Supplement Materials: Collaborative Analysis for Paired A/B Testing Experiments

Qiong Zhang[†], Lulu Kang[‡] and Xinwei Deng[*]

[†] *School of Mathematical and Statistical Sciences, Clemson University*

[‡] *Department of Mathematics and Statistics, University of Massachusetts Amherst*

[*] *Department of Statistics, Virginia Tech, Blacksburg, VA*

Supplementary Material

In the supplementary materials, we provide all the technical proofs for the main results of the paper, and supplementary numerical results.

S1. Proof of Proportion 1

Let $\boldsymbol{y}_1 = (y_{1,1}, \ldots, y_{n,1})^\top$ and $\boldsymbol{y}_2 = (y_{1,2}, \ldots, y_{n,2})^\top$ be the response vectors collected from the first and second experiments. According to the true model (2.1), the covariance matrix of the combined response vector $\boldsymbol{y} = (\boldsymbol{y}_1^\top, \boldsymbol{y}_2^\top)^\top$ is given by

$$\boldsymbol{V} = \begin{bmatrix} (\sigma_1^2+\tau^2)\boldsymbol{I}_n & \tau^2\boldsymbol{I}_n \\ \tau^2\boldsymbol{I}_n & (\sigma_2^2+\tau^2)\boldsymbol{I}_n \end{bmatrix}$$

with

$$\boldsymbol{V}^{-1} = \frac{1}{(\sigma_1^2+\tau^2)(\sigma_2^2+\tau^2)-\tau^4}\begin{bmatrix} (\sigma_2^2+\tau^2)\boldsymbol{I}_n & -\tau^2\boldsymbol{I}_n \\ -\tau^2\boldsymbol{I}_n & (\sigma_1^2+\tau^2)\boldsymbol{I}_n \end{bmatrix}$$

and the covariates matrix is

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{1}_n & \boldsymbol{0} & \boldsymbol{x}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{1}_n & \boldsymbol{0} & \boldsymbol{x}_2 \end{pmatrix}$$

with $\boldsymbol{x}_k = (x_{1,k}, \ldots, x_{n,k})^\top$ be the design of the k-th experiment for $k = 1, 2$. Therefore, the weighted least squared estimator of the linear coefficients $\boldsymbol{\theta} = (\alpha_1, \alpha_2, \beta_1, \beta_2)^\top$ is

$$\hat{\boldsymbol{\theta}}^{wls} = (\hat{\alpha}_1^{wls}, \hat{\alpha}_2^{wls}, \hat{\beta}_1^{wls}, \hat{\beta}_2^{wls})^\top = (\boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{y}, \tag{S1.1}$$

with variance-covariance matrix

$$\mathrm{Var}(\hat{\boldsymbol{\theta}}^{wls}) = (\boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}. \tag{S1.2}$$

Under the orthogonal assumption

$$\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{X} = \frac{1}{(\sigma_1^2+\tau^2)(\sigma_2^2+\tau^2)-\tau^4} \begin{pmatrix} n(\sigma_2^2+\tau^2) & -n\tau^2 & 0 & 0 \\ -n\tau^2 & n(\sigma_1^2+\tau^2) & 0 & 0 \\ 0 & 0 & n(\sigma_2^2+\tau^2) & 0 \\ 0 & 0 & 0 & n(\sigma_1^2+\tau^2) \end{pmatrix}.$$

Also,

$$\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{y} = \frac{1}{(\sigma_1^2+\tau^2)(\sigma_2^2+\tau^2)-\tau^4} \begin{pmatrix} (\sigma_2^2+\tau^2)\sum_{i=1}^n y_{i,1} - \tau^2 \sum_{i=1}^n y_{i,2} \\ (\sigma_1^2+\tau^2)\sum_{i=1}^n y_{i,2} - \tau^2 \sum_{i=1}^n y_{i,1} \\ (\sigma_2^2+\tau^2)\sum_{i=1}^n x_{i,1}y_{i,1} - \tau^2 \sum_{i=1}^n x_{i,1}y_{i,2} \\ (\sigma_1^2+\tau^2)\sum_{i=1}^n x_{i,2}y_{i,2} - \tau^2 \sum_{i=1}^n x_{i,2}y_{i,1} \end{pmatrix}$$

Therefore,

$$\hat{\beta}_1^{wls} = \frac{(\sigma_2^2+\tau^2)\sum_{i=1}^n x_{i,1}y_{i,1} - \tau^2 \sum_{i=1}^n x_{i,1}y_{i,2}}{n(\sigma_2^2+\tau^2)} = \frac{\tau^2 \hat{\beta}_1^p + \sigma_2^2 \hat{\beta}_1^s}{\sigma_2^2+\tau^2} = \hat{\beta}_1^c,$$

and

$$\hat{\beta}_2^{wls} = \frac{(\sigma_1^2+\tau^2)\sum_{i=1}^n x_{i,1}y_{i,2} - \tau^2 \sum_{i=1}^n x_{i,1}y_{i,1}}{n(\sigma_1^2+\tau^2)} = \frac{\tau^2 \hat{\beta}_2^p + \sigma_1^2 \hat{\beta}_2^s}{\sigma_1^2+\tau^2} = \hat{\beta}_2^c.$$

Therefore, the hybrid estimators are equivalent to the weighted least squared estimators, i.e., the best linear unbiased estimator under the given assumptions. Thus, the conclusion in part (i) holds.

Notice that

$$\hat{\beta}_1^s = \beta_1 + \frac{1}{n}\sum_{i=1}^n x_{i,1}(\varepsilon_{i,1} + u_i), \quad \hat{\beta}_1^p = \beta_1 + \frac{1}{n}\sum_{i=1}^n x_{i,1}(\varepsilon_{i,1} - \varepsilon_{i,2})$$
$$\hat{\beta}_2^s = \beta_2 + \frac{1}{n}\sum_{i=1}^n x_{i,2}(\varepsilon_{i,2} + u_i), \quad \hat{\beta}_2^p = \beta_2 + \frac{1}{n}\sum_{i=1}^n x_{i,2}(\varepsilon_{i,2} - \varepsilon_{i,1}).$$

Consider $(x_{i,1}(\varepsilon_{i,1}+u_i), x_{i,1}(\varepsilon_{i,1}-\varepsilon_{i,2}), x_{i,2}(\varepsilon_{i,2}+u_i), x_{i,2}(\varepsilon_{i,2}-\varepsilon_{i,1}))^\top$'s as independent random vectors, the Lindbergh-Feller multivariate central limit theorem gives that, as $n \to \infty$

$$\frac{1}{\sqrt{n}}\left(\hat{\beta}_1^s - \beta_1, \hat{\beta}_1^p - \beta_1, \hat{\beta}_2^s - \beta_2, \hat{\beta}_2^p - \beta_2\right)^\top \to \mathcal{N}_4(\mathbf{0}_4, \Sigma_4),$$

in distribution with

$$\Sigma_4 = \begin{bmatrix} \sigma_1^2 + \tau^2 & \sigma_1^2 & 0 & 0 \\ \sigma_1^2 & \sigma_1^2 + \sigma_2^2 & 0 & 0 \\ 0 & 0 & \sigma_2^2 + \tau^2 & \sigma_2^2 \\ 0 & 0 & \sigma_2^2 & \sigma_1^2 + \sigma_2^2 \end{bmatrix}.$$

We can directly find the joint asymptotic normal distribution of $\hat{\beta}_1^c - \beta_1$ and $\hat{\beta}_2^c - \beta_2$ as the linear combination of $\left(\hat{\beta}_1^s - \beta_1, \hat{\beta}_1^p - \beta_1, \hat{\beta}_2^s - \beta_2, \hat{\beta}_2^p - \beta_2\right)^\top$. This completes the proof of part (ii).

S2. Proof of Proportion 2

We first provide the weighted least squared estimators of β_1 and β_2. Define the following notations:

- $\boldsymbol{y}_{1,s}$: first experiment's output for the shared users with the 2nd experiment.
- $\boldsymbol{y}_{1,ns}$: first experiment's output for the non-shared users.
- $\boldsymbol{y}_{2,s}$: 2nd experiment's output for the shared users with the first experiment.
- $\boldsymbol{y}_{2,ns}$: 2nd experiment's output for the non-shared users.

