6 Appendix

6.1 Detailed Proof for Lemma 2

From Equation (16), we can see that $\hat{\boldsymbol{\beta}}_{\mathcal{A}_z}(z)$ is a function of z. For a real value z, there exists t_z^1 such that for any real value z' in $[z, z+t_z^1)$, all elements of $\hat{\boldsymbol{\beta}}_{\mathcal{A}_{z'}}(z')$ remain the same signs with $\hat{\boldsymbol{\beta}}_{\mathcal{A}_z}(z)$. Similarly, from Equation (17), we can see that $\boldsymbol{s}_{\mathcal{A}_z^c}(z)$ is a function of z. Then, for a real value z, there exists t_z^2 such that for any real value z' in $[z, z+t_z^2)$, all elements of $\boldsymbol{s}_{\mathcal{A}_{z'}^c}(z')$ are smaller than 1 in absolute value. Finally, by taking $t_z = \min\{t_z^1, t_z^2\}$, we obtain the interval in which the active set and signs of lasso solution remain the same. The remaining task is to compute t_z^1 and t_z^2.

We first show how to derive t_z^1. From Equation (16), we have

$$\hat{\boldsymbol{\beta}}_{\mathcal{A}_z}(z') - \hat{\boldsymbol{\beta}}_{\mathcal{A}_z}(z) = \boldsymbol{\psi}_{\mathcal{A}_z}(z) \times (z' - z).$$

To guarantee $\hat{\boldsymbol{\beta}}_{\mathcal{A}_z}(z')$ and $\hat{\boldsymbol{\beta}}_{\mathcal{A}_z}(z)$ have the same signs,

$$s_j(z') = s_j(z), \quad \forall j \in \mathcal{A}_z. \tag{27}$$

For a specific $j \in \mathcal{A}_z$, we consider the following cases:

- If $\hat{\beta}_j(z) > 0$, then $\hat{\beta}_j(z') = \hat{\beta}_j(z) + \psi_j(z) \times (z' - z) > 0$.
  - If $\psi_j(z) > 0$, then $z' - z > -\frac{\hat{\beta}_j(z)}{\psi_j(z)}$ (This inequality always holds since the left hand side is positive while the right hand side is negative).
  - If $\psi_j(z) < 0$, then $z' - z < -\frac{\hat{\beta}_j(z)}{\psi_j(z)}$.
- If $\hat{\beta}_j(z) < 0$, then $\hat{\beta}_j(z') = \hat{\beta}_j(z) + \psi_j(z) \times (z' - z) < 0$.
  - If $\psi_j(z) > 0$, then $z' - z < -\frac{\hat{\beta}_j(z)}{\psi_j(z)}$.
  - If $\psi_j(z) < 0$, then $z' - z > -\frac{\hat{\beta}_j(z)}{\psi_j(z)}$ (This inequality always holds since the left hand side is positive while the right hand side is negative).

Finally, for satisfying the condition in Equation (27),

$$z' - z < \min_{j \in \mathcal{A}_z} \left(-\frac{\hat{\beta}_j(z)}{\psi_j(z)} \right)_{++} = t_z^1.$$

We next show how to derive t_z^2. From Equation (17), we have

$$\lambda \boldsymbol{s}_{\mathcal{A}_z^c}(z') - \lambda \boldsymbol{s}_{\mathcal{A}_z^c}(z) = \boldsymbol{\gamma}_{\mathcal{A}_z^c}(z) \times (z' - z).$$

To guarantee $\|\lambda \boldsymbol{s}_{\mathcal{A}_z^c}(z')\|_\infty = \|\lambda \boldsymbol{s}_{\mathcal{A}_z^c}(z) + \boldsymbol{\gamma}_{\mathcal{A}_z^c}(z) \times (z' - z)\|_\infty < \lambda$,

$$-\lambda < \lambda s_j(z) + \gamma_j(z) \times (z' - z) < \lambda, \quad \forall j \in \mathcal{A}_z^c. \tag{28}$$

For a specific $j \in \mathcal{A}_z^c$, we have the following cases:

- If $\gamma_j(z) > 0$, then $\frac{-\lambda - \lambda s_j(z)}{\gamma_j(z)} < z' - z < \frac{\lambda - \lambda s_j(z)}{\gamma_j(z)}$.
- If $\gamma_j(z) < 0$, then $\frac{\lambda - \lambda s_j(z)}{\gamma_j(z)} < z' - z < \frac{-\lambda - \lambda s_j(z)}{\gamma_j(z)}$.

