Biometrics 000, 000–000 DOI: 000
000 0000

Web-based Supplementary Materials for "Estimating Individualized Treatment Rules for Ordinal Treatments" by

Jingxiang Chen[1], Haoda Fu[4], Xuanyao He[4], Michael R. Kosorok[1,2], and Yufeng Liu[1,2,3,*]

[1]Department of Biostatistics, University of North Carolina at Chapel Hill

[2]Department of Statistics and Operations Research, University of North Carolina at Chapel Hill

[3]Department of Genetics, Carolina Center for Genome Sciences,

Lineberger Comprehensive Cancer Center, University of North Carolina at Chapel Hill

[4]Eli Lilly and Company

email: yfliu@email.unc.edu

This paper has been submitted for consideration for publication in *Biometrics*

1. Key Assumptions for Individual Treatment Rule Estimation

We use the notation $R^*(a)$ to denote the outcome under the treatment $a \in \mathcal{A}$, and $p(A|\boldsymbol{X})$ to denote the conditional probability of A given $\boldsymbol{X}$. Recall that the ITR is a map $\mathcal{D} : \mathbb{R}^p \to \mathcal{A}$ which assigns a patient with $\boldsymbol{X}$ to the treatment $\mathcal{D}(\boldsymbol{X})$. The three key assumptions mentioned in the beginning of Section 2 of the main paper are listed as below:

- Assumption 1 (Positivity). There exists a positive ϵ that makes $p(a|\boldsymbol{X}) \geqslant \epsilon$ hold for $\forall a \in \mathcal{A}$ with probability 1.
- Assumption 2 (Strong ignorability). The potential outcome reward $R^*(a) : a \in \mathcal{A}$ are conditionally independent of A given $\boldsymbol{X}$.
- Assumption 3 (Consistency). The potential outcome satisfies $R^*(A) = R$.

These assumptions aim to build the connections between the observed outcome and the potential data.

2. Computational Algorithm for GOWL

Recall the main optimization problem

$$\sum_{i=1}^{n}\sum_{k=1}^{K-1}\frac{|r_i|}{P(a_i|\boldsymbol{x}_i)}\left[I(r_i \geqslant 0)\left[1-a_i^{(k)}f(\boldsymbol{x}_i^{(k)})\right]_+ + I(r_i<0)\left[1+a_i^{(k)}f(\boldsymbol{x}_i^{(k)})\right]_+\right]+\lambda||f||^2. \quad (1)$$

We now introduce our algorithm to solve (1). Due to the convexity of the objective function in (1), we generalize the primal-dual method Vazirani (2013) used in SVM to estimate the classifier $f(\boldsymbol{x}_i^{(k)})$. Starting from (1), by introducing a series of slack variable $\xi_i^{(k)}$ and $\psi_i^{(k)}$ for all observations $i = 1, \cdots, n$ and all duplicates $k = 1, \cdots, K-1$, we rewrite the minimization in (1) by minimizing the following objective function with respect to f and all slack variables,

$$\sum_{i=1}^{n}\sum_{k=1}^{K-1}\frac{\left|r_i^{(k)}\right|}{P(a_i|\boldsymbol{x}_i)}\left[I(r_i^{(k)} \geqslant 0)\xi_i^{(k)} + I(r_i^{(k)}<0)\psi_i^{(k)}\right]+\lambda||f||^2, \quad (2)$$

with $\xi_i^{(k)} \geqslant 0, \psi_i^{(k)} \geqslant 0, \xi_i^{(k)} \geqslant 1 - a_i^{(k)}f(\boldsymbol{x}_i^{(k)})$, and $\psi_i^{(k)} \geqslant 1 + a_i^{(k)}f(\boldsymbol{x}_i^{(k)})$.

Next, we discuss how to solve (2) for the linear case in Section 2.1 and the non-linear case in Section 2.2 of the main paper.