Given Table 1, the sizes of the four vectors are n_0, n_1, n_0, and n_2, respectively. Similarly, we define the random effects $\boldsymbol{u}_s$ as the ones for the shared users, of size n_0, and $\boldsymbol{u}_{1,ns}$ and $\boldsymbol{u}_{2,ns}$ are the random effects of the non-shared groups for the two experiments, of size n_1 and n_2, respectively. There are also four random noise, $\boldsymbol{\epsilon}_{1,s}$,

$\boldsymbol{\epsilon}_{1,ns}$, $\boldsymbol{\epsilon}_{2,s}$, and $\boldsymbol{\epsilon}_{2,ns}$. Based on the model assumptions in (2.1), we have that

$$\begin{bmatrix} \boldsymbol{y}_{1,s} \\ \boldsymbol{y}_{1,ns} \\ \boldsymbol{y}_{2,s} \\ \boldsymbol{y}_{2,ns} \end{bmatrix} = \begin{bmatrix} \mathbf{1}_{n_0} & \mathbf{0} & \boldsymbol{x}_{1,s} & \mathbf{0} \\ \mathbf{1}_{n_1} & \mathbf{0} & \boldsymbol{x}_{1,ns} & \\ \mathbf{0} & \mathbf{1}_{n_0} & \mathbf{0} & \boldsymbol{x}_{2,s} \\ \mathbf{0} & \mathbf{1}_{n_2} & \mathbf{0} & \boldsymbol{x}_{2,ns} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \boldsymbol{u}_s \\ \boldsymbol{u}_{1,ns} \\ \boldsymbol{u}_s \\ \boldsymbol{u}_{2,ns} \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_{1,s} \\ \boldsymbol{\epsilon}_{1,ns} \\ \boldsymbol{\epsilon}_{2,s} \\ \boldsymbol{\epsilon}_{2,ns} \end{bmatrix}$$

The covariance matrix of the responses is

$$\boldsymbol{V} = \begin{bmatrix} (\sigma_1^2+\tau^2)\boldsymbol{I}_{n_0} & \mathbf{0} & \tau^2\boldsymbol{I}_{n_0} & \mathbf{0} \\ \mathbf{0} & (\sigma_1^2+\tau^2)\boldsymbol{I}_{n_1} & \mathbf{0} & \mathbf{0} \\ \tau^2\boldsymbol{I}_{n_0} & \mathbf{0} & (\sigma_2^2+\tau^2)\boldsymbol{I}_{n_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & (\sigma^2+\tau^2)\boldsymbol{I}_{n_2} \end{bmatrix}$$

with inverse $\boldsymbol{V}^{-1}$

$$\begin{bmatrix} \left(\sigma_1^2+\frac{\sigma_2^2\tau^2}{\sigma_2^2+\tau^2}\right)^{-1}\boldsymbol{I}_{n_0} & \mathbf{0} & -\left[(\sigma_1^2+\tau^2)(\sigma_2^2+\tau^2)/\tau^2-\tau^2\right]^{-1}\boldsymbol{I}_{n_0} & \mathbf{0} \\ \mathbf{0} & (\sigma_1^2+\tau^2)^{-1}\boldsymbol{I}_{n_1} & \mathbf{0} & \mathbf{0} \\ -\left[(\sigma_1^2+\tau^2)(\sigma_2^2+\tau^2)/\tau^2-\tau^2\right]^{-1}\boldsymbol{I}_{n_0} & \mathbf{0} & \left[(\sigma_2^2+\tau^2)-\frac{\tau^4}{\sigma_1^2+\tau^2}\right]^{-1}\boldsymbol{I}_{n_0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & (\sigma_2^2+\tau^2)^{-1}\boldsymbol{I}_{n_2} \end{bmatrix}$$

Then we have that $\boldsymbol{X}^\top \boldsymbol{V}^{-1}\boldsymbol{X} =$

$$\begin{bmatrix} (an_0+bn_1) & en_0 & a\mathbf{1}_{n_0}^\top \boldsymbol{x}_{1,s} + b\mathbf{1}_{n_1}^\top \boldsymbol{x}_{1,ns} & e\mathbf{1}_{n_0}^\top \boldsymbol{x}_{2,s} \\ en_0 & cn_0 + dn_2 & e\mathbf{1}_{n_0}^\top \boldsymbol{x}_{1,s} & c\mathbf{1}_{n_0}^\top \boldsymbol{x}_{2,s} + dc\mathbf{1}_{n_2}^\top \boldsymbol{x}_{2,ns} \\ a\mathbf{1}_{n_0}^\top \boldsymbol{x}_{1,s} + b\mathbf{1}_{n_1}^\top \boldsymbol{x}_{1,ns} & e\mathbf{1}_{n_0}^\top \boldsymbol{x}_{1,s} & (an_0+bn_1) & e\boldsymbol{x}_{1,s}^\top \boldsymbol{x}_{2,s} \\ e\mathbf{1}_{n_2}^\top \boldsymbol{x}_{2,s} & c\mathbf{1}_{n_0}^\top \boldsymbol{x}_{2,s} + dc\mathbf{1}_{n_2}^\top \boldsymbol{x}_{2,ns} & e\boldsymbol{x}_{1,s}^\top \boldsymbol{x}_{2,s} & cn_0 + dn_2 \end{bmatrix}.$$

Here $a = \left(\sigma_1^2 + \frac{\sigma_2^2\tau^2}{\sigma_2^2+\tau^2}\right)^{-1}$, $b = (\sigma_1^2+\tau^2)^{-1}$, $c = \left[(\sigma_2^2+\tau^2) - \frac{\tau^4}{\sigma_1^2+\tau^2}\right]^{-1}$, $d = (\sigma_2^2+\tau^2)^{-1}$, and $e = -\left[(\sigma_1^2+\tau^2)(\sigma_2^2+\tau^2)/\tau^2 - \tau^2\right]^{-1}$. Also, we have that

$$\boldsymbol{X}^\top \boldsymbol{V}^{-1}\boldsymbol{y} = \begin{bmatrix} a\mathbf{1}_{n_0}^\top \boldsymbol{y}_{1,s} + b\mathbf{1}_{n_1}^\top \boldsymbol{y}_{1,ns} + e\mathbf{1}_{n_0}^\top \boldsymbol{y}_{2,s} \\ e\mathbf{1}_{n_0}^\top \boldsymbol{y}_{1,s} + c\mathbf{1}_{n_0}^\top \boldsymbol{y}_{2,s} + d\mathbf{1}_{n_2}^\top \boldsymbol{y}_{2,ns} \\ a\boldsymbol{x}_{1,s}^\top \boldsymbol{y}_{1,s} + b\boldsymbol{x}_{1,ns}^\top \boldsymbol{y}_{1,ns} + e\boldsymbol{x}_{1,s}^\top \boldsymbol{y}_{2,s} \\ e\boldsymbol{x}_{2,s}^\top \boldsymbol{y}_{1,s} + c\boldsymbol{x}_{2,s}^\top \boldsymbol{y}_{2,s} + d\boldsymbol{x}_{2,ns}^\top \boldsymbol{y}_{2,ns} \end{bmatrix}.$$

Then the weighted least squared estimator is given by

$$(\hat{\alpha}_1^{wls}, \hat{\alpha}_2^{wls}, \hat{\beta}_1^{wls}, \hat{\beta}_2^{wls})^\top = \left(\boldsymbol{X}^\top \boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1} \boldsymbol{X}^\top \boldsymbol{V}^{-1}\boldsymbol{y}.$$

Let

$$\boldsymbol{A} = \begin{bmatrix} an_0 + bn_1 & en_0 \\ en_0 & cn_0 + dn_2 \end{bmatrix}$$

$$\boldsymbol{B} = \begin{bmatrix} a\mathbf{1}_{n_0}^\top \boldsymbol{x}_{1,s} + b\mathbf{1}_{n_1}^\top \boldsymbol{x}_{1,ns} & e\mathbf{1}_{n_0}^\top \boldsymbol{x}_{2,s} \\ e\mathbf{1}_{n_0}^\top \boldsymbol{x}_{1,s} & c\mathbf{1}_{n_0}^\top \boldsymbol{x}_{2,s} + dc\mathbf{1}_{n_2}^\top \boldsymbol{x}_{2,ns} \end{bmatrix}$$

$$\boldsymbol{D} = \begin{bmatrix} an_0 + bn_1 & e\boldsymbol{x}_{1,s}^\top \boldsymbol{x}_{2,s} \\ e\boldsymbol{x}_{1,s}^\top \boldsymbol{x}_{2,s} & cn_0 + dn_2 \end{bmatrix}.$$

We have that

$$\mathrm{Var}\left(\begin{bmatrix} \hat{\beta}_1^{wle} \\ \hat{\beta}_2^{wle} \end{bmatrix}\right) = \left(\boldsymbol{D} - \boldsymbol{B}^\top \boldsymbol{A}^{-1} \boldsymbol{B}\right)^{-1}$$

Thus

$$\left[n\mathrm{Var}\left(\begin{bmatrix} \hat{\beta}_1^{wle} \\ \hat{\beta}_2^{wle} \end{bmatrix}\right)\right]^{-1} = \frac{1}{n}\boldsymbol{D} - \left(\frac{1}{n}\boldsymbol{B}^\top\right)\left(\frac{1}{n}\boldsymbol{A}\right)^{-1}\left(\frac{1}{n}\boldsymbol{B}\right)$$

Note that, under Assumption 2, as $n \to \infty$

$$\frac{1}{n}\boldsymbol{A} \to \begin{bmatrix} ar_0 + br_1 & er_0 \\ er_0 & cr_0 + dr_2 \end{bmatrix}$$

and

$$\frac{1}{n}\boldsymbol{D} \rightarrow \begin{bmatrix} ar_0 + br_1 & 0 \\ 0 & cr_0 + dr_2 \end{bmatrix}$$

which are constant matrices. Under Assumption 3, we have that

$$\frac{1}{n}\boldsymbol{B} \rightarrow \boldsymbol{0}$$

as $n \rightarrow \infty$. Therefore,

$$\left[n\mathrm{Var}\left(\begin{bmatrix} \hat{\beta}_1^{wle} \\ \hat{\beta}_2^{wle} \end{bmatrix} \right) \right]^{-1} \rightarrow \begin{bmatrix} ar_0 + br_1 & 0 \\ 0 & cr_0 + dr_2 \end{bmatrix}$$

and

$$n\mathrm{Var}\left(\begin{bmatrix} \hat{\beta}_1^{wle} \\ \hat{\beta}_2^{wle} \end{bmatrix} \right) \rightarrow \begin{bmatrix} 1/(ar_0 + br_1) & 0 \\ 0 & 1/(cr_0 + dr_2) \end{bmatrix} \tag{S2.3}$$

as $n \rightarrow \infty$. Also, we have that

$$\begin{bmatrix} \hat{\beta}_1^{wle} \\ \hat{\beta}_2^{wle} \end{bmatrix} = \mathrm{Var}\left(\begin{bmatrix} \hat{\beta}_1^{wle} \\ \hat{\beta}_2^{wle} \end{bmatrix} \right) \begin{bmatrix} -\boldsymbol{B}^\top \boldsymbol{A}^{-1} & \boldsymbol{I}_2 \end{bmatrix} \boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{y} \tag{S2.4}$$

$$= n\mathrm{Var}\left(\begin{bmatrix} \hat{\beta}_1^{wle} \\ \hat{\beta}_2^{wle} \end{bmatrix} \right) \begin{bmatrix} -\left(n^{-1}\boldsymbol{B}^\top\right)\left(n^{-1}\boldsymbol{A}\right)^{-1} & \boldsymbol{I}_2 \end{bmatrix} \frac{1}{n}\boldsymbol{X}^\top \boldsymbol{V}^{-1} \boldsymbol{y} \tag{S2.5}$$