Note that the first inequalities of the above two cases always hold since the left hand side is negative while the right hand side is positive). Then, for satisfying the condition in Equation (28),

$$z' - z < \min_{j \in \mathcal{A}_z^c} \left(\lambda \frac{\mathrm{sign}(\gamma_j(z)) - s_j(z)}{\gamma_j(z)} \right)_{++} = t_z^2.$$

Finally, we can compute t_z by taking $t_z = \min\left\{t_z^1, t_z^2\right\}$.

6.2 Derivations of the Proposed Method for Various Settings

6.2.1 Elastic Net

In some cases, the lasso solutions are unstable. One way to stabilize them is to add an ℓ_2 penalty to the objective function, resulting in the elastic net (Zou and Hastie, 2005). Therefore, we extend our proposed method and provide detailed derivation for testing the selected features in elastic net case. We now consider the optimization problem with parametrized response vector $\boldsymbol{y}(z)$ for $z \in \mathbb{R}$ as follows

$$\hat{\boldsymbol{\beta}}(z) = \underset{\boldsymbol{\beta}\in\mathbb{R}^p}{\arg\min} \frac{1}{2n}\|\boldsymbol{y}(z) - X\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1 + \frac{1}{2}\delta\|\boldsymbol{\beta}\|_2^2. \tag{29}$$

For any z in $\mathbb{R}$, the optimality condition is given by

$$\frac{1}{n}X^\top\Big(X\hat{\boldsymbol{\beta}}(z) - \boldsymbol{y}(z)\Big) + \lambda\boldsymbol{s}(z) + \delta\hat{\boldsymbol{\beta}}(z) = 0,\ \ \boldsymbol{s}(z) \in \partial\|\hat{\boldsymbol{\beta}}(z)\|_1. \tag{30}$$

Similar to lasso case, to construct the truncation region $\mathcal{Z}$, we have to 1) compute the entire path of $\hat{\boldsymbol{\beta}}(z)$ in Equation (29), and 2) identify a set of intervals of z on which $\mathcal{A}(\boldsymbol{y}(z)) = \mathcal{A}(\boldsymbol{y}^{\text{obs}})$.

Lemma 3. *Let us consider two real values z' and z ($z' > z$). If $\hat{\boldsymbol{\beta}}_{\mathcal{A}_z}(z)$ and $\hat{\boldsymbol{\beta}}_{\mathcal{A}_{z'}}(z')$ have the same active set and the same signs, then we have*

$$\hat{\boldsymbol{\beta}}_{\mathcal{A}_z}(z') - \hat{\boldsymbol{\beta}}_{\mathcal{A}_z}(z) = \boldsymbol{\psi}_{\mathcal{A}_z}(z) \times (z' - z), \tag{31}$$

$$\lambda\boldsymbol{s}_{\mathcal{A}_z^c}(z') - \lambda\boldsymbol{s}_{\mathcal{A}_z^c}(z) = \boldsymbol{\gamma}_{\mathcal{A}_z^c}(z) \times (z' - z), \tag{32}$$

where $\boldsymbol{\psi}_{\mathcal{A}_z}(z) = (X_{\mathcal{A}_z}^\top X_{\mathcal{A}_z} + n\delta I_{|\mathcal{A}_z|})^{-1}X_{\mathcal{A}_z}^\top\boldsymbol{b}$, and $\boldsymbol{\gamma}_{\mathcal{A}_z^c}(z) = \frac{1}{n}(X_{\mathcal{A}_z^c}^\top\boldsymbol{b} - X_{\mathcal{A}_z^c}^\top X_{\mathcal{A}_z}\boldsymbol{\psi}_{\mathcal{A}_z}(z))$.

