbol{\Delta}}^{\mathrm{WLS}} - \boldsymbol{\Delta}^*\|^2_{\mathbf{H}^{-1/2}}$*, with equality holding if and only if* $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Delta}}^{\mathrm{WLS}}$.

Proof. We have

$$\widehat{\boldsymbol{\beta}} = \widehat{\boldsymbol{\beta}}^{\mathrm{WLS}} + (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{A}^\top \widehat{\boldsymbol{\lambda}}.$$

This and the fact that $\widehat{\boldsymbol{\beta}}^{\mathrm{WLS}} \sim N_d(\boldsymbol{\beta}^*, (\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1})$ implies that

$$\begin{aligned}
\mathbf{A}\widehat{\boldsymbol{\beta}} = \mathbf{A}\widehat{\boldsymbol{\beta}}^{\mathrm{WLS}} + \mathbf{A}(\mathbf{X}^\top \mathbf{W} \mathbf{X})^{-1} \mathbf{A}^\top \widehat{\boldsymbol{\lambda}} &\Rightarrow \widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Delta}}^{\mathrm{WLS}} + \mathbf{H}\widehat{\boldsymbol{\lambda}} \\
&\Rightarrow \widehat{\boldsymbol{\Delta}} = \boldsymbol{\Delta}^* + \boldsymbol{\varsigma} + \mathbf{H}\widehat{\boldsymbol{\lambda}}, \quad \boldsymbol{\varsigma} \sim N_q(\mathbf{0}, \mathbf{H}). \qquad \text{(S.2)}
\end{aligned}$$

It follows that

$$\begin{aligned}
& \widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}^* = \boldsymbol{\varsigma} + \mathbf{H}\widehat{\boldsymbol{\lambda}} \\
\Rightarrow \quad & \mathbf{H}^{-1/2}(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}^*) = \mathbf{g} + \mathbf{H}^{1/2}\widehat{\boldsymbol{\lambda}}, \quad \mathbf{g} \sim N_q(\mathbf{0}, \mathbf{I}). \qquad \text{(S.3)}
\end{aligned}$$

Now note that the dual optimization problem (11) can be rewritten as

$$\min_{\boldsymbol{\lambda} \in \mathbb{R}^q_+} \frac{1}{2} \boldsymbol{\lambda}^\top \mathbf{H} \boldsymbol{\lambda} + \boldsymbol{\lambda}^\top \widehat{\boldsymbol{\Delta}}^{\mathrm{WLS}}.$$

Using the variable transformation $\boldsymbol{\gamma} = \mathbf{H}^{1/2}\boldsymbol{\lambda} \Leftrightarrow \boldsymbol{\lambda} = \mathbf{H}^{-1/2}\boldsymbol{\gamma}$, we obtain the equivalent optimization problem

$$\min_{\mathbf{H}^{-1/2}\boldsymbol{\gamma} \in \mathbb{R}^q_+} \frac{1}{2} \|\boldsymbol{\gamma}\|_2^2 + \boldsymbol{\gamma}^\top \mathbf{H}^{-1/2} \widehat{\boldsymbol{\Delta}}^{\mathrm{WLS}}.$$

Using (S.2) and (S.3), we obtain another equivalent optimization problem

$$\begin{aligned}
& \min_{\mathbf{H}^{-1/2}\boldsymbol{\gamma} \in \mathbb{R}^q_+} \left\{ \frac{1}{2} \|\boldsymbol{\gamma}\|_2^2 + \boldsymbol{\gamma}^\top \mathbf{H}^{-1/2} \boldsymbol{\Delta}^* + \boldsymbol{\gamma}^\top \mathbf{g} + \frac{1}{2} \|\mathbf{g}\|_2^2 \right\} \\
= & \min_{\mathbf{H}^{-1/2}\boldsymbol{\gamma} \in \mathbb{R}^q_+} \left\{ \boldsymbol{\gamma}^\top \mathbf{H}^{-1/2} \boldsymbol{\Delta}^* + \frac{1}{2} \|\mathbf{g} + \boldsymbol{\gamma}\|_2^2 \right\}.
\end{aligned}$$

Let $\widehat{\boldsymbol{\gamma}}$ denote the corresponding minimizer. Note that $\widehat{\boldsymbol{\gamma}}^\top \mathbf{g} < 0$ unless $\widehat{\boldsymbol{\gamma}} = \mathbf{0}$ (otherwise, $\widehat{\boldsymbol{\gamma}} = \mathbf{0}$ would be the optimal solution, since $\boldsymbol{\Delta}^* > \mathbf{0}$ and hence for any feasible non-zero $\boldsymbol{\gamma}$, we must have $\boldsymbol{\gamma}^\top \mathbf{H}^{-1/2} \boldsymbol{\Delta}^* > 0$). The latter observation implies that $\|\widehat{\boldsymbol{\gamma}} + \mathbf{g}\|_2^2 \leq \|\mathbf{g}\|_2^2$. At the same time, we note that in view of (S.2) and (S.3)

$$\mathbf{H}^{-1/2}(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}^*) = \mathbf{g} + \widehat{\boldsymbol{\gamma}}, \qquad \mathbf{H}^{-1/2}(\widehat{\boldsymbol{\Delta}}^{\mathrm{WLS}} - \boldsymbol{\Delta}^*) = \mathbf{g}.$$

Since $\|\mathbf{g} + \widehat{\boldsymbol{\gamma}}\|_2^2 \leq \|\mathbf{g}\|_2^2$ as shown above, with equality holding (with probability one) if and only if $\widehat{\boldsymbol{\gamma}} = \mathbf{0} \;\Leftrightarrow \widehat{\boldsymbol{\lambda}} = \mathbf{0}$ (and thus $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Delta}}^{\mathrm{WLS}}$), we conclude that

$$\|\mathbf{H}^{-1/2}(\widehat{\boldsymbol{\Delta}} - \boldsymbol{\Delta}^*)\|_2^2 \leq \|\mathbf{H}^{-1/2}(\widehat{\boldsymbol{\Delta}}^{\mathrm{WLS}} - \boldsymbol{\Delta}^*)\|_2^2,$$

with equality holding if and only if $\widehat{\boldsymbol{\Delta}} = \widehat{\boldsymbol{\Delta}}^{\mathrm{WLS}}$. □

The third lemma is used in the Proof of Proposition 1.

Lemma 3. *For $s > 0$, consider the quadratic program*

$$\min_{\boldsymbol{\lambda} \in \mathbb{R}_+^d} \frac{1}{2} \boldsymbol{\lambda}^\top \mathbf{Q} \boldsymbol{\lambda} - s \cdot \mathbf{h}^\top \boldsymbol{\lambda}, \tag{S.4}$$

for a symmetric positive definite matrix $\mathbf{Q}$, and let $\widehat{\boldsymbol{\lambda}}(s)$ be its minimizer. We then have $\widehat{\boldsymbol{\lambda}}(s) = s\widehat{\boldsymbol{\lambda}}(1)$.