Also, we can express the collaborative estimators by

$$\hat{\beta}_1^c = \frac{-e\boldsymbol{x}_{1,s}^\top(\boldsymbol{y}_{1,s} - \boldsymbol{y}_{2,s}) + b\boldsymbol{x}_{1,ns}^\top\boldsymbol{y}_{1,ns} + (a+e)\boldsymbol{x}_{1,s}^\top\boldsymbol{y}_{1,s}}{an_0 + bn_1} = \frac{a\boldsymbol{x}_{1,s}^\top\boldsymbol{y}_{1,s} + b\boldsymbol{x}_{1,ns}^\top\boldsymbol{y}_{1,ns} + e\boldsymbol{x}_{1,s}^\top\boldsymbol{y}_{2,s}}{an_0 + bn_1},$$

and

$$\hat{\beta}_2^c = \frac{-e\boldsymbol{x}_{2,s}^\top(\boldsymbol{y}_{2,s} - \boldsymbol{y}_{1,s}) + d\boldsymbol{x}_{2,ns}^\top\boldsymbol{y}_{2,ns} + (c+e)\boldsymbol{x}_{2,s}^\top\boldsymbol{y}_{2,s}}{cn_0 + dn_2} = \frac{e\boldsymbol{x}_{2,s}^\top\boldsymbol{y}_{1,s} + c\boldsymbol{x}_{2,s}^\top\boldsymbol{y}_{2,s} + d\boldsymbol{x}_{2,ns}^\top\boldsymbol{y}_{2,ns}}{cn_0 + dn_2}.$$

We can then express

$$\begin{bmatrix} \hat{\beta}_1^c \\ \hat{\beta}_2^c \end{bmatrix} = n \begin{bmatrix} 0 & 0 & 1/(an_0 + bn_1) & 0 \\ 0 & 0 & 0 & 1/(cn_0 + dn_2) \end{bmatrix} \frac{1}{n}\boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{y}$$

Let

$$\boldsymbol{M}_n = n\mathrm{Var}\left(\begin{bmatrix} \hat{\beta}_1^{wle} \\ \hat{\beta}_2^{wle} \end{bmatrix}\right) \begin{bmatrix} \left(n^{-1}\boldsymbol{B}^\top\right)\left(n^{-1}\boldsymbol{A}\right)^{-1} & \boldsymbol{I}_2 \end{bmatrix} - n \begin{bmatrix} 0 & 0 & 1/(an_0 + bn_1) & 0 \\ 0 & 0 & 0 & 1/(cn_0 + dn_2) \end{bmatrix}$$

Then $\boldsymbol{M}_n \to 0$ as $n \to \infty$. Therefore,

$$\begin{bmatrix} \hat{\beta}_1^{wle} \\ \hat{\beta}_2^{wle} \end{bmatrix} - \begin{bmatrix} \hat{\beta}_1^c \\ \hat{\beta}_2^c \end{bmatrix} = \boldsymbol{M}_n \cdot \frac{1}{n}\boldsymbol{X}^\top\boldsymbol{V}^{-1}\boldsymbol{y} \to 0$$

as $n \to \infty$, which shows the asymptotic equivalence between the weighted least squared estimators and the collaborative estimators. This completes the proof of part (i).

For part (ii), let $r_{i,k}$ be a 0-1 value indicating whether or not the k-th response from the i-th user is available. Therefore, we can express

$$
\begin{aligned}
n_0^{-1}\sum_{i=1}^{n_0} x_{i,1}y_{i,1} &= \frac{n}{n_0}\cdot n^{-1}\sum_{i=1}^{n} x_{i,1}r_{i,1}r_{i,2}y_{i,1} \\
&= \beta_1 + \frac{n}{n_0}\cdot\frac{\sum_{i=1}^{n_0} x_{i,1}}{n}\alpha + \frac{n}{n_0}\cdot n^{-1}\sum_{i=1}^{n} x_{i,1}r_{i,1}r_{i,2}(\varepsilon_{i,1}+u_i) \\
n_0^{-1}\sum_{i=1}^{n_0} x_{i,1}z_i &= \beta_1 + \frac{n}{n_0}\cdot\frac{\sum_{i=1}^{n_0} x_{i,1}}{n}\alpha - \frac{n}{n_0}\cdot\frac{\sum_{i=1}^{n_0} x_{i,1}x_{i,2}}{n}\beta_2 + \frac{n}{n_0}\cdot n^{-1}\sum_{i=1}^{n} x_{i,1}r_{i,1}r_{i,2}(\varepsilon_{i,1}-\varepsilon_{i,2}) \\
n_1^{-1}\sum_{i=n_0+1}^{n_0+n_1} x_{i,1}y_{i,1} &= \beta_1 + \frac{n}{n_1}\cdot\frac{\sum_{i=n_0+1}^{n_0+n_1} x_{i,1}}{n}\alpha + \frac{n}{n_1}\cdot n^{-1}\sum_{i=1}^{n} x_{i,1}r_{i,1}(1-r_{i,2})(\varepsilon_{i,1}+u_i)
\end{aligned}
$$

also, for estimating β_2, we have

$$
\begin{aligned}
n_0^{-1}\sum_{i=1}^{n_0} x_{i,2}y_{i,2} &= \frac{n}{n_0}\cdot n^{-1}\sum_{i=1}^{n} x_{i,2}r_{i,1}r_{i,2}y_{i,2} \\
&= \beta_2 + \frac{n}{n_0}\cdot\frac{\sum_{i=1}^{n_0} x_{i,2}}{n}\alpha + \frac{n}{n_0}\cdot n^{-1}\sum_{i=1}^{n} x_{i,2}r_{i,1}r_{i,2}(\varepsilon_{i,2}+u_i) \\
n_0^{-1}\sum_{i=1}^{n_0} x_{i,2}z_i &= \beta_2 + \frac{n}{n_0}\cdot\frac{\sum_{i=1}^{n_0} x_{i,2}}{n}\alpha - \frac{n}{n_0}\cdot\frac{\sum_{i=1}^{n_0} x_{i,1}x_{i,2}}{n}\beta_1 + \frac{n}{n_0}\cdot n^{-1}\sum_{i=1}^{n} x_{i,2}r_{i,1}r_{i,2}(\varepsilon_{i,2}-\varepsilon_{i,1}) \\
n_2^{-1}\sum_{i=n_0+n_1+1}^{n_0+n_1+n_2} x_{i,2}y_{i,2} &= \beta_2 + \frac{n}{n_2}\cdot\frac{\sum_{i=n_0+n_1+1}^{n_0+n_1+n_2} x_{i,2}}{n}\alpha + \frac{n}{n_2}\cdot n^{-1}\sum_{i=1}^{n} x_{i,2}r_{i,2}(1-r_{i,1})(\varepsilon_{i,2}+u_i)
\end{aligned}
$$

First, we apply the Lindbergh-Feller's multivariate central limit theorem to $n^{-1}\sum_{i=1}^{n} x_{i,1}r_{i,1}r_{i,2}(\varepsilon_{i,1}+u_i)$, $n^{-1}\sum_{i=1}^{n} x_{i,1}r_{i,1}r_{i,2}(\varepsilon_{i,1}-\varepsilon_{i,2})$, $n^{-1}\sum_{i=1}^{n} x_{i,1}r_{i,1}(1-r_{i,2})(\varepsilon_{i,1}+u_i)$, $n^{-1}\sum_{i=1}^{n} x_{i,2}r_{i,1}r_{i,2}(\varepsilon_{i,2}+u_i)$, $n^{-1}\sum_{i=1}^{n} x_{i,2}r_{i,1}r_{i,2}(\varepsilon_{i,2}-\varepsilon_{i,1})$, and $n^{-1}\sum_{i=1}^{n} x_{i,2}r_{i,2}(1-r_{i,1})(\varepsilon_{i,2}+u_i)$. Under Assumptions 2-3, the resulting multivariate asymptotic distribution has mean zero and variance-covariance matrix:

$$\begin{bmatrix} r_0(\sigma_1^2+\tau^2) & r_0\sigma_1^2 & 0 & 0 & 0 & 0 \\ r_0\sigma_1^2 & r_0(\sigma_1^2+\sigma_2^2) & 0 & 0 & 0 & 0 \\ 0 & 0 & r_1(\sigma_1^2+\tau^2) & 0 & 0 & 0 \\ 0 & 0 & 0 & r_0(\sigma_2^2+\tau^2) & r_0\sigma_2^2 & 0 \\ 0 & 0 & 0 & r_0\sigma_2^2) & r_0(\sigma_1^2+\sigma_2^2) & 0 \\ 0 & 0 & 0 & 0 & 0 & r_2(\sigma_2^2+\tau^2) \end{bmatrix}.$$