2.1 *Linear Decision Function Estimation*

Suppose that the decision function $f(\boldsymbol{x}_i^{(k)})$ above is a linear function of $\boldsymbol{x}_i^{(k)}$ with the slope $\tilde{\beta}$ and an intercept $\tilde{b}$, i.e. $f(\boldsymbol{x}_i^{(k)}) = \left[\boldsymbol{x}_i^{(k)}\right]^T \tilde{\beta} + \tilde{b}$. Before introducing the algorithm, we express $f(\boldsymbol{x}_i^{(k)}) = \left[\boldsymbol{x}_i^{(k)}\right]^T \tilde{\beta} + \tilde{b} = \boldsymbol{x}_i\beta + b_k$ by denoting $\boldsymbol{x}_i^{(k)} = (\boldsymbol{x}_i^T, \mathbf{e}_k^T)^T$, where $\mathbf{e}_k^T$ is a $K-1$ dimensional row vector whose kth element is 1 while others are zeros. Note that $\tilde{\beta}^T = (\beta^T, b_1 - \tilde{b}, \cdots, b_{K-1} - \tilde{b})$. In other words, the decision function on the duplicated covariate set $\boldsymbol{x}_i^{(k)}$ can also be understood as a varying intercept function of $\boldsymbol{x}_i$, i.e. $f(\boldsymbol{x}_i^{(k)}) = g(\boldsymbol{x}_i) + b_k$. On one hand, such a form of the decision function constructs $K-1$ parallel boundaries in the original sample space to avoid contradicting classifying results. On the other hand, for the ordinal treatment scenario, it is usually desirable to have the $K-1$ intercepts monotonic along the treatment group in terms of the interpretation, i.e. $b_i < (>) b_{i+1}$ for all $i = 1, \cdots, K-2$ when $K \geqslant 3$. We show in Section 4 of the paper that GOWL enjoys such a property under a reasonable condition. When the assumption of parallel linear boundaries becomes too strong, one can use nonlinear learning techniques to achieve more flexible boundaries as in Section 3.3.2 in the main paper.

To solve (2) with a linear decision function, we plug the expression of $f(\boldsymbol{x}_i^{(k)})$ above back into (2) and reparamatrize the formula as:

$$\min_{\tilde{\beta},\xi,\psi} \left\{ \frac{1}{2}||\tilde{\beta}||^2 + C \sum_{i=1}^{K-1} \sum_{k=1}^{K-1} \frac{\left|r_i^{(k)}\right|}{P(a_i|\boldsymbol{x}_i)} \left[I(r_i^{(k)} \geqslant 0)\xi_i^{(k)} + I(r_i^{(k)} < 0)\psi_i^{(k)} \right] \right\},$$

with $\xi_i^{(k)} \geqslant 0, \psi_i^{(k)} \geqslant 0, \xi_i^{(k)} \geqslant 1 - a_i^{(k)} f(\boldsymbol{x}_i^{(k)}),\ \psi_i^{(k)} \geqslant 1 + a_i^{(k)} f(\boldsymbol{x}_i^{(k)})$, and (ξ,ψ) denote all slack variables. By introducing Lagrange multipliers, we can derive the Lagrange function for the primal

problem as:

$$\begin{aligned} L_P &= \frac{1}{2}||\tilde{\beta}||^2 + C\sum_{i=1}^{n}\sum_{k=1}^{K-1}\frac{\left|r_i^{(k)}\right|}{P(a_i|\boldsymbol{x}_i)}\left[I(r_i^{(k)} \geqslant 0)\xi_i^{(k)} + I(r_i^{(k)} < 0)\psi_i^{(k)}\right] \\ &- \sum_{i=1}^{n}\sum_{k=1}^{K-1}\mu_i^{(k)}\xi_i^{(k)} - \sum_{i=1}^{n}\sum_{k=1}^{K-1}\nu_i^{(k)}\psi_i^{(k)} - \sum_{i=1}^{n}\sum_{k=1}^{K-1}\alpha_i^{(k)}\left[a_i^{(k)}f(\boldsymbol{x}_i^{(k)}) + \xi_i^{(k)} - 1\right] \\ &- \sum_{i=1}^{n}\sum_{k=1}^{K-1}\eta_i^{(k)}\left[-a_i^{(k)}f(\boldsymbol{x}_i^{(k)}) + \psi_i^{(k)} - 1\right]. \end{aligned}$$