Proof. Let $\mathcal{A}(1) = \{1 \leq j \leq d : \;\; \widehat{\lambda}_j(1) > 0\}$ be the active set of $\widehat{\boldsymbol{\lambda}}(1)$. In particular, $\widehat{\boldsymbol{\lambda}}_{\mathcal{A}(1)}(1) = (\mathbf{Q}_{\mathcal{A}(1)\mathcal{A}(1)})^{-1} \mathbf{h}_{\mathcal{A}(1)}$, where the subscripts $\mathcal{A}(1)\mathcal{A}(1)$ and $\mathcal{A}(1)$ refer to the principal submatrix and subvector, respectively, corresponding to $\mathcal{A}(1)$. We shall demonstrate that for any $s > 0$, it holds that $\mathcal{A}(s) = \mathcal{A}(1)$. For this purpose, observe that

$$\begin{aligned}
\mathbf{Q}_{\mathcal{A}(1)\mathcal{A}(1)}(\widehat{\boldsymbol{\lambda}}_{\mathcal{A}(1)}(1) \cdot s) = s\mathbf{h}_{\mathcal{A}(1)} \quad &\Leftrightarrow \mathbf{Q}_{\mathcal{A}(1)\mathcal{A}(1)} \widehat{\boldsymbol{\lambda}}_{\mathcal{A}(1)}(1) = \mathbf{h}_{\mathcal{A}(1)}, \\
\mathbf{Q}_{\mathcal{A}(1)^c\mathcal{A}(1)}(\widehat{\boldsymbol{\lambda}}_{\mathcal{A}(1)}(1) \cdot s) \geq s\mathbf{h}_{\mathcal{A}(1)^c} \;\; &\Leftrightarrow \mathbf{Q}_{\mathcal{A}(1)^c\mathcal{A}(1)} \widehat{\boldsymbol{\lambda}}_{\mathcal{A}(1)}(1) \geq \mathbf{h}_{\mathcal{A}(1)^c},
\end{aligned}$$

Noting that the left hand sides in the above display constitute the necessary and sufficient optimality conditions for optimization problem (S.4) (cf. Chen and Plemmons, 2010, p. 5) we conclude that $\mathcal{A}(s) = \mathcal{A}(1)$ and $\widehat{\boldsymbol{\lambda}}(s) = s\widehat{\boldsymbol{\lambda}}(1)$. $\square$

E Additional simulation studies

In this section, we complement the results presented in §5 in the following ways.

E.1 We examine the relative efficiency (RE) concerning estimation of the mean difference $\Delta^*(x)$ and the covariate-specific ROC curve $\mathrm{ROC}_x^*(u)$ in dependence of *various combinations* of (i) the specific value taken by the covariate x and (ii) the FAR u.

E.2 We complete the result of our simulation for the mean difference and the covariate-specific ROC curves in §5 using Table.

E.3 We visualize the results of the simulation study in §5 using multiple figures as an alternative to the Tables 1 and 2.

E.4 We shed light on the performance of our method in the presence of *multiple*, possibly highly correlated covariates.

We recall that values of RE larger than one are equivalent to a smaller MSE of the proposed order-constrained method relative to the unconstrained method.

E.1 Relative efficiency of mean differences and covariate-specific ROC curves

RE of mean differences. In Figure 6, the RE attains its minimum when $x = 0.5$ for both choices of the sample size ($N = 20$ and $N = 100$) under consideration. For all four studies, the RE for $N = 100$ is always smaller than for $N = 20$; in all cases, the RE exceeds the baseline of one except for study 4 in which the RE is close to one.

RE of covariate-specific ROC curves. Figures 7 to 10 display the RE for the estimated covariate-specific ROC curves. Figures 7 and 8 graph RE versus x for a set of fixed values of the FAR u, different sample sizes, and different error distributions. The RE attains its minimum when $x = 0.5$. Moreover, the RE increases with u. In Figures 9 and 10, the roles of x and u are swapped compared to the two previous figures. It should be emphasized that the RE is above one in all figures except for two cases in study 4 in which the RE is slightly below one.

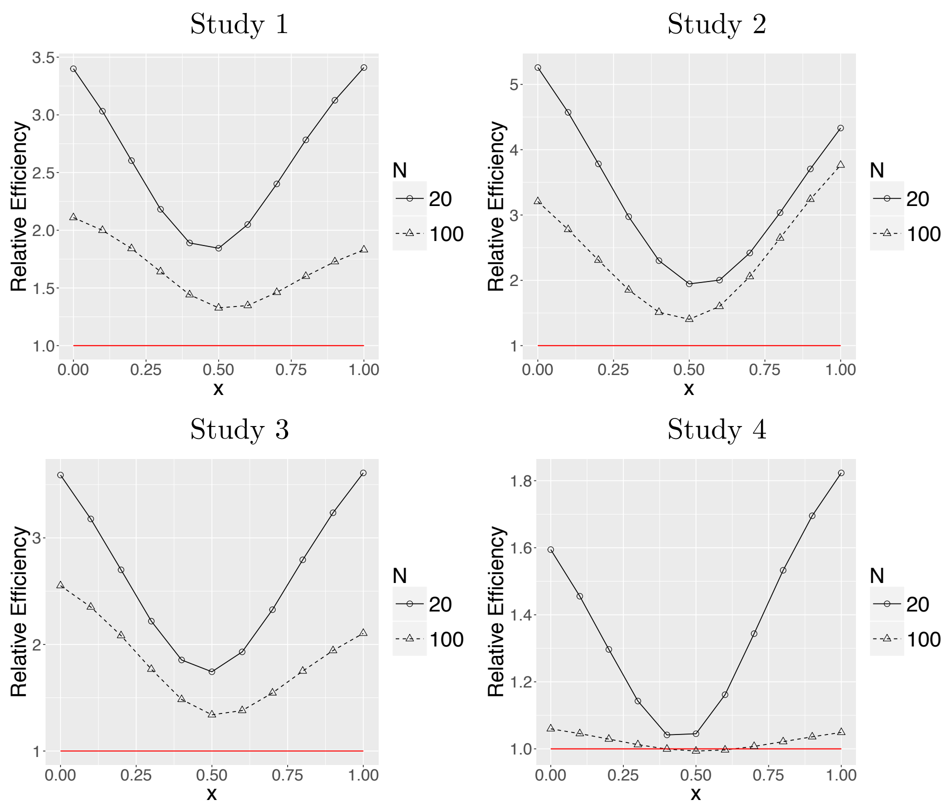

Figure 6: Relative efficiency (RE) of the estimated mean difference in dependence of x. The red horizontal line corresponds to an RE of one.

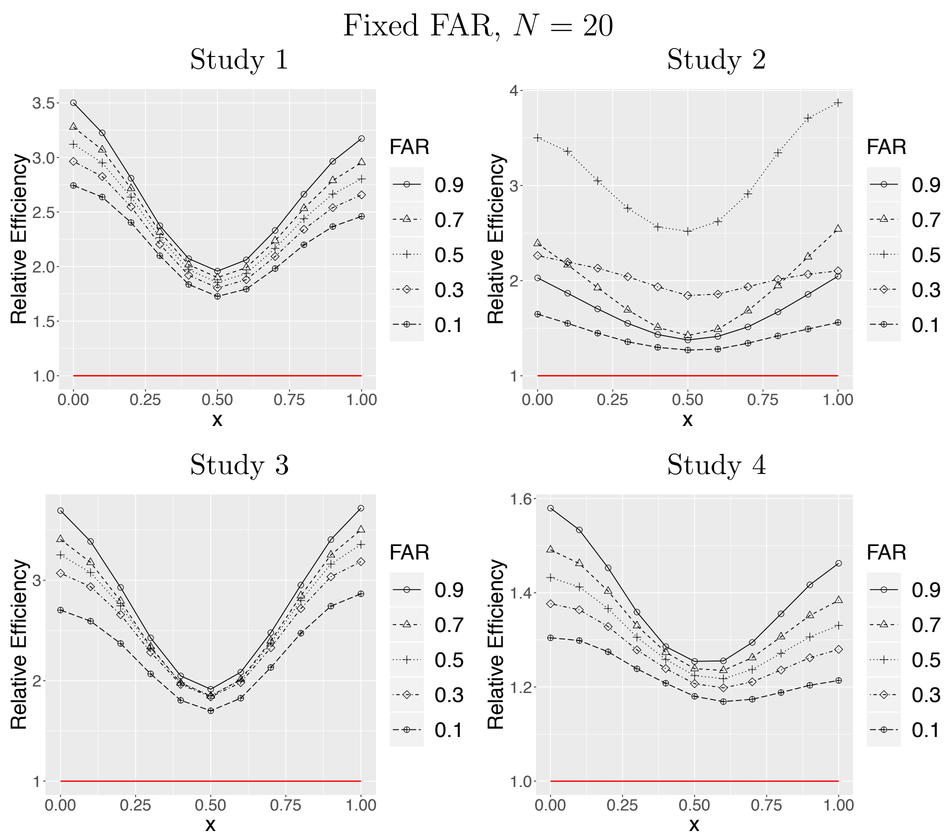

Figure 7: Relative efficiency (RE) of the estimated ROC in dependence of x for different values of FAR when $N = 20$. The red horizontal line corresponds to an RE of one.