Then we have that

$$\sqrt{n}\begin{bmatrix} \frac{\sum_{i=1}^{n_0} x_{i,1}y_{i,1}}{n_0} - \beta_1 \\ \frac{\sum_{i=1}^{n_0} x_{i,1}z_i}{n_0} - \beta_1 \\ \frac{\sum_{i=n_0+1}^{n_0+n_1} x_{i,1}y_{i,1}}{n_1} - \beta_1 \\ \frac{\sum_{i=1}^{n_0} x_{i,2}y_{i,2}}{n_0} - \beta_2 \\ \frac{\sum_{i=1}^{n_0} x_{i,2}z_i}{n_0} - \beta_2 \\ \frac{\sum_{i=n_0+n+1+1}^{n_0+n_1+n_2} x_{i,2}y_{i,2}}{n_1} - \beta_2 \end{bmatrix} \xrightarrow{d} \mathcal{N}_6\left(\mathbf{0}_6, \boldsymbol{V}_6\right),$$

with

$$\begin{bmatrix} r_0^{-1}(\sigma_1^2+\tau^2) & r_0^{-1}\sigma_1^2 & 0 & 0 & 0 & 0 \\ r_0^{-1}\sigma_1^2 & r_0^{-1}(\sigma_1^2+\sigma_2^2) & 0 & 0 & 0 & 0 \\ 0 & 0 & r_1^{-1}(\sigma_1^2+\tau^2) & 0 & 0 & 0 \\ 0 & 0 & 0 & r_0^{-1}(\sigma_2^2+\tau^2) & r_0^{-1}\sigma_2^2 & 0 \\ 0 & 0 & 0 & r_0^{-1}\sigma_2^2 & r_0^{-1}(\sigma_1^2+\sigma_2^2) & 0 \\ 0 & 0 & 0 & 0 & 0 & r_2^{-1}(\sigma_2^2+\tau^2) \end{bmatrix},$$

which indicates the joint asymptotic normal distribution of $\hat{\beta}_1^c$ and $\hat{\beta}_2^c$.

S3. Bias and Variance for Different Types of Responses

In Figures 1-3, we provide the estimated bias corresponding to Figures 3-5 in the main paper. Following (4.18) and (4.21), the estimated bias is given by

$$\text{Bias} = 100^{-1}\sum_{l=1}^{100}(\hat{\beta}_1^l - \beta_1)$$

for continuous responses in Figure 1, and

$$\text{Bias} = 100^{-1}\sum_{l=1}^{100}(\hat{\beta}_1^l - \tilde{\beta}_1)$$

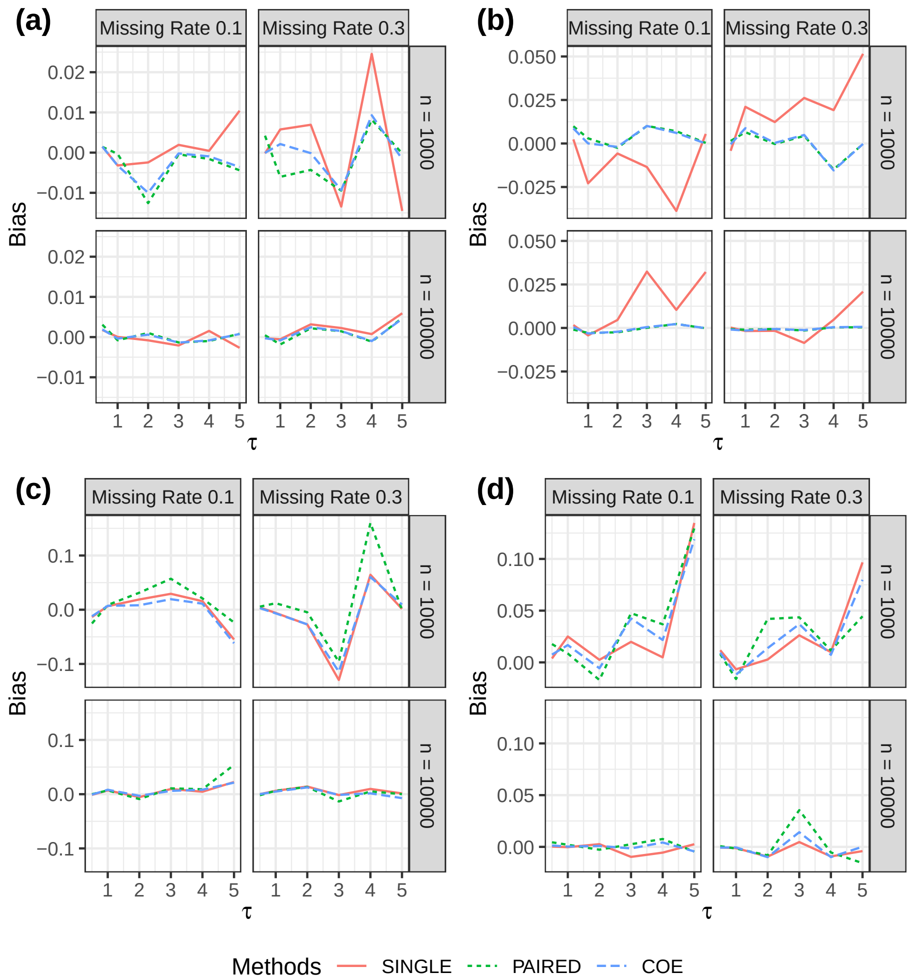

Figure 1: Bias under user effect settings (a)-(d) for continuous responses.

for binary or count responses in Figure 2 or 3, respectively. These figures show that bias of the three methods are around zero for all three methods. Overall, no method appears to exhibit a clear advantage or disadvantage in terms of bias.

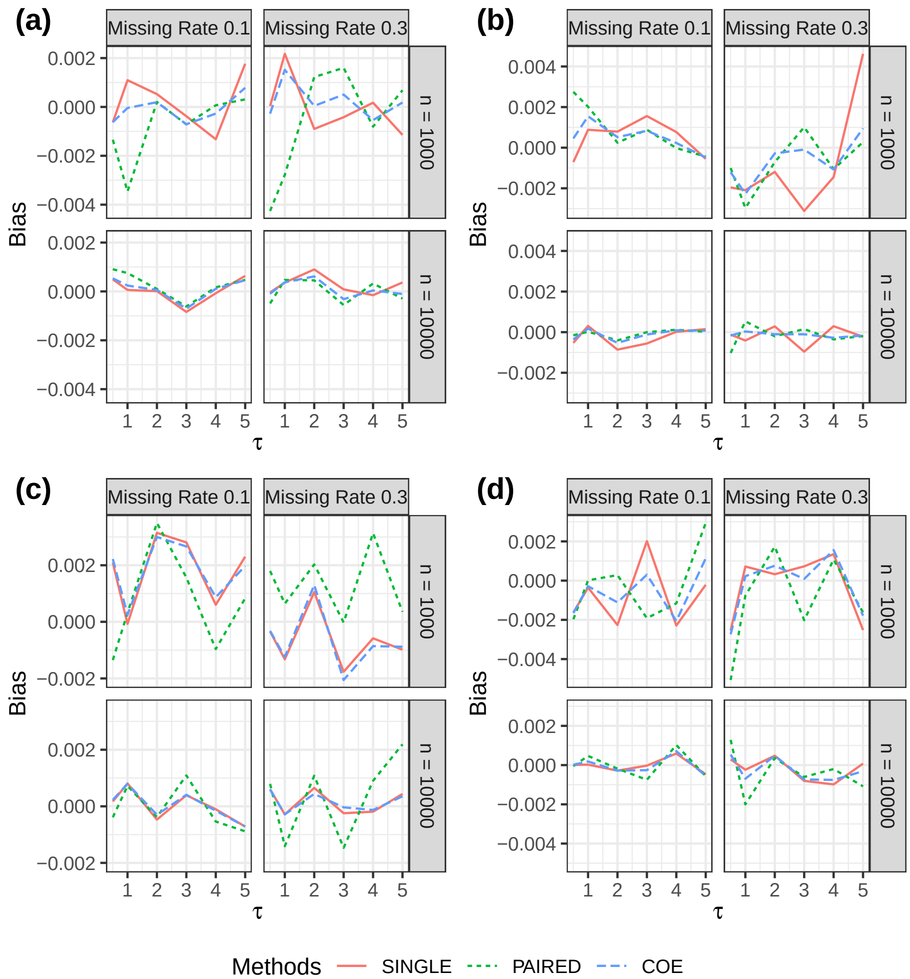

Figure 2: Bias under user effect settings (a)-(d) for binary responses.