Proof. From the optimality conditions of the elastic net (30) , we have

$$(X_{\mathcal{A}_z}^\top X_{\mathcal{A}_z} + n\delta I_{|\mathcal{A}_z|})\ \hat{\boldsymbol{\beta}}_{\mathcal{A}_z}(z) - X_{\mathcal{A}_z}^\top\boldsymbol{y}(z) + n\lambda\boldsymbol{s}_{\mathcal{A}_z}(z) = 0, \tag{33}$$

$$(X_{\mathcal{A}_{z'}}^\top X_{\mathcal{A}_{z'}} + n\delta I_{|\mathcal{A}_{z'}|})\ \hat{\boldsymbol{\beta}}_{\mathcal{A}_{z'}}(z') - X_{\mathcal{A}_{z'}}^\top\boldsymbol{y}(z') + n\lambda\boldsymbol{s}_{\mathcal{A}_{z'}}(z') = 0. \tag{34}$$

By substracting (33) from (34) and $\mathcal{A}_z = \mathcal{A}_{z'}$, we have

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\mathcal{A}_z}(z') - \hat{\boldsymbol{\beta}}_{\mathcal{A}_z}(z) &= (X_{\mathcal{A}_z}^\top X_{\mathcal{A}_z} + n\delta I_{|\mathcal{A}_z|})^{-1}X_{\mathcal{A}_z}^\top(\boldsymbol{y}(z') - \boldsymbol{y}(z)) \\
&= (X_{\mathcal{A}_z}^\top X_{\mathcal{A}_z} + n\delta I_{|\mathcal{A}_z|})^{-1}X_{\mathcal{A}_z}^\top(\boldsymbol{a} + \boldsymbol{b}z' - \boldsymbol{a} - \boldsymbol{b}z) \\
&= (X_{\mathcal{A}_z}^\top X_{\mathcal{A}_z} + n\delta I_{|\mathcal{A}_z|})^{-1}X_{\mathcal{A}_z}^\top\boldsymbol{b} \times (z' - z).
\end{aligned}$$

Thus, we achieve Equation (31). Similarly, we can write the optimality conditions with $X_{\mathcal{A}_z^c}$ for z and z', and easily obtain Equation (32). □

Now, we can see that $\hat{\boldsymbol{\beta}}_{\mathcal{A}_z}(z)$ and $\boldsymbol{s}_{\mathcal{A}_z^c}(z)$ are functions of z. Then, for a real value z, there exists t_z such that for any real value z' in $[z, z + t_z)$, all elements of $\hat{\boldsymbol{\beta}}_{\mathcal{A}_{z'}}(z')$ remain the same signs with $\hat{\boldsymbol{\beta}}_{\mathcal{A}_z}(z)$, and all elements of $\boldsymbol{s}_{\mathcal{A}_{z'}^c}(z')$ are strictly smaller than 1 in absolute value. The value of t_z can be computed by Lemma 2 as in lasso case.

6.2.2 Full Target Case

In the full target case, as discussed in Liu et al. (2018), the data is used to choose the interesting features but it is *not* used for summarizing the relation between the response and the selected features. Therefore, we can always use *all* the features to define the direction of interest

$$\boldsymbol{\eta}_j = X(X^\top X)^{-1}\boldsymbol{e}_j,$$

where $\boldsymbol{e}_j \in \mathbb{R}^p$ is a zero vector with one at its j^{th} coordinate. The conditional inference is defined as

$$\boldsymbol{\eta}_j^\top\boldsymbol{Y} \mid \left\{j \in \mathcal{A}(\boldsymbol{Y}), \boldsymbol{q}(\boldsymbol{Y}) = \boldsymbol{q}(\boldsymbol{y}^{\text{obs}})\right\}. \tag{35}$$

In Liu et al. (2018), the authors proposed a solution to conduct conditional inference for a specific case when $p < n$, and there is no solution for the case when $p > n$. With the proposed parametric programming method, we can solve this problem. We first re-write the conditional inference in (35) as the problem of characterizing the sampling distribution of

$$Z \mid \{Z \in \mathcal{Z}\} \text{ where } \mathcal{Z} = \{z \in \mathbb{R} \mid j \in \mathcal{A}(\boldsymbol{y}(z))\}. \tag{36}$$

The $\boldsymbol{y}(z)$ in (36) is defined as in (11). Then, to identify $\mathcal{Z}$, we only need to obtain the path of Lasso solution $\hat{\boldsymbol{\beta}}(z)$ as we proposed in §3, and simply check the intervals in which j is an element of the active set corresponding to $\hat{\boldsymbol{\beta}}(z)$ along the path. Finally, after having $\mathcal{Z}$, we can easily compute the selective p-value or selective confidence interval.