Fixed FAR, $N = 100$

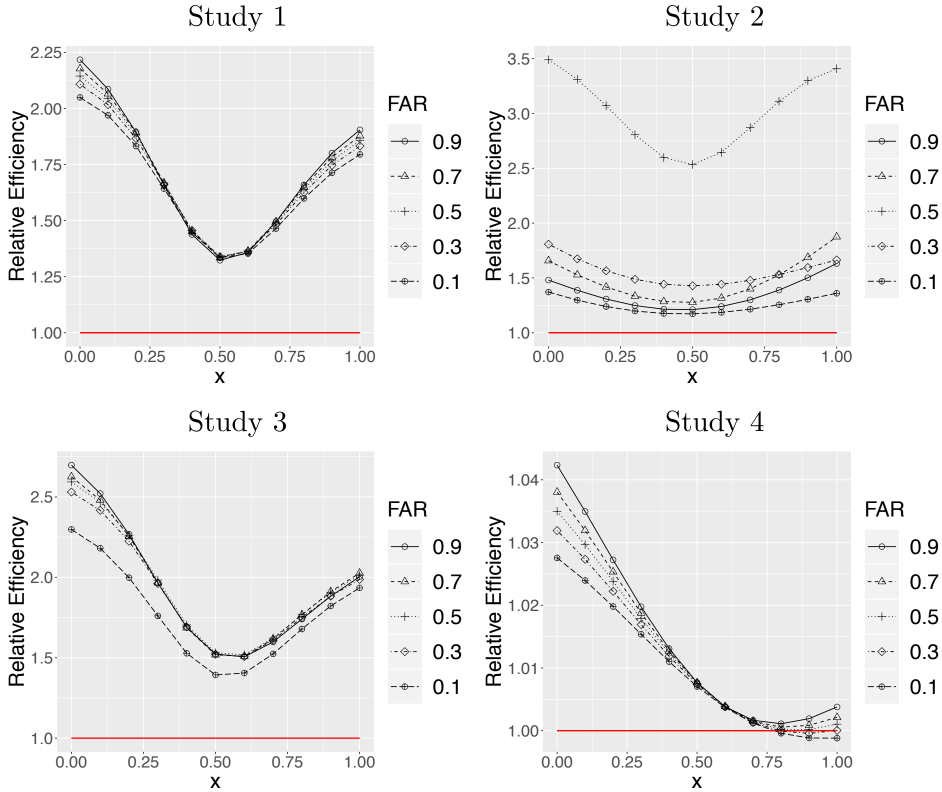

Figure 8: Relative efficiency (RE) of the estimated ROC in dependence of x for different values of FAR when $N = 100$. The red horizontal line corresponds to an RE of one.

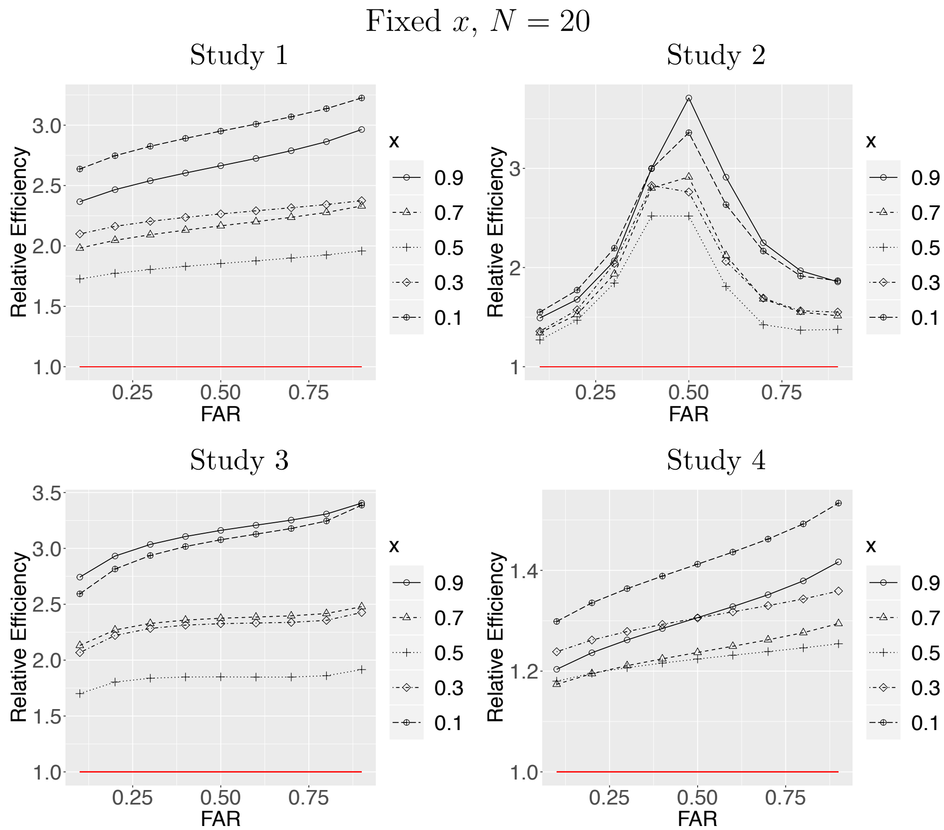

Figure 9: Relative efficiency (RE) of the estimated ROC in dependence of FAR for different values of x when $N = 20$. The red horizontal line corresponds to an RE of one.