In Figures 4-6, we provide the sample variance ratio with respect to the single experiment estimator associated with Figures 3-5 in the main paper. Following (4.18) and (4.21), the variance ratio is given by

$$\text{Var.ratio} = \frac{99^{-1}\sum_{l=1}^{100}(\hat{\beta}_1^l - 100^{-1}\sum_{l=1}^{100}\hat{\beta}_1^l)^2}{99^{-1}\sum_{l=1}^{100}(\hat{\beta}_1^{s,l} - 100^{-1}\sum_{l=1}^{100}\hat{\beta}_1^{s,l})^2},$$

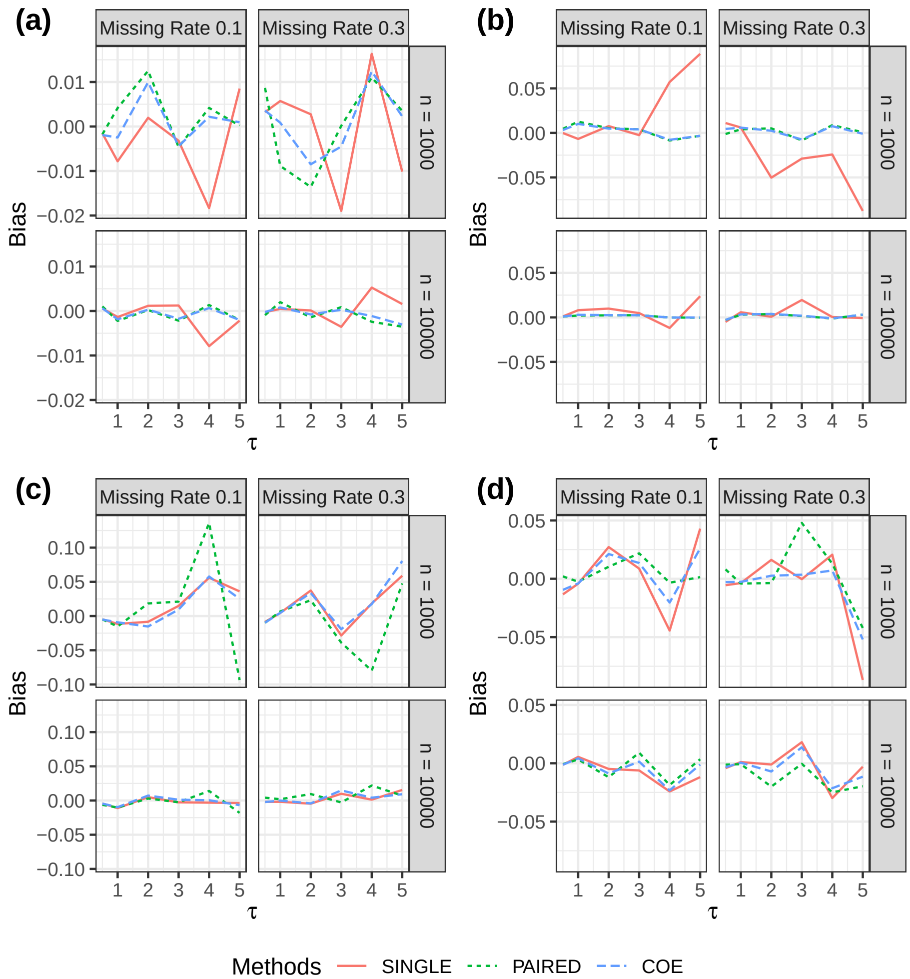

Figure 3: Bias under user effect settings (a)-(d) for integer/count responses.

for continuous, binary and count responses in Figures 4-6. The variance behavior of the three methods is similar to the patterns observed for the mean squared errors in Figures 3-5 of the main paper.

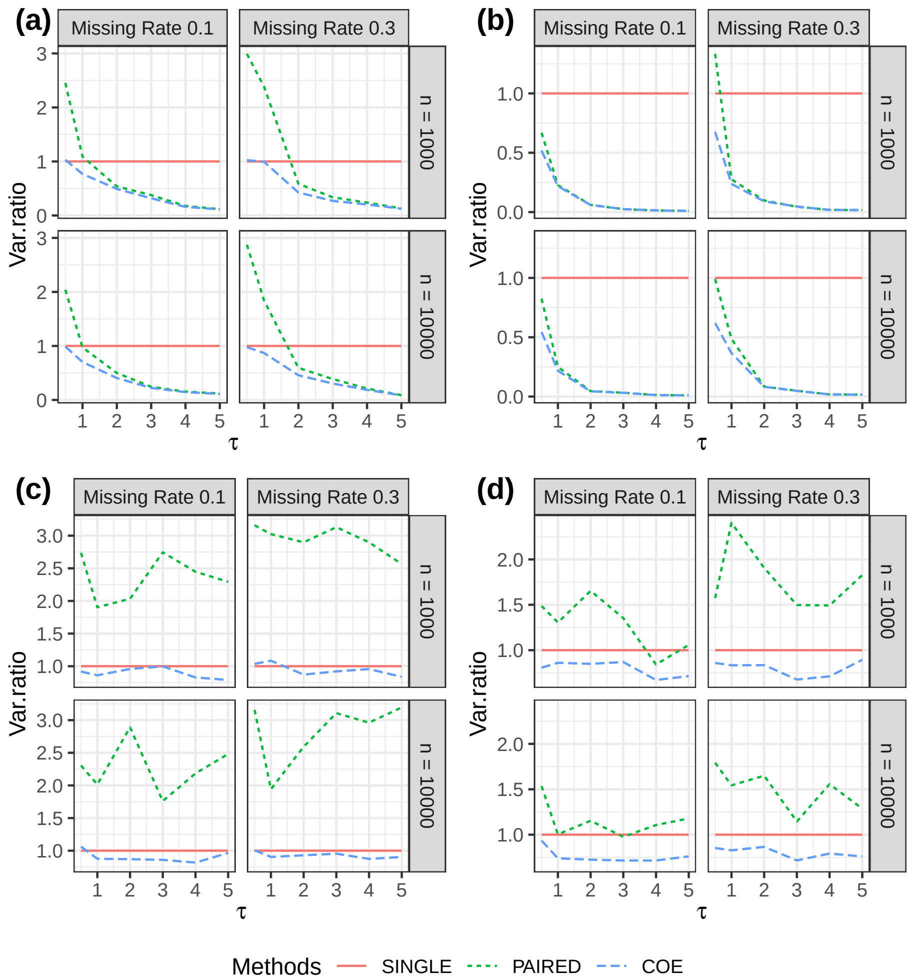

Figure 4: Var.ratio under user effect settings (a)-(d) for continuous responses.

S4. Confidence Intervals

We provide statistical inference results for SINGLE, PAIRED and COE estimators. Under the asymptotic normal distributions of those estimators, the confidence interval can be easily constructed with estimated variances given in Section 3 of the main paper. In Figures 7-9, we report the coverage rates of 95% confidence intervals

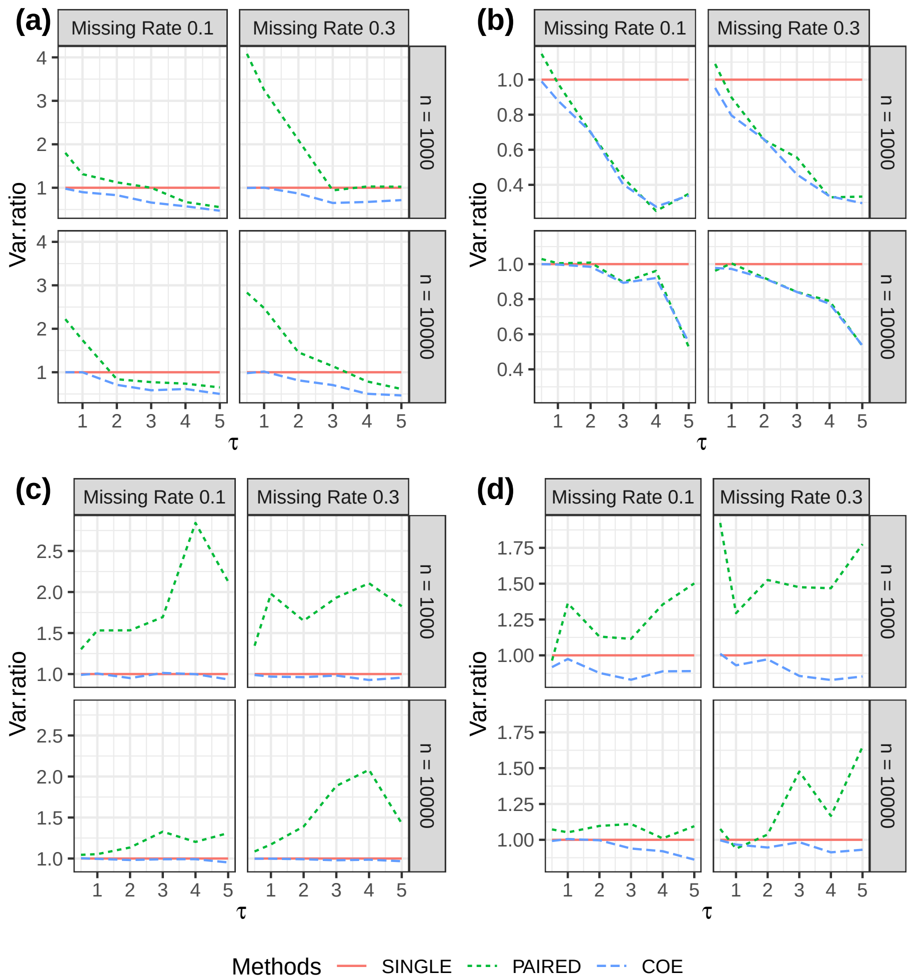

Figure 5: Var.ratio under user effect settings (a)-(d) for binary responses.

across 100 micro-replications associated with Figures 3-5 in the main paper. The coverage rates of the confidence intervals for all three methods are approximately 95%, indicating that statistical inference based on these methods is effective. However, since COE has the smallest variance, it provides the shortest confidence intervals among the three methods.