6.2.3 Stable Partial Target Case

In the stable partial target case, as discussed in Liu et al. (2018), we only allow stable features to influence the formation of the test-statistic. The stable features are those with very strong signals and we would not to miss out. We will choose a set $\mathcal{H}_{\text{obs}}$ of stable features. Then, for any $j \in \mathcal{H}_{\text{obs}}, j \in \mathcal{A}_{\text{obs}}$,

$$\boldsymbol{\eta}_j = X_{\mathcal{H}_{\text{obs}}}(X_{\mathcal{H}_{\text{obs}}}^\top X_{\mathcal{H}_{\text{obs}}})^{-1}\boldsymbol{e}_j.$$

And, for any $j \notin \mathcal{H}_{\text{obs}}, j \in \mathcal{A}_{\text{obs}}$,

$$\boldsymbol{\eta}_j = X_{\mathcal{H}_{\text{obs}}\cup\{j\}}(X_{\mathcal{H}_{\text{obs}}\cup\{j\}}^\top X_{\mathcal{H}_{\text{obs}}\cup\{j\}})^{-1}\boldsymbol{e}_j.$$

We next show how to construct $\mathcal{H}_{\text{obs}}$ according to Liu et al. (2018).

Stable target formation by setting higher value of λ (TN-ℓ_1). In this case, $\mathcal{H}_{\text{obs}}$ is the lasso active set but with a higher value of λ than the one was used to select $\mathcal{A}_{\text{obs}}$. We denote $\mathcal{H}_{\text{obs}} = \mathcal{H}(\boldsymbol{y}^{\text{obs}})$, the conditional inference is then defined as

$$\boldsymbol{\eta}_j^\top \boldsymbol{Y} \mid \left\{j \in \mathcal{A}(\boldsymbol{Y}), \mathcal{H}(\boldsymbol{Y}) = \mathcal{H}(\boldsymbol{y}^{\text{obs}}), \boldsymbol{q}(\boldsymbol{Y}) = \boldsymbol{q}(\boldsymbol{y}^{\text{obs}})\right\}. \tag{37}$$

The main drawback of the method in Liu et al. (2018) is that they have to consider all $2^{|\mathcal{H}_{\text{obs}}|}$ sign vectors, which requires huge computation time when $|\mathcal{H}_{\text{obs}}|$ is large. With our piecewise-linear homotopy computation, we can easily overcome this drawback. We first re-write the conditional inference in (37) as the problem of characterizing the sampling distribution of

$$Z \mid \{Z \in \mathcal{Z}\} \text{ where } \mathcal{Z} = \{z \in \mathbb{R} \mid j \in \mathcal{A}(\boldsymbol{y}(z)), \mathcal{H}(\boldsymbol{y}(z)) = \mathcal{H}(\boldsymbol{y}^{\text{obs}})\}. \tag{38}$$

We now can easily identify $\mathcal{Z} = \mathcal{Z}_1 \cap \mathcal{Z}_2$, where $\mathcal{Z}_1 = \{z \in \mathbb{R} \mid j \in \mathcal{A}(\boldsymbol{y}(z))\}$ which is the same with full target case, and $\mathcal{Z}_2 = \{z \in \mathbb{R} \mid \mathcal{H}(\boldsymbol{y}(z)) = \mathcal{H}(\boldsymbol{y}^{\text{obs}})\}$ which we can simply obtain by using the proposed method in §3 of the main paper.

Stable target formation by setting a cutoff value c (TN-Custom). In this case, we choose $\mathcal{H}_{\text{obs}}$ by setting a cutoff value c for choosing β_j such that $|\beta_j| \geq c$ [1]. The set $\mathcal{H}_{\text{obs}}$ is defined as

$$\mathcal{H}_{\text{obs}} = \{j \in \mathcal{A}_{\text{obs}}, |\beta_j| \geq c\},$$

where $\beta_j = \boldsymbol{e}_j^\top (X_{\mathcal{A}_{\text{obs}}}^\top X_{\mathcal{A}_{\text{obs}}})^{-1} X_{\mathcal{A}_{\text{obs}}}^\top \boldsymbol{y}^{\text{obs}}$. We denote $\mathcal{H}_{\text{obs}} = \mathcal{H}(\mathcal{A}_{\text{obs}}) \subset \mathcal{A}_{\text{obs}}$, the conditional inference is then formulated as

$$\boldsymbol{\eta}_j^\top \boldsymbol{Y} \mid \{\mathcal{H}(\mathcal{A}(\boldsymbol{Y})) = \mathcal{H}(\mathcal{A}_{\text{obs}}), \mathcal{A}(\boldsymbol{Y}) = \mathcal{A}_{\text{obs}}\}. \tag{39}$$