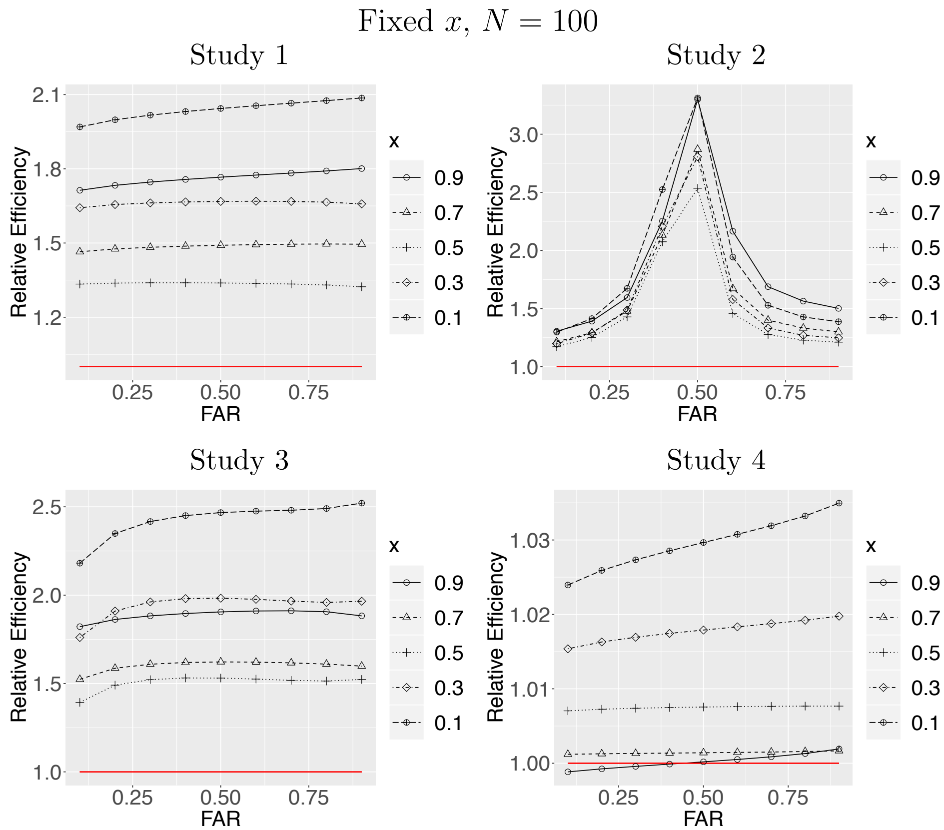

Figure 10: Relative efficiency (RE) of the estimated ROC in dependence of FAR for different values of x when $N = 100$. The red horizontal line corresponds to an RE of one.

E.2 Bias and MSE of the mean differences and the covariate-specific ROC curves

Table 1 shows the bias and MSE in estimating the mean difference $\Delta^*(x)$ for $x = 0.5$ for all four studies, in dependence of the sample size and different values for ψ and ϕ. As expected, the bias of the constrained method exceeds that without constraint. On the other hand, the MSE of the former is smaller, i.e., the relative efficiency (RE) is larger than one throughout. As the sample size increases, the RE gets closer to one, which is anticipated in light of Corollary 1. Furthermore, it can be seen that small values of ψ and large values of ϕ lead to an increase of the RE. The largest REs tend to be attained in Study 1 (normal errors), which is not surprising given the optimality of least squares regression in this case.

We further investigate the bias and the MSE in estimating $\mathrm{ROC}_x^*(u)$. Specifically, we fix $u = 0.5$ and $x = 0.5$. Table 2 depicts the results in studies 1 to 4, respectively. Small sample sizes yield large values for the RE in alignment with the results in Table 1. The RE noticeably exceeds one, showing advantages of the proposed use of order constraints. In Study 4, unconstrained estimation performs on par with constrained estimation.

E.3 Visualizations of the results

Figure 11 and Figure 12 provide an alternative representation of the results reported in Table 1 and 2, respectively. Recall that according to the data-generating model (19) underlying the simulations, we denote by ψ the location parameter and by ϕ the scale parameter. It can be seen from Figure 11 that the RE decreases and approaches one as the sample size increases. The REs for the mean difference also increase with small values of ψ and large values of ϕ. The RE exceeds one in each figure, which shows that our method reduces the MSE.

				Study 1			Study 2			Study 3			Study 4		
N	ψ	$\sqrt{\phi}$	Method	B	MSE	RE(SE)	B	MSE	RE(SE)	B	MSE	RE(SE)	B	MSE	RE(SE)
20	0.5	8	w/o	0.13	13.70	2.52(0.79)	-1.48	4759.14	2.07(0.49)	-0.33	14.61	1.83(0.18)	0.08	4.44	1.79(0.42)
			w/	1.14	5.43		18.43	2299.81		1.35	8.00		0.54	2.49	
		16	w/o	0.26	52.25	2.56(0.81)	-2.75	18878.61	2.07(0.49)	-0.67	57.93	1.85(0.18)	0.16	17.58	1.85(0.44)
			w/	2.59	20.20		36.06	9106.17		3.06	31.38		1.39	9.52	
		32	w/o	0.51	216.32	2.55(0.80)	-5.30	75390.17	2.08(0.50)	-1.34	231.48	1.84(0.18)	0.33	70.11	1.85(0.44)
			w/	5.55	84.88		71.46	36257.42		6.78	125.88		3.17	37.95	
	1	8	w/o	0.13	13.70	2.45(0.76)	-1.48	4759.14	2.07(0.49)	-0.33	14.61	1.80(0.17)	0.08	4.44	1.69(0.39)
			w/	0.99	5.59		18.22	2297.45		1.16	8.12		0.41	2.63	
		16	w/o	0.26	54.25	2.55(0.80)	-2.75	18878.61	2.07(0.49)	-0.67	57.93	1.85(0.18)	0.16	17.58	1.82(0.43)
			w/	2.41	21.25		35.83	9101.16		2.85	31.32		1.21	9.67	
		32	w/o	0.51	216.32	2.56(0.81)	-5.30	75390.17	2.08(0.50)	-1.34	231.48	1.85(0.18)	0.33	70.11	1.85(0.44)
			w/	5.35	84.65		71.23	36246.83		6.32	125.31		2.96	37.94	
100	0.5	8	w/o	-0.05	1.33	1.43(0.10)	1.01	682.21	1.43(0.13)	-0.06	2.60	1.54(0.11)	-0.02	0.48	1.16(0.08)
			w/	0.26	0.93		12.44	476.73		0.48	1.69		0.09	0.42	
		16	w/o	-0.08	5.22	1.56(0.11)	2.09	2687.60	1.42(0.13)	-0.13	10.23	1.60(0.12)	-0.04	1.89	1.33(0.10)
			w/	0.79	3.34		25.00	1887.30		1.27	6.38		0.35	1.42	
		32	w/o	-0.16	20.77	1.59(0.12)	4.24	10698.28	1.42(0.13)	-0.24	40.69	1.60(0.12)	-0.08	7.54	1.43(0.11)
			w/	1.94	13.08		50.26	7526.15		2.92	25.45		1.02	5.28	
	1	8	w/o	-0.05	1.33	1.29(0.08)	1.01	682.21	1.44(0.13)	-0.06	2.60	1.45(0.10)	-0.02	0.48	1.07(0.07)
			w/	0.16	1.03		12.21	475.33		0.36	1.79		0.03	0.45	
		16	w/o	-0.08	5.22	1.50(0.11)	2.09	2687.60	1.43(0.13)	-0.12	10.23	1.58(0.11)	-0.04	1.89	1.23(0.09)
			w/	0.64	3.48		24.77	1884.01		1.10	6.47		0.23	1.54	
		32	w/o	-0.16	20.77	1.58(0.12)	4.24	10698.28	1.42(0.13)	-0.24	40.69	1.61(0.12)	-0.08	7.54	1.39(0.11)
			w/	1.75	13.13		50.03	7519.18		2.72	25.35		0.84	5.42	

Table 1: Bias (B) and MSE of the mean difference $\Delta^*(x)$ for $x = 0.5$. (w/o: linear regression without order constraint; w/: linear regression with order constraint.)