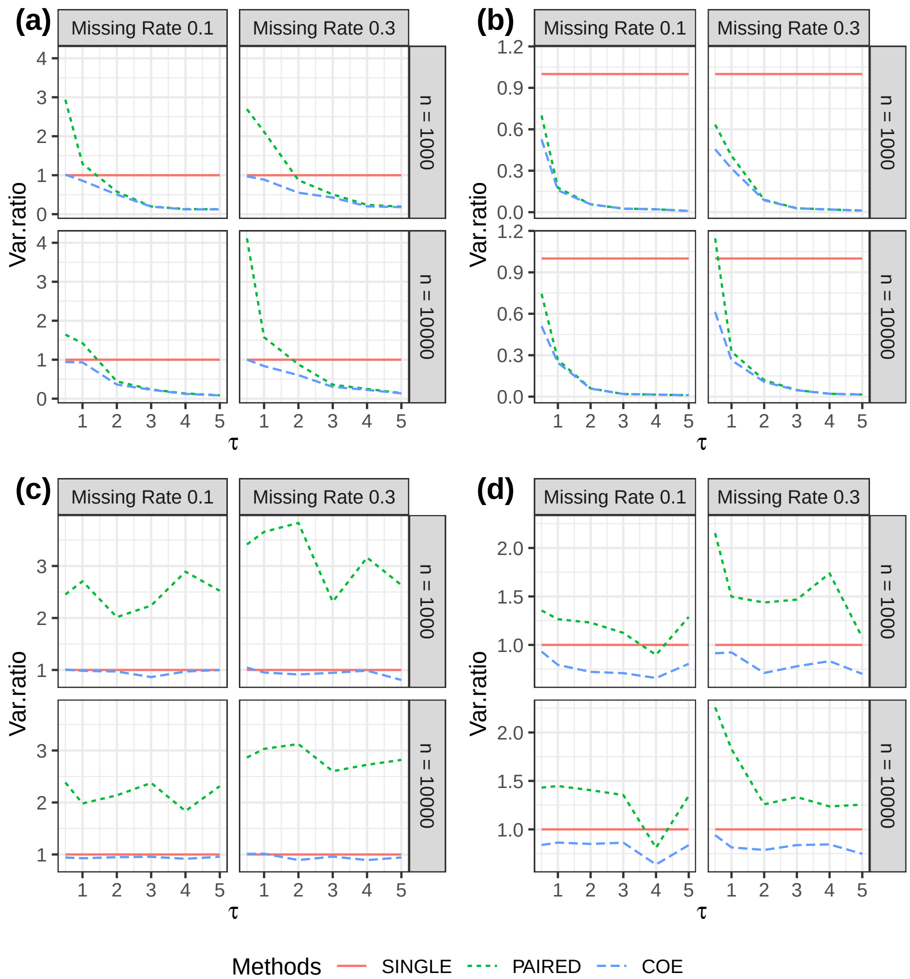

Figure 6: Var.ratio under user effect settings (a)-(d) for integer/count responses.

S5. Collaborative experiments with mixed responses

In this section, we evaluate the performance of COE with mixed responses. Particularly, we generate a pair of experiments under the model assumption in (2.1). The outcomes of the first experiment are converted to binary responses as in (4.19), whereas the outcomes of the second experiments are maintained the same as the

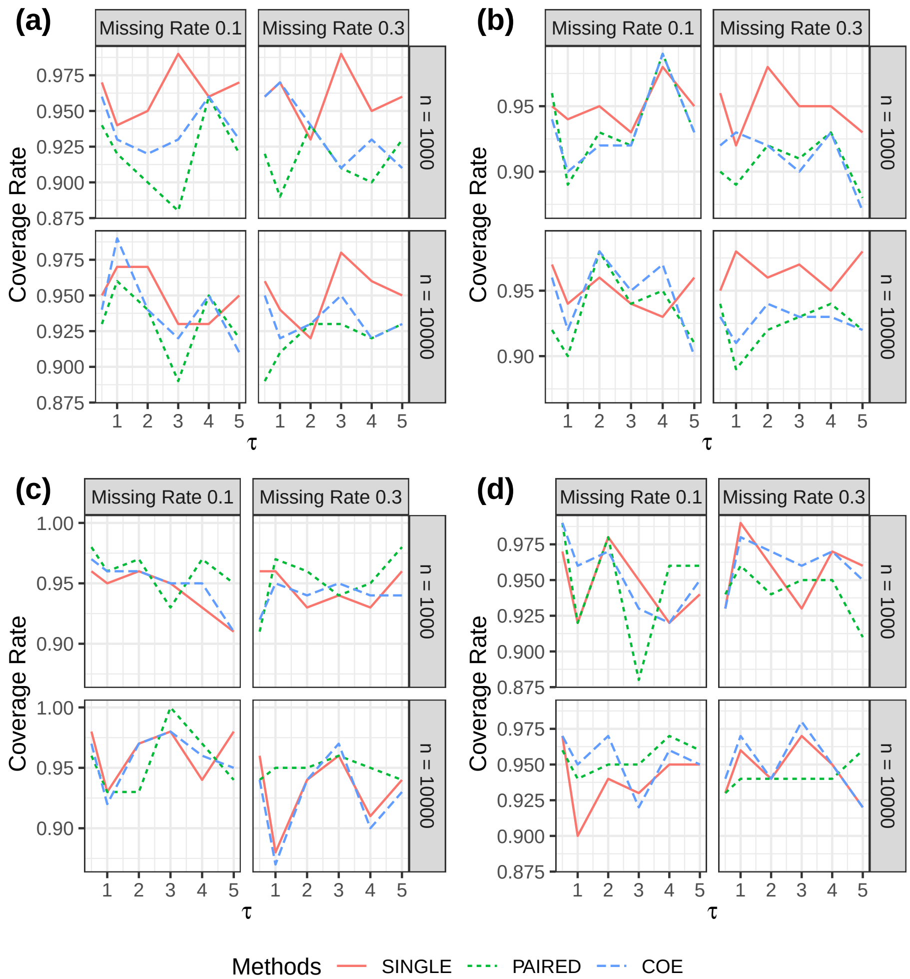

Figure 7: Coverage rate under user effect settings (a)-(d) for continuous responses.

original continuous values. The results on the logarithm of MSE.ratio in (4.18) for the estimated treatment effects for binary responses (first experiment) and continuous responses (second experiment) are provided in Figure 10. Overall, COE has the smallest mean squared errors among three methods.

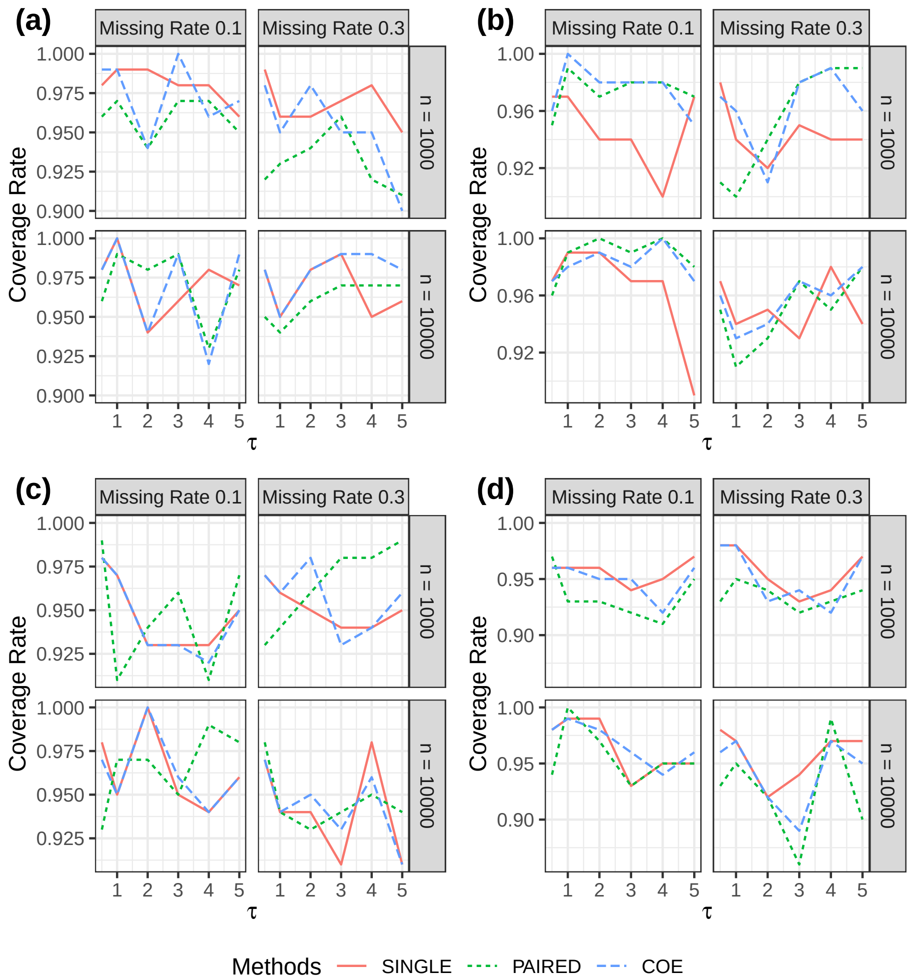

Figure 8: Coverage rate under user effect settings (a)-(d) for binary responses.