The main drawback of the method in Liu et al. (2018) is that they still require conditioning on $\{\mathcal{A}(\boldsymbol{Y}) = \mathcal{A}_{\text{obs}}\}$, which is computationally intractable when $|\mathcal{A}_{\text{obs}}|$ is large because the enumeration of $2^{|\mathcal{A}_{\text{obs}}|}$ sign vectors is required. With our proposed method, we can easily overcome this drawback.

[1] We note that our formulation is slightly different but more general than the one in Liu et al. (2018).

6.2.4 Marginal Model

In the case of marginal model, we can always decide a priori to investigate the marginal relationship between the column j of feature matrix X and the observed response vector $\boldsymbol{y}^{\text{obs}}$ if j is selected. The conditional inference is defined as

$$\boldsymbol{\eta}_j^\top \boldsymbol{Y} \mid \left\{ j \in \mathcal{A}(\boldsymbol{Y}), \boldsymbol{q}(\boldsymbol{Y}) = \boldsymbol{q}(\boldsymbol{y}^{\text{obs}}) \right\}, \tag{40}$$

where $\boldsymbol{\eta}_j = X_j(X_j^\top X_j)^{-1}\boldsymbol{e}_j$. The solution for conducting this conditional inference is the same with the full target case. The only difference between marginal model case and full target case is the formulation of $\boldsymbol{\eta}_j$.

6.2.5 Interaction Model

Firstly, we apply Lasso on $\{X, \boldsymbol{y}^{\text{obs}}\}$ to obtain the active set $\mathcal{A}_{\text{obs}} = \mathcal{A}(\boldsymbol{y}^{\text{obs}})$. Next, we construct a feature matrix for interaction model as

$$X_{\text{inter}} = (X_i X_j)_{i,j \in \mathcal{A}_{\text{obs}}, i<j} \in \mathbb{R}^{n \times d},$$

where $d = 0.5|\mathcal{A}_{\text{obs}}|(|\mathcal{A}_{\text{obs}}| - 1)$. Then, the Lasso optimization problem for the interaction model is given by

$$\hat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^d}{\arg\min} \frac{1}{2}\|\boldsymbol{y}^{\text{obs}} - X_{\text{inter}}\boldsymbol{\beta}\|_2^2 + \lambda\|\boldsymbol{\beta}\|_1.$$

Let us denote $\mathcal{A}_{\text{inter}} = \mathcal{A}_{\text{inter}}(\boldsymbol{y}^{\text{obs}})$ be the active set of the interaction model with $\boldsymbol{y}^{\text{obs}}$, the conditional inference on the j^{th} selected feature in $\mathcal{A}_{\text{inter}}$ is defined as

$$\boldsymbol{\eta}_j^\top \boldsymbol{Y} \mid \{ j \in \mathcal{A}_{\text{inter}}(\boldsymbol{Y}), \mathcal{A}(\boldsymbol{Y}) = \mathcal{A}(\boldsymbol{y}^{\text{obs}}), \boldsymbol{q}(\boldsymbol{Y}) = \boldsymbol{q}(\boldsymbol{y}^{\text{obs}}) \}, \tag{41}$$

where $\eta_j = X_{\text{inter}}(X_{\text{inter}}^\top X_{\text{inter}})^{-1}\boldsymbol{e}_j$ in which $\boldsymbol{e}_j \in \mathbb{R}^d$. We note that $\mathcal{A}_{\text{inter}}(\boldsymbol{Y})$ is different from $\mathcal{A}(\boldsymbol{Y})$ which is the active set when we apply Lasso on data $\{X, \boldsymbol{Y}\}$. By restricting the response vector to a line as in (11), the conditional inference in (41) is re-defined as

$$Z \mid \{Z \in \mathcal{Z}\} \text{ where } \mathcal{Z} = \{ z \in \mathbb{R} \mid j \in \mathcal{A}_{\text{inter}}(\boldsymbol{y}(z)), \mathcal{A}(\boldsymbol{y}(z)) = \mathcal{A}(\boldsymbol{y}^{\text{obs}}) \}.$$

From now on, the process of identifying $\mathcal{Z}$ is straightforward which is based on the method we proposed in §3 of the main paper and the extension for full target case in the Appendix.