				Study 1			Study 2			Study 3			Study 4		
N	ψ	$\sqrt{\phi}$	Method	B	MSE	RE(SE)	B	MSE	RE(SE)	B	MSE	RE(SE)	B	MSE	RE(SE)
20	0.5	8	w/o	0.78	2.82	1.73(0.13)	-1.89	4.88	2.21(0.17)	-1.44	2.43	1.91(0.15)	-43.01	21.15	1.39(0.04)
			w/	6.48	1.63		6.89	2.21		4.96	1.27		-37.69	15.22	
		16	w/o	0.63	2.81	1.76(0.14)	-0.87	3.64	1.96(0.15)	-1.23	2.40	1.91(0.15)	-47.88	25.77	1.47(0.04)
			w/	7.18	1.60		7.54	1.86		5.97	1.25		-40.77	17.48	
		32	w/o	0.55	2.81	1.76(0.14)	-0.37	2.89	1.85(0.14)	-1.33	2.39	1.89(0.15)	-50.29	28.24	1.53(0.04)
			w/	7.59	1.60		7.59	1.57		6.52	1.27		-42.09	18.49	
	1	8	w/o	0.91	2.77	1.68(0.12)	-2.81	4.98	2.36(0.18)	-1.64	2.42	1.89(0.14)	-38.21	16.97	1.31(0.04)
			w/	5.75	1.65		5.44	2.11		3.96	1.28		-34.40	12.94	
		16	w/o	0.71	2.80	1.74(0.13)	-1.26	3.67	2.03(0.16)	-1.34	2.40	1.93(0.15)	-45.44	23.38	1.42(0.04)
			w/	6.76	1.61		6.82	1.80		5.41	1.25		-39.37	16.38	
		32	w/o	0.59	2.81	1.67(0.14)	-0.54	2.90	1.89(0.14)	-1.19	2.39	1.90(0.15)	-49.09	26.99	1.50(0.04)
			w/	7.36	1.60		7.23	1.54		6.22	1.25		-41.47	18.01	
100	0.5	8	w/o	-0.16	0.34	1.42(0.10)	-3.84	0.71	2.66(0.32)	-1.16	0.36	1.80(0.11)	-42.27	18.13	1.05(0.01)
			w/	1.38	0.24		-0.49	0.27		0.77	0.20		-41.29	17.25	
		16	w/o	-0.16	0.34	1.55(0.12)	-1.84	0.44	2.37(0.20)	-0.69	0.34	1.82(0.13)	-46.54	21.97	1.09(0.01)
			w/	2.03	0.22		1.37	0.19		1.82	0.19		-44.70	20.15	
		32	w/o	-0.17	0.34	1.58(0.12)	-0.85	0.37	1.95(0.12)	-0.46	0.33	1.73(0.12)	-48.69	24.04	1.12(0.01)
			w/	2.47	0.21		2.31	0.19		2.41	0.19		-46.11	21.40	
	1	8	w/o	-0.12	0.34	1.28(0.09)	-5.80	0.92	2.55(0.22)	-1.61	0.37	1.68(0.10)	-38.05	14.70	1.03(0.01)
			w/	0.90	0.27		-2.60	0.36		-0.13	0.22		-37.60	14.33	
		16	w/o	-0.15	0.34	1.49(0.11)	-2.83	0.50	2.77(0.23)	-0.92	0.34	1.84(0.12)	-44.39	19.98	1.06(0.01)
			w/	1.65	0.23		0.30	0.18		1.26	0.19		-43.13	18.78	
		32	w/o	-0.16	0.34	1.58(0.12)	-1.38	0.38	2.24(0.14)	-0.57	0.34	1.79(0.13)	-47.61	22.98	1.10(0.01)
			w/	2.24	0.21		1.77	0.17		2.10	0.19		-45.45	20.81	

Table 2: Bias (B) and MSE of $\mathrm{ROC}_x^*(u)$ for $x = 0.5$ and $\mathrm{FAR} = 0.5$. (w/o: linear regression without order constraint; w/: linear regression with order constraint). All values of B and MSE have been multiplied by 100.

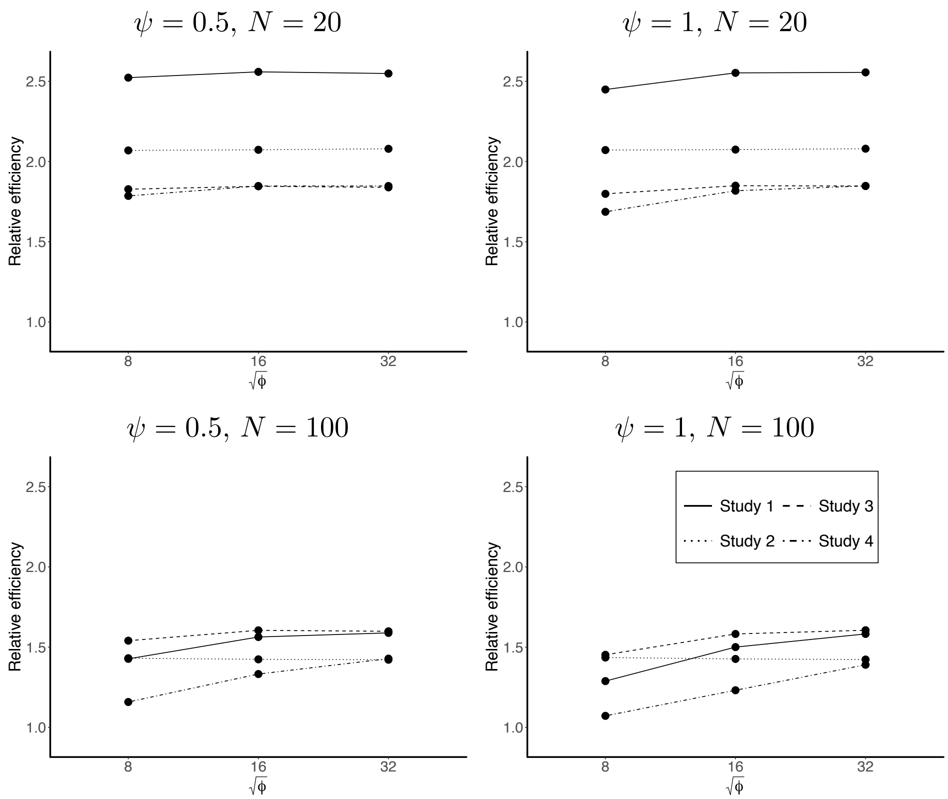

Figure 11: Relative efficiency (RE) of the estimated mean difference in dependence of ψ for different values of ϕ when $x = 0.5$.