S6. Robust Analysis to Model Misspecification

In the main body of the paper, we assume that the underlying model of the outcome is the model in (2.1), i.e.,

$$y_{i,k} = u_i + \alpha_k + x_{i,k}\beta_k + \epsilon_{i,k}, \quad k = 1, 2,$$

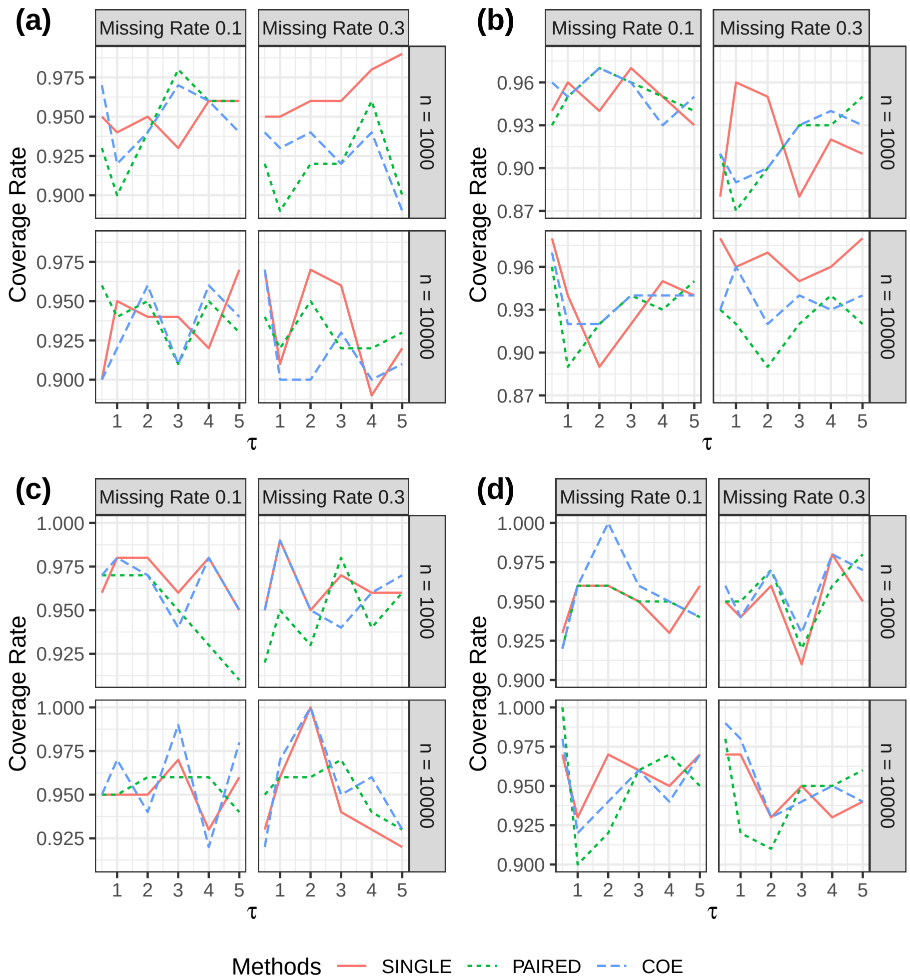

Figure 9: Coverage rate of 95% Confidence Intervals under user effect settings (a)-(d) for integer/count responses.

where the individual effect u_i is a random effect with mean zero and variance τ^2. In Section 4.2, to show the robustness of the COE to random effect, we assume four different models (a)–(d) that generate the data. Our simulation results show that the proposed COE has smaller MSE compared to the SINGLE and PAIRED estimators.

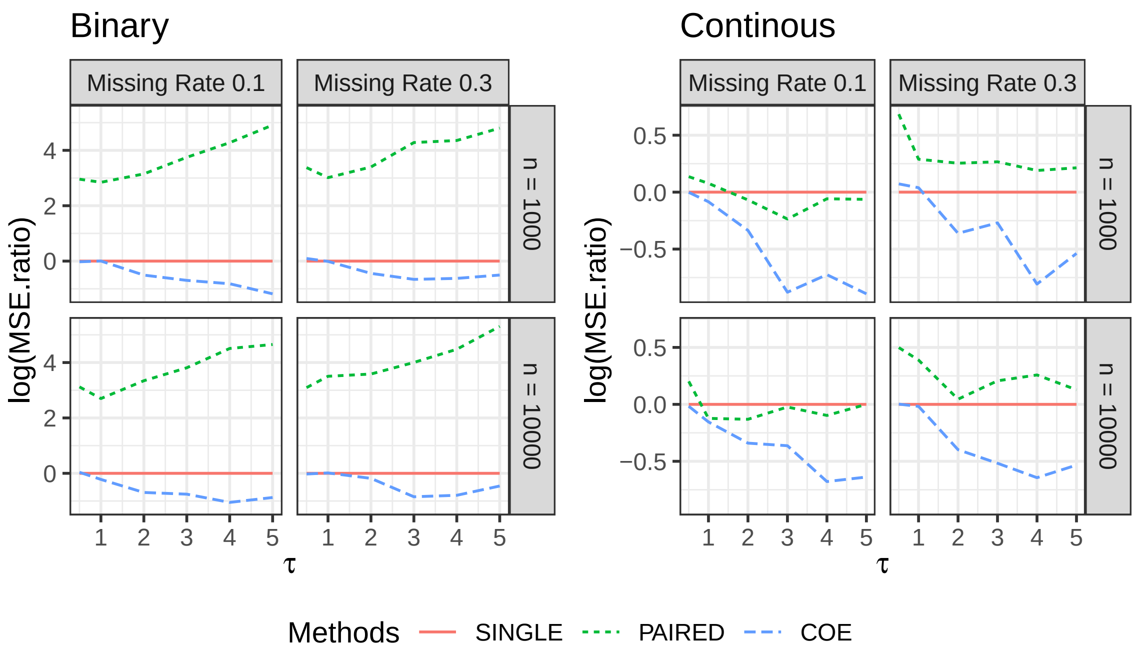

Figure 10: Logarithm of MSE.ratio in (4.18) for a paired of experiments with binary responses (left) and continuous responses (right).

Here, we provide a theoretical justification on the robustness of the COE with respect to the model misspecification on the random effect assumption.

To facilitate the discussion, we only focus on the case when the two experiments are exactly overlapped, i.e., $r_0 = 1$ and $n_0 = n$ in Assumption 2. We assume a more general model

$$y_{i,k} = u_{ik+}I(x_{i,k} = 1) + u_{ik-}I(x_{ik} = -1) + \alpha_k + x_{i,k}\beta_k + \epsilon_{i,k}, \quad k = 1, 2, \qquad \text{(S6.6)}$$

where u_{ik+} and u_{ik-} are independent random effects across different users and the joint distribution of $\boldsymbol{u}_i = (u_{i1+}, u_{i1-}, u_{i2+}, u_{i2-})^\top$ has mean zero and covariance Σ_i with off-diagonal (l, h)-th entry $v_{i,lh}$ and diagonal entries $\tau^2_{i,1+}, \tau^2_{i,1-}, \tau^2_{i,2+}, \tau^2_{i,2-}$. Other terms of (S6.6) are defined in the same way as in model (2.1), such as $\text{Var}(\epsilon_{i,k}) = \sigma^2_k$ for $k = 1, 2$. It is straightforward to see that this model (S6.6) covers all the scenarios in (a)–(d) in Section 4.2.

Note that the two experiments are exactly overlapped. Similar to the **Single Analysis** in Section 2, we can easily derive the SINGLE estimator and its variance under (S6.6)

$$\begin{aligned}
\hat{\beta}^s_1 &= n^{-1} \sum_{i=1}^n x_{i,1} y_{i,1}, \\
\text{Var}(\hat{\beta}^s_1) &= \frac{\sigma^2_1}{n} + n^{-2} \sum_{i=1}^n (\tau^2_{i,1+} I(x_{i1} = 1) + \tau^2_{i,1-} I(x_{i1} = -1)) \\
&= \frac{\sigma^2_1}{n} + n^{-2} \sum_{i=1}^n \boldsymbol{a}_i^\top \Sigma_i \boldsymbol{a}_i,
\end{aligned}$$

where $\boldsymbol{a}_i = (I(x_{i,1} = 1), I(x_{i1} = -1), 0, 0)^\top$. Also, similarly to the **Paired Analysis** in Section 2, the difference model is

$$z_i = \alpha + x_{i,1}\beta_1 - x_{i,2}\beta_2 + \delta_i \tag{S6.7}$$

where

$$\delta_i = \varepsilon_{i1} - \varepsilon_{i2} + u_{i1+}I(x_{i,1} = 1) + u_{i1-}I(x_{i1} = -1) - u_{i2+}I(x_{i,2} = 1) - u_{i2-}I(x_{i2} = -1),$$

$$= \varepsilon_{i1} - \varepsilon_{i2} + \boldsymbol{c}_i^\top \boldsymbol{u}_i,$$

where $\boldsymbol{c}_i = (I(x_{i,1} = 1), I(x_{i1} = -1), -I(x_{i,2} = 1), -I(x_{i2} = -1))^\top$. Thus, the paired analysis estimator is given by

$$\hat{\beta}_1^p = n^{-1} \sum_{i=1}^n x_{i,1} z_i. \tag{S6.8}$$

with

$$\mathrm{Var}\left(\hat{\beta}_1^p\right) = \frac{\sigma_1^2 + \sigma_2^2}{n} + \frac{1}{n^2} \sum_{i=1}^n \boldsymbol{c}_i^\top \Sigma_i \boldsymbol{c}_i$$

Furthermore,

$$\mathrm{cov}\left(\hat{\beta}_1^s, \hat{\beta}_1^p\right) = \frac{\sigma_1^2}{n} + \frac{1}{n^2} \sum_{i=1}^n \boldsymbol{c}_i^\top \Sigma_i \boldsymbol{a}_i.$$

Therefore, we have that the joint distribution of $(\hat{\beta}_1^s, \hat{\beta}_1^p)$ has mean $\beta_1 \mathbf{1}_2$ and covariance matrix

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12} & \Sigma_{22} \end{bmatrix} = n^{-1} \begin{bmatrix} \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 + \sigma_2^2 \end{bmatrix} + n^{-2} \sum_{i=1}^n [\boldsymbol{a}_i \quad \boldsymbol{c}_i]^\top \Sigma_i [\boldsymbol{a}_i \quad \boldsymbol{c}_i].$$