6.3 Additional Experiments.

For the experiments, we executed the code on Intel(R) Xeon(R) CPU E5-2687W v4 @ 3.00GHz.

Efficiency of the proposed method. We checked the computation time of our extension for elastic net when applying on synthetic data. The results are shown in Figure 9.

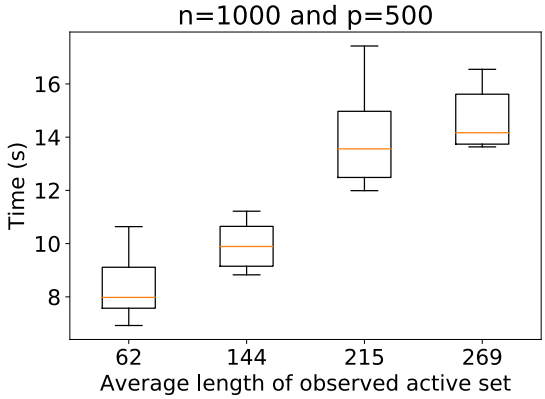


Figure 9: Computation time of our proposed method in elastic net case.

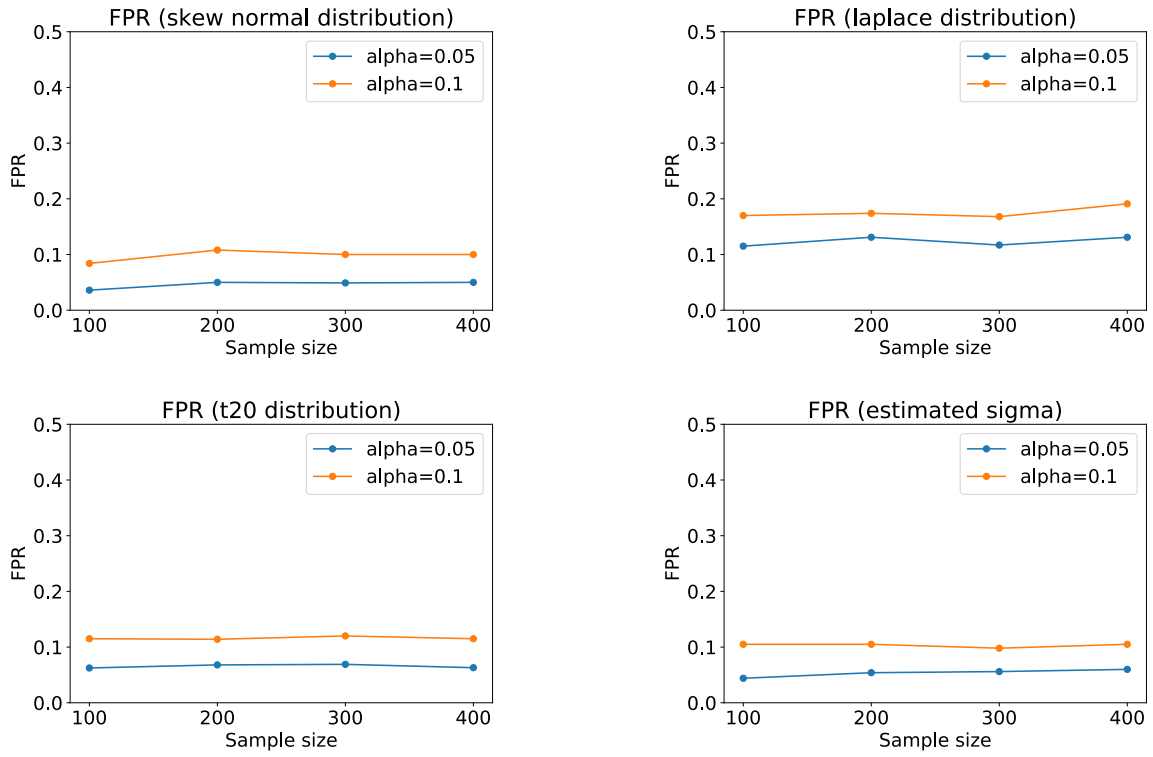
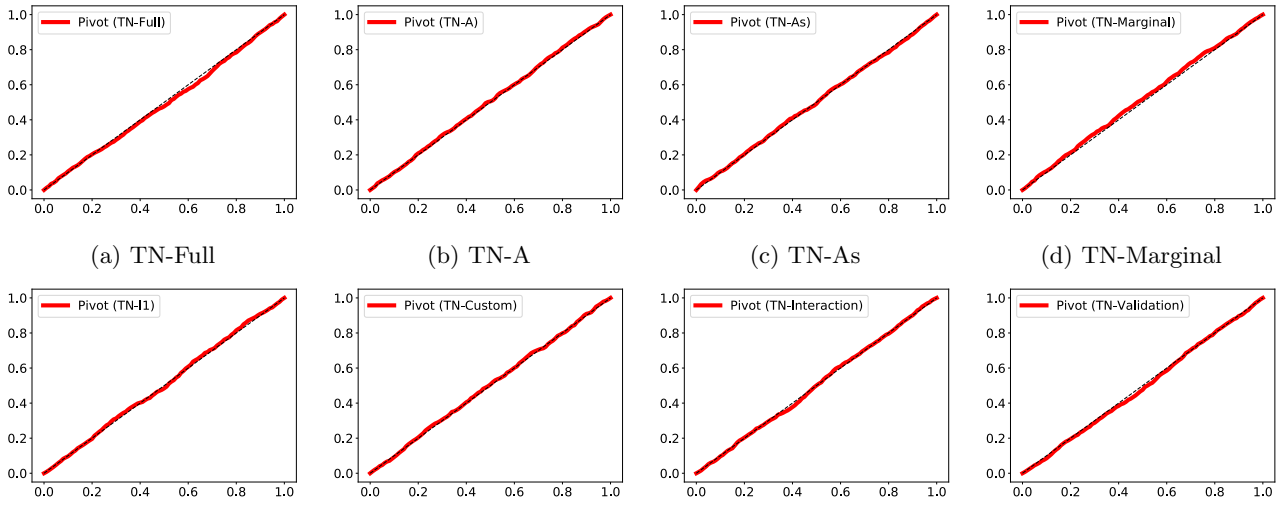