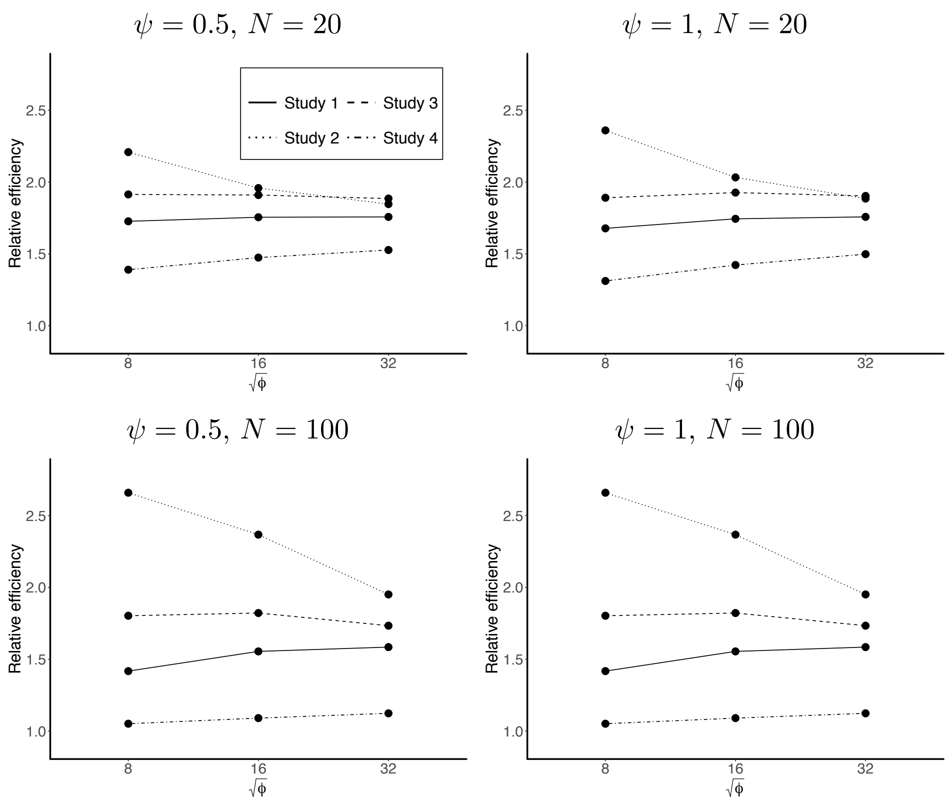

Figure 12: Relative efficiency (RE) of the estimated ROC in dependence of ψ for different values of ϕ when $x = 0.5$ and FAR $= 0.5$.

E.4 Bias and MSE of mean difference and covariate-specific ROC curves with multiple predictors

We here consider the case of multiple ($p > 1$), possibly highly correlated covariates. Extending the simulation model (19), we generate data as follows:

$$T_i = 1 + \mathbf{x}_i^\top \boldsymbol{\theta} + D_i \times \psi + D_i \times \mathbf{x}_i^\top \boldsymbol{\theta} + e_{1i} D_i \times \sqrt{\phi} + e_{0i}(1 - D_i),$$

- $\mathbf{x}_i = (x_{i1}, \ldots, x_{ip})^\top$ is such that $x_{ij} = \Phi(g_{ij})$ follows a uniform distribution, $j = 1, \ldots, p$, where

  $$\mathbf{g}_i = (g_{i1}, \ldots, g_{ip})^\top, \quad \mathbf{g}_i \sim N_d(\mathbf{0}, \boldsymbol{\Sigma}_\rho), \quad 1 \leq i \leq N.$$

  The covariance matrix $\boldsymbol{\Sigma}_\rho$ has entries one on its diagonal, and $0 \leq \rho < 1$ otherwise.

- The entries of $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)^\top$ are all set to $1/p$.

- All other quantities remain unchanged compared to the simulations for $p = 1$.

The results of the bias and MSE with regard to estimation of the mean difference and the covariate-specific ROC curve for (i) $\mathbf{x} = (0.5, \ldots, 0.5)^\top$ and (ii) $\mathbf{x} = (1, 0, \ldots, 0)^\top$ in dependence of different values of the sample size N, the number of covariates p, and the correlation coefficient ρ.

Table 3 shows the bias and MSE in estimating the mean difference for $\mathbf{x} = (0.5, \ldots, 0.5)^\top$. The RE increases with the number of predictors and ρ but decreases with the sample size. All REs are larger than the baseline 1, which implies that the proposed method outperforms the traditional method. Table 4 shows the bias and MSE in estimating the conditional ROC for $\mathbf{x} = (0.5, \ldots, 0.5)^\top$. In Table 4, with the exception of the two cases in study 1 when the sample size is 50, most of the REs are larger than 1.

Table 5 and Table 6 demonstrate the bias and MSE in estimating the mean difference and ROC for $\mathbf{x} = (1, 0, \ldots, 0)^\top$. Comparing with Table 3 and 4, it can be seen that the values of both bias and MSE of the proposed method change slightly as $\mathbf{x}$ changes, where the MSE of the conventional method is changes dramatically when $\mathbf{x}$ changes. All values of RE are larger than 1 in these two tables.

				Study 1			Study 2			Study 3			Study 4		
N	p	ρ	Method	B	MSE	RE(SE)	B	MSE	RE (SE)	B	MSE	RE(SE)	B	MSE	RE(SE)
50	1	-	w/o	-0.03	2.73	1.41(0.10)	0.87	1441.26	1.40(0.19)	0.13	6.09	1.48(0.12)	-0.01	1.04	1.15(0.08)
			w/	0.39	1.93		15.37	1026.54		0.82	4.11		0.13	0.90	
	2	0.5	w/o	0.01	3.20	1.36(0.09)	-2.05	1422.08	1.66(0.19)	-0.09	6.15	1.47(0.11)	0.02	1.20	1.15(0.10)
			w/	0.51	2.36		13.93	854.38		0.75	4.20		0.22	1.05	
		0.9	w/o	0.01	3.19	1.39(0.09)	-2.03	1427.52	1.78(0.20)	-0.09	6.14	1.53(0.11)	0.02	1.19	1.18(0.10)
			w/	0.46	2.29		13.37	802.42		0.68	4.02		0.19	1.01	
	5	0.5	w/o	-0.03	3.17	1.41(0.09)	-1.50	1560.73	1.57(0.17)	0.10	7.03	1.55(0.12)	-0.02	1.16	1.18(0.09)
			w/	0.60	2.25		16.11	996.03		0.96	4.52		0.27	1.00	
		0.9	w/o	-0.03	3.17	1.53(0.10)	-1.24	1550.42	1.79(0.20)	0.10	6.99	1.72(0.13)	-0.02	1.16	1.26(0.09)
			w/	0.46	2.07		14.59	864.54		0.76	4.07		0.18	0.92	
	10	0.5	w/o	0.06	5.34	2.04(0.14)	-0.01	2263.82	1.77(0.20)	-0.02	9.30	1.87(0.16)	0.04	1.92	1.82(0.14)
			w/	0.75	2.62		19.15	1279.04		1.07	4.97		0.35	1.06	
		0.9	w/o	0.04	5.30	2.31(0.15)	-0.41	2335.25	2.37(0.27)	0.02	9.30	2.19(0.20)	0.03	1.91	2.01(0.14)
			w/	0.54	2.29		16.37	985.06		0.77	4.26		0.23	0.95	
100	1	-	w/o	-0.05	1.33	1.29(0.08)	1.01	682.21	1.44(0.13)	-0.06	2.60	1.45(0.10)	-0.02	0.48	1.07(0.07)
			w/	0.16	1.03		12.21	475.33		0.36	1.79		0.03	0.45	
	2	0.5	w/o	-0.02	1.42	1.23(0.08)	1.08	762.43	1.46(0.13)	0.07	2.70	1.23(0.09)	-0.01	0.52	1.06(0.07)
			w/	0.27	1.16		12.99	520.79		0.53	2.21		0.10	0.49	
		0.9	w/o	-0.02	1.43	1.24(0.08)	1.10	759.93	1.54(0.14)	0.07	2.70	1.28(0.10)	-0.01	0.52	1.07(0.07)
			w/	0.23	1.15		12.43	493.24		0.47	2.12		0.09	0.48	
	5	0.5	w/o	0.07	1.43	1.19(0.08)	0.19	736.11	1.36(0.13)	0.02	2.92	1.31(0.09)	0.03	0.51	1.06(0.07)
			w/	0.40	1.20		13.28	540.22		0.59	2.22		0.17	0.48	
		0.9	w/o	0.07	1.44	1.24(0.08)	0.09	743.19	1.62(0.16)	0.02	2.91	1.43(0.10)	0.03	0.51	1.09(0.08)
			w/	0.32	1.16		11.77	459.01		0.46	2.04		0.13	0.47	
	10	0.5	w/o	-0.01	1.64	1.37(0.09)	-0.02	815.33	1.46(0.14)	0.10	3.48	1.42(0.10)	-0.01	0.59	1.26(0.09)
			w/	0.42	1.20		13.74	559.49		0.73	2.46		0.18	0.47	
		0.9	w/o	-0.02	1.62	1.50(0.10)	-0.15	819.44	1.88(0.19)	0.09	3.42	1.57(0.11)	-0.01	0.59	1.34(0.09)
			w/	0.27	1.08		11.40	436.93		0.52	2.17		0.10	0.44	

Table 3: Bias (B) and MSE of the mean difference $\Delta^*(\mathbf{x})$ for $\mathbf{x} = (0.5, \ldots, 0.5)^\top$, where $\mathbf{x} \in \mathbb{R}^p$. (w/o: linear regression without order constraint; w/: linear regression with order constraint.)