According to Lemma 1 in the main body of the paper, we can obtain the best linear unbiased estimator of β_1 (denoted by $\hat{\beta}_1^c$) as a linear combination of $\hat{\beta}_1^s$ and $\hat{\beta}_1^p$, with weights $\Sigma_{22} - \Sigma_{12}$ and $\Sigma_{11} - \Sigma_{12}$ respectively. Note that

$$n(\Sigma_{22} - \Sigma_{12}) = \sigma_2^2 + n^{-1}\sum_{i=1}^{n} \boldsymbol{c}_i^\top \Sigma_i \boldsymbol{c}_i - n^{-1}\sum_{i=1}^{n} \boldsymbol{c}_i^\top \Sigma_i \boldsymbol{a}_i$$

and

$$n(\Sigma_{11} - \Sigma_{12}) = n^{-1}\sum_{i=1}^{n} \boldsymbol{a}_i^\top \Sigma_i \boldsymbol{a}_i - n^{-1}\sum_{i=1}^{n} \boldsymbol{c}_i^\top \Sigma_i \boldsymbol{a}_i$$

Following the assumption (S6.6), we have that $\frac{S_{1+}^2 + S_{1-}^2}{2}$ is an estimator of $\sigma_1^2 + n^{-1}\sum_{i=1}^{n} \boldsymbol{a}_i^\top \Sigma_i \boldsymbol{a}_i$ (denoted by $\tau_{n,1}^2$ for short), $\frac{S_{2+}^2 + S_{2-}^2}{2}$ is an estimator of

$$\sigma_2^2 + n^{-1}\sum_{i=1}^{n}(\boldsymbol{c}_i - \boldsymbol{a}_i)^\top \Sigma_i (\boldsymbol{c}_i - \boldsymbol{a}_i) = \sigma_2^2 + n^{-1}\sum_{i=1}^{n} \boldsymbol{c}_i^\top \Sigma_i \boldsymbol{c}_i + n^{-1}\sum_{i=1}^{n} \boldsymbol{a}_i^\top \Sigma_i \boldsymbol{a}_i - 2n^{-1}\sum_{i=1}^{n} \boldsymbol{a}_i^\top \Sigma_i \boldsymbol{c}_i,$$

(denoted by $\tau_{n,2}^2$ for short), and $\frac{S_{++}^2 + S_{+-}^2 + S_{-+}^2 + S_{--}^2}{4}$ is an estimator of $\sigma_1^2 + \sigma_2^2 + n^{-1}\sum_{i=1}^{n} \boldsymbol{c}_i^\top \Sigma_i \boldsymbol{c}_i$ (denoted by $\tau_{n,3}^2$). Here S_{++}^2, S_{+-}^2, S_{-+}^2 and S_{--}^2 are the sample variances of z_i's under experimental setting $x_{i,1}$ and $x_{i,2}$ with the sub-index representing the signs of $x_{i,1}$ and $x_{i,2}$. Note that

$$\frac{1}{2}(\tau_{n,1}^2 + \tau_{n,2}^2 + \tau_{n,3}^2) = \sigma_1^2 + \sigma_2^2 + n^{-1}\sum_{i=1}^{n} \boldsymbol{c}_i^\top \Sigma_i \boldsymbol{c}_i + n^{-1}\sum_{i=1}^{n} \boldsymbol{a}_i^\top \Sigma_i \boldsymbol{a}_i - n^{-1}\sum_{i=1}^{n} \boldsymbol{c}_i^\top \Sigma_i \boldsymbol{a}_i.$$

Then one can have

$$\frac{1}{2}\begin{bmatrix} 1 & 1 & -1 \\ 1 & -1 & 1 \\ -1 & 1 & 1 \end{bmatrix}\begin{bmatrix} \frac{S_{1+}^2+S_{1-}^2}{2} \\ \frac{S_{2+}^2+S_{2-}^2}{2} \\ \frac{S_{++}^2+S_{+-}^2+S_{-+}^2+S_{--}^2}{4} \end{bmatrix} = \begin{bmatrix} n^{-1}\sum_{i=1}^n \boldsymbol{a}_i^\top \Sigma_i \boldsymbol{a}_i - n^{-1}\sum_{i=1}^n \boldsymbol{c}_i^\top \Sigma_i \boldsymbol{a}_i \\ \sigma_1^2 - n^{-1}\sum_{i=1}^n \boldsymbol{c}_i^\top \Sigma_i \boldsymbol{a}_i \\ \sigma_2^2 + n^{-1}\sum_{i=1}^n \boldsymbol{c}_i^\top \Sigma_i \boldsymbol{c}_i - n^{-1}\sum_{i=1}^n \boldsymbol{c}_i^\top \Sigma_i \boldsymbol{a}_i \end{bmatrix}. \tag{S6.9}$$

Therefore, in the implementation, we use the weights of $\sigma_2^2 + n^{-1}\sum_{i=1}^n \boldsymbol{c}_i^\top \Sigma_i \boldsymbol{c}_i - n^{-1}\sum_{i=1}^n \boldsymbol{c}_i^\top \Sigma_i \boldsymbol{a}_i$ and $n^{-1}\sum_{i=1}^n \boldsymbol{a}_i^\top \Sigma_i \boldsymbol{a}_i - n^{-1}\sum_{i=1}^n \boldsymbol{c}_i^\top \Sigma_i \boldsymbol{a}_i$ for $\hat{\beta}^s$ and $\hat{\beta}^p$, which are approximation of the best linear unbiased weights.

We should compare (2.11) and the robust version (S6.9). It is easy to see that the weights for $\hat{\beta}_1^s$ and $\hat{\beta}_1^p$ of the estimator $\hat{\beta}_1^c$ under the model (2.1) and (S6.6) are exactly the same despite of the different model assumptions. Essentially, the more general model (S6.6) overparameterized the random effect assumption. For the estimator $\hat{\beta}_1^c$, such overparameterization is not needed. Following the same derivation, $\hat{\beta}_1^c$ under model (S6.6) is BLUE and asymptotically unbiased. This explains why under different data generating models (a)–(d) in Section 2, the proposed COE can lead to smaller MSE compared to the other two estimators.

S7. Case Study of Customer Campaign

This case study is pseudo-study based on real data on customer campaign. We use this dataset to extract users' effects (a linear combination of user covariates and additional random effects) and add a synthetic additive treatment effect to generate new responses. This pseudo-study allows us to replicate experiments with user effects extracted from real data.

We create a case study based on the customer personality dataset in Kaggle (Patel, 2021). This dataset contains users' information and their responses to five campaigns. We take 2216 users with complete records of all the features. Through variable selection, we keep income, number of kids at home, and number of teens at home as effective user features in model fitting. The response for each of the five campaigns is a binary outcome, indicating accept (1) or not (0). Let $\text{Accept}_{i,k}$ be the status of the i-th user to the k-th campaign, and income_i, kids_i and teens_i be the covariates of i-th user. We model the binary campaign records by a generalized mixed-effect model:

$$\Phi^{-1}\left\{\mathbb{P}(\text{Accept}_{i,k}=1)\right\} = \alpha_k + \gamma_1 \times \text{income}_i + \gamma_2 \times \text{kids}_i + \gamma_3 \times \text{teens}_i + \tilde{u}_i,$$

where α_k's and γ_j's are the fixed effects and $\tilde{u}_i$'s are the random user effects with mean zero and variance τ^2. By fitting this model, we extract the user effect u_i be the conditional mean of $\gamma_1 \times \text{income}_i + \gamma_2 \times \text{kids}_i + \gamma_3 \times \text{teens}_i + \tilde{u}_i$ given data. We then use the extracted user effects to generate responses under (4.17) and convert to binary and count responses by (4.19) and (4.20), respectively. The results are shown

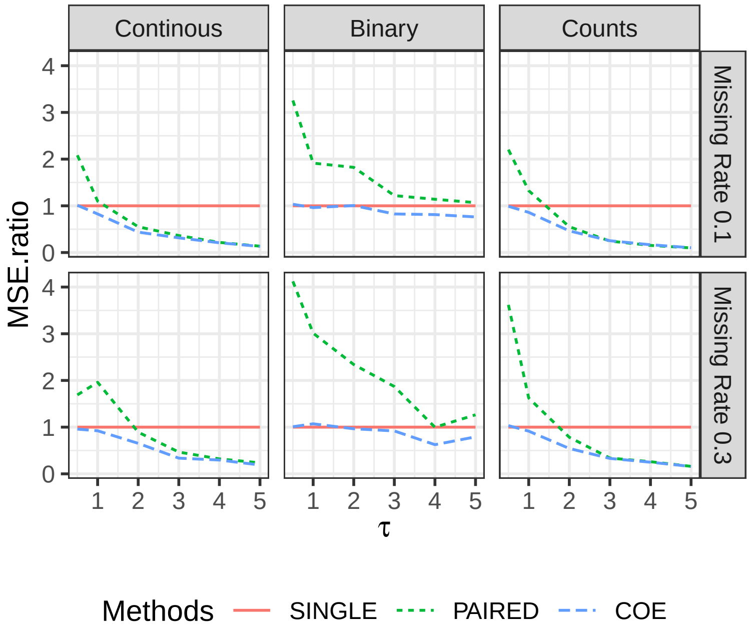

Figure 11: MSE.ratio in (4.18) (for continuous responses) or modified MSE.ratio in (4.21) (for binary and count responses) with user effect extracted from real data.

in Figure 11, which demonstrates a similar comparison as in Section 4.

References

Patel, A. (2021), "Customer Personality Analysis," https://www.kaggle.com/datasets/ imakash3011/customer-personality-analysis.