Figure 10: The robustness of the proposed method in terms of the FPR control.



(a) TN-Full (b) TN-A (c) TN-As (d) TN-Marginal

(e) TN-ℓ_1 (f) TN-Custom (g) TN-Interaction (h) TN-Validation

Figure 11: Uniform QQ-plot of the pivotal quantity.

The robustness of the proposed method in terms of the FPR control. We applied our proposed method to the case when the data follows Laplace distribution, skew normal distribution (skewness coefficient 10), and t_{20} distribution. We also conducted experiments when σ^2 is also estimated from the data. We generated n outcomes as $y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i$, $i = 1, ..., n$, where $p = 5, \boldsymbol{x}_i \sim \mathbb{N}(0, I_p)$, and ε_i follows Laplace distribution, skew normal distribution, or t_{20} distribution with zero mean and standard deviation was set to 1. In the case of estimated σ^2, $\varepsilon_i \sim \mathbb{N}(0, 1)$. We set all elements of $\boldsymbol{\beta}$ to 0, and set $\lambda = 0.5$. For each case, we ran 1,200 trials for each $n \in \{100, 200, 300, 400\}$. The FPR results are shown in Figure 10.

Uniformity verification of the pivotal quantity. We generated $n = 100$ outcomes as $y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i$, $i = 1, ..., n$, where $p = 5, \boldsymbol{x}_i \sim \mathbb{N}(0, I_p)$, and $\varepsilon_i \sim \mathbb{N}(0, 1)$. We set the first two elements of $\boldsymbol{\beta}$ to 2, and set $\lambda = 5$. We applied our method and ran 1,200 trials for each case of conditioning: TN-Full, TN-A, TN-As, TN-Marginal (marginal model), TN-ℓ_1, TN-Custom, TN-Interaction (interaction model), and TN-Validation (considering validation selection event). For stable partial target formation, to identify $\mathcal{H}_{\text{obs}}$, we set the value of higher λ to 15 in the case of TN-ℓ_1, and cutoff value c is set to 1 in the case of TN-Custom. We set $\Lambda = \{2^{-1}, 2^0, 2^1\}$ and performed 5-fold cross-validation in the case of TN-Validation. The results are shown in Figure 11.