				Study 1			Study 2			Study 3			Study 4		
N	p	ρ	Method	B	MSE	RE(SE)	B	MSE	RE(SE)	B	MSE	RE(SE)	B	MSE	RE(SE)
50	1	-	w/o	-0.01	0.71	1.43(0.09)	-4.50	2.25	2.40(0.23)	-0.78	0.77	1.67(0.11)	-38.12	15.05	1.08(0.02)
			w/	2.10	0.50		0.89	0.94		1.69	0.46		-36.85	13.95	
	2	0.5	w/o	-1.00	0.62	1.18(0.08)	-5.35	1.72	2.83(0.27)	-2.40	0.64	1.41(0.09)	-39.82	16.32	1.18(0.02)
			w/	2.27	0.52		-0.54	0.61		1.15	0.45		-36.72	13.84	
		0.9	w/o	-2.56	0.47	0.96(0.06)	-5.75	1.32	2.83(0.25)	-3.46	0.51	1.18(0.08)	-42.23	18.20	1.28(0.02)
			w/	1.86	0.49		-1.46	0.47		0.75	0.43		-37.19	14.18	
	5	0.5	w/o	-4.40	0.35	0.96(0.06)	-6.20	0.82	2.70(0.19)	-4.46	0.37	1.09(0.06)	-45.19	20.57	1.40(0.01)
			w/	1.65	0.36		-2.11	0.31		0.95	0.34		-37.99	14.70	
		0.9	w/o	-5.95	0.40	1.37(0.07)	-6.59	0.57	1.92(0.08)	-5.76	0.38	1.38(0.06)	-47.63	22.73	1.41(0.01)
			w/	0.09	0.29		-3.37	0.30		-0.60	0.27		-39.76	16.07	
	10	0.5	w/o	-6.13	0.41	1.88(0.09)	-6.73	0.55	2.05(0.10)	-6.03	0.40	1.87(0.11)	-48.00	23.08	1.40(0.01)
			w/	0.01	0.22		-3.07	0.27		-0.64	0.21		-40.32	16.45	
		0.9	w/o	-6.89	0.48	2.36(0.08)	-6.94	0.50	1.63(0.05)	-6.58	0.44	2.02(0.06)	-49.15	24.16	1.26(0.01)
			w/	-2.79	0.20		-4.48	0.31		-3.15	0.22		-43.56	19.12	
100	1	-	w/o	-0.12	0.34	1.28(0.09)	-5.80	0.92	2.55(0.22)	-1.61	0.37	1.68(0.10)	-38.05	14.70	1.03(0.01)
			w/	0.90	0.27		-2.60	0.36		-0.13	0.22		-37.60	14.33	
	2	0.5	w/o	-0.74	0.31	1.16(0.07)	-5.98	0.81	2.91(0.17)	-1.80	0.31	1.38(0.08)	-39.17	15.55	1.10(0.01)
			w/	1.18	0.26		-2.83	0.28		0.26	0.22		-37.43	14.18	
		0.9	w/o	-1.95	0.28	1.08(0.06)	-6.16	0.70	2.61(0.13)	-2.77	0.28	1.29(0.08)	-41.13	17.12	1.19(0.01)
			w/	0.91	0.25		-3.23	0.27		0.01	0.22		-37.72	14.40	
	5	0.5	w/o	-3.15	0.23	1.04(0.07)	-6.31	0.60	2.40(0.11)	-3.85	0.28	1.42(0.08)	-43.19	18.77	1.29(0.01)
			w/	1.22	0.22		-3.34	0.25		0.11	0.20		-37.88	14.49	
		0.9	w/o	-5.17	0.31	1.57(0.09)	-6.64	0.51	1.85(0.06)	-5.29	0.33	1.75(0.08)	-46.38	21.55	1.41(0.01)
			w/	0.32	0.20		-4.18	0.27		-0.75	0.19		-38.94	15.32	
	10	0.5	w/o	-5.47	0.33	2.45(0.12)	-6.77	0.51	2.04(0.06)	-5.38	0.32	2.25(0.10)	-46.78	21.91	1.38(0.01)
			w/	0.02	0.14		-3.94	0.25		-0.54	0.14		-39.72	15.88	
		0.9	w/o	-6.57	0.44	3.11(0.12)	-6.92	0.49	1.59(0.03)	-6.26	0.40	2.55(0.09)	-48.57	23.59	1.33(0.01)
			w/	-1.93	0.14		-4.95	0.31		-2.34	0.16		-41.97	17.71	

Table 4: Bias (B) and MSE of $\mathrm{ROC}^*_{\mathbf{x}}(u)$ for $\mathbf{x} = (0.5, \ldots, 0.5)^\top$ where $\mathbf{x} \in \mathbb{R}^p$, and FAR = 0.5. (w/o: linear regression without order constraint; w/: linear regression with order constraint. All values of B and MSE have been multiplied by 100.)

N	p	ρ	Method	Study 1 B	Study 1 MSE	Study 1 RE(SE)	Study 2 B	Study 2 MSE	Study 2 RE (SE)	Study 3 B	Study 3 MSE	Study 3 RE(SE)	Study 4 B	Study 4 MSE	Study 4 RE(SE)
50	1	-	w/o	-0.08	11.51	2.22(0.16)	-0.53	5848.48	3.19(0.63)	0.09	23.14	2.45(0.22)	-0.06	4.37	1.49(0.12)
			w/	0.09	5.17		14.99	1026.54		0.53	9.46		-0.09	2.94	
	2	0.5	w/o	0.66	39.78	7.55(0.65)	1.05	17610.01	12.36(1.75)	-0.51	77.70	10.07(0.99)	0.42	14.84	5.24(0.52)
			w/	0.57	5.27		15.04	1424.95		0.67	7.72		0.29	2.83	
		0.9	w/o	1.42	186.12	35.77(3.12)	3.87	80450.89	61.87(8.99)	-1.00	363.31	47.74(4.82)	0.90	67.48	23.49(2.09)
			w/	0.50	5.20		14.13	1300.42		0.60	7.61		0.24	2.87	
	5	0.5	w/o	-0.63	75.36	25.19(2.29)	-7.62	40461.72	47.37(8.82)	0.85	160.14	34.32(3.42)	-0.32	26.31	15.18(1.41)
			w/	0.01	2.99		8.96	854.15		0.29	4.67		-0.16	1.73	
		0.9	w/o	-1.29	344.14	117.32(10.76)	-11.03	174154.67	210.68(37.12)	1.84	746.54	157.16(16.50)	-0.66	120.37	73.30(6.52)
			w/	0.13	2.93		9.39	826.64		0.43	3.43		-0.06	1.64	
	10	0.5	w/o	0.22	132.60	55.88(5.39)	-10.25	74462.32	188.71(39.49)	-0.13	265.72	77.58(10.24)	0.15	47.56	31.85(2.83)
			w/	-0.27	2.37		5.01	394.60		-0.20	3.43		-0.40	1.49	
		0.9	w/o	0.28	612.98	236.57(22.26)	-15.03	343743.72	721.07(137.15)	0.11	1297.94	330.54(47.27)	0.21	219.34	152.08(13.42)
			w/	-0.01	2.59		7.04	476.72		0.11	3.93		-0.18	1.43	
100	1	-	w/o	0.05	5.19	1.75(0.11)	-0.59	2500.40	3.25(0.41)	-0.08	11.11	2.25(0.16)	0.05	1.91	1.24(0.09)
			w/	0.03	2.96		11.04	768.61		0.09	4.93		-0.01	1.54	
	2	0.5	w/o	-0.15	18.22	6.39(0.48)	-5.16	8271.16	9.68(1.16)	-0.15	37.05	7.70(0.67)	-0.07	6.53	4.02(0.29)
			w/	0.24	2.85		13.56	854.81		0.51	4.81		0.10	1.62	
		0.9	w/o	-0.30	82.07	27.77(2.12)	9.20	36276.85	45.27(5.63)	-0.38	167.71	35.25(3.15)	-0.13	29.59	16.64(1.18)
			w/	0.19	2.96		12.84	801.28		0.46	4.76		0.08	1.78	
	5	0.5	w/o	0.11	31.87	14.59(1.12)	-0.01	16781.19	39.30(5.57)	0.10	65.46	21.43(1.80)	0.11	11.97	9.30(0.71)
			w/	0.60	2.18		7.16	427.02		0.11	3.05		-0.09	1.29	
		0.9	w/o	0.08	141.32	65.04(4.90)	-0.87	74247.59	192.56(27.69)	0.08	293.56	99.50(8.64)	0.14	53.52	42.12(3.16)
			w/	0.15	2.17		7.29	385.57		0.20	2.95		0.03	1.27	
	10	0.5	w/o	-0.24	41.53	24.86(2.03)	-6.04	25492.42	149.22(26.39)	0.36	84.74	35.79(3.99)	-0.12	15.43	13.99(0.98)
			w/	-0.40	1.67		3.22	170.83		-0.32	2.37		-0.45	1.10	
		0.9	w/o	-0.25	179.81	117.23(8.48)	-10.55	107392.76	502.03(79.35)	0.74	363.77	145.88(15.32)	-0.13	66.57	66.82(4.81)
			w/	-0.19	1.53		4.90	213.91		-0.02	2.49		-0.21	1.00	

Table 5: Bias (B) and MSE of the mean difference $\Delta^*(\mathbf{x})$ for $\mathbf{x} = (1, 0, \dots, 0)^\top$, where $\mathbf{x} \in \mathbb{R}^p$. (w/o: linear regression without order constraint; w/: linear regression with order constraint.)

				Study 1			Study 2			Study 3			Study 4		
N	p	ρ	Method	B	MSE	RE(SE)	B	MSE	RE(SE)	B	MSE	RE(SE)	B	MSE	RE(SE)
50	1	-	w/o	-0.92	2.52	2.30(0.14)	-6.57	5.00	3.50(0.25)	-1.86	2.62	2.52(0.16)	-35.44	14.74	1.07(0.03)
			w/	0.20	1.10		-1.78	1.43		-0.14	1.04		-35.37	13.80	
	2	0.5	w/o	0.28	3.69	3.61(0.22)	-5.71	6.09	6.50(0.50)	-3.51	3.84	4.96(0.31)	-38.86	18.61	1.28(0.04)
			w/	2.30	1.02		-0.60	0.94		0.68	0.77		-36.70	14.54	
		0.9	w/o	-1.15	5.83	5.94(0.35)	-6.14	8.34	10.84(0.79)	-4.95	6.11	8.12(0.50)	-41.15	22.65	1.51(0.05)
			w/	1.81	0.98		-1.40	0.77		0.30	0.75		-37.31	15.01	
	5	0.5	w/o	-4.66	2.21	4.21(0.26)	-5.41	3.75	10.30(0.68)	-2.73	2.19	4.70(0.28)	-47.55	24.61	1.30(0.03)
			w/	-0.65	0.52		-2.84	0.36		-0.63	0.47		-42.82	18.98	
		0.9	w/o	-5.82	2.20	5.19(0.31)	-5.55	3.58	11.68(0.65)	-3.70	2.12	5.71(0.35)	-49.24	26.13	1.38(0.03)
			w/	-0.80	0.42		-3.25	0.31		-0.92	0.37		-42.90	18.95	
	10	0.5	w/o	-4.38	0.86	2.64(0.15)	-5.49	1.72	6.45(0.47)	-4.64	0.89	3.29(0.19)	-48.45	24.15	1.12(0.01)
			w/	-2.55	0.33		-3.82	0.27		-2.89	0.27		-46.15	21.64	
		0.9	w/o	-5.01	0.76	3.12(0.17)	-5.52	1.53	6.71(0.48)	-5.09	0.81	3.95(0.19)	-49.31	24.84	1.14(0.01)
			w/	-2.87	0.24		-3.87	0.23		-3.08	0.20		-46.44	21.78	
100	1	-	w/o	0.04	1.20	1.74(0.10)	-7.86	2.46	2.99(0.20)	-2.30	1.39	2.15(0.12)	-33.96	12.49	1.00(0.02)
			w/	0.09	0.69		-4.92	0.82		-1.66	0.65		-34.34	12.54	
	2	0.5	w/o	-1.65	2.46	4.13(0.24)	-5.41	3.61	8.20(0.48)	-2.58	2.51	4.88(0.29)	-40.18	18.47	1.23(0.03)
			w/	0.88	0.59		-2.78	0.44		0.13	0.51		-37.81	15.02	
		0.9	w/o	-3.26	4.71	7.83(0.43)	-5.51	6.00	14.63(0.66)	-3.89	4.79	9.29(0.51)	-42.59	22.69	1.48(0.04)
			w/	0.59	0.60		-3.21	0.41		-0.12	0.52		-38.12	15.33	
	5	0.5	w/o	-2.61	1.80	4.24(0.24)	-5.90	2.60	9.35(0.52)	-2.98	1.84	5.39(0.29)	-44.68	21.70	1.19(0.02)
			w/	-0.38	0.42		-3.66	0.28		-1.14	0.34		-41.98	18.17	
		0.9	w/o	-4.17	2.03	5.44(0.31)	-6.26	2.75	10.89(0.51)	-4.08	1.96	6.66(0.35)	-47.07	24.03	1.35(0.03)
			w/	-0.17	0.37		-3.85	0.25		-1.15	0.29		-41.68	17.86	
	10	0.5	w/o	-4.39	0.79	2.74(0.14)	-5.04	1.52	6.50(0.41)	-3.83	0.76	3.02(0.16)	-48.04	23.69	1.11(0.01)
			w/	-2.69	0.29		-4.26	0.23		-2.82	0.25		-45.88	21.36	
		0.9	w/o	-5.06	0.75	3.38(0.17)	-5.19	1.33	6.05(0.40)	-4.42	0.70	3.54(0.18)	-49.17	24.68	1.18(0.01)
			w/	-2.50	0.22		-4.17	0.22		-2.63	0.20		-45.38	20.84	

Table 6: Bias (B) and MSE of $\mathrm{ROC}^*_{\mathbf{x}}(u)$ for $\mathbf{x} = (1, 0, \ldots, 0)^\top$ where $\mathbf{x} \in \mathbb{R}^p$, and FAR = 0.5. (w/o: linear regression without order constraint; w/: linear regression with order constraint. All values of B and MSE have been multiplied by 100.)