The corresponding dual problem can be derived by taking partial derivatives with respect to $(\tilde{\beta}, \xi, \psi)$ and simplifying the results using the Karush–Kuhn–Tucker conditions. Then, the dual problem becomes maximizing L_D with respect to the slack variables $\{\alpha_i^{(k)}, \eta_i^{(k)}; i = 1, \ldots, n; k = 1, \ldots, K-1\}$, where

$$\begin{aligned} L_D &= \sum_{i=1}^{n}\sum_{k=1}^{K-1}\alpha_i^{(k)} + \sum_{i=1}^{n}\sum_{k=1}^{K-1}\eta_i^{(k)} - \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{K-1}\sum_{j=1}^{n}\sum_{h=1}^{K-1}\alpha_i^{(k)}\alpha_j^{(h)}a_i^{(k)}a_j^{(k)}\left(\left[\boldsymbol{x}_i^{(k)}\right]^T \cdot \left[\boldsymbol{x}_j^{(h)}\right]\right) \\ &- \frac{1}{2}\sum_{i=1}^{n}\sum_{k=1}^{K-1}\sum_{j=1}^{n}\sum_{h=1}^{K-1}\eta_i^{(k)}\eta_j^{(h)}a_i^{(k)}a_j^{(k)}\left(\left[\boldsymbol{x}_i^{(k)}\right]^T \cdot \left[\boldsymbol{x}_j^{(h)}\right]\right), \end{aligned}$$

with $0 \leqslant \alpha_i^{(k)} \leqslant \frac{C \cdot r_i^{(k)}}{P(a_i|\boldsymbol{x}_i)} I(r_i^{(k)} \geqslant 0), 0 \leqslant \eta_i^{(k)} \leqslant \frac{C \cdot r_i^{(k)}}{P(a_i|\boldsymbol{x}_i)} I(r_i^{(k)} < 0)$, and $\sum_{i=1}^{n}(\alpha_i^{(k)} - \eta_i^{(k)})a_i^{(k)} = 0$. Note that the parameters in the dual problem above can be solved by applying standard quadratic programming with linear constrains. Furthermore, the slope estimate can be obtained via $\hat{\tilde{\beta}} = \sum_{i=1}^{n}\sum_{k=1}^{K-1}(\hat{\alpha}_i^{(k)}a_i^{(k)}\text{sign}(r_i^{(k)} \geqslant 0)\boldsymbol{x}_i^{(k)})$. The intercept vector $\{b_1, \cdots, b_{K-1}\}$ can be estimated by plugging $\hat{\tilde{\beta}}$ back into the original maximization in (1) and solving a standard linear programming problem with linear constraints (Vazirani (2013)). Because there are $2n(K-1)$ parameters in the dual problem above, with a finite K, the computational complexity of (1) is the same as that of the standard primal-dual problem in the SVM.

2.2 *Nonlinear Decision Function Estimation*

The previous subsection solves (2) for the linear case. However, in practice, the linear assumption can be too strong for some problems. To make our model more flexible, we

perform nonlinear learning by applying the kernel learning approach in Reproducing Kernel Hilbert Spaces (RKHS). Kernel learning in RKHS is flexible and has achieved great successes in many nonlinear learning studies (Kimeldorf & Wahba, 1970; Hastie et al., 2011).

Under the binary treatment case, we can show by the Representer Theorem (Kimeldorf & Wahba (1970)) that under some regularity conditions, the decision function on the data $(\boldsymbol{x}_i^{(1)}, a_i^{(1)}, r_i^{(1)})$ can be written in the form $f(\boldsymbol{x}_i^{(1)}) = \sum_{j=1}^n k(\boldsymbol{x}_i, \boldsymbol{x}_j)c_j + \tilde{b}$, where $k(\cdot,\cdot)$ is the standard kernel function associated with the RKHS $\mathcal{H}$. When the treatment is extended to an ordinal variable, we need to define an extended version of the kernel function on the duplicated covariates $\boldsymbol{x}_i^{(k)}$ to construct the decision function. In particular, we have $f(\boldsymbol{x}_i^{(k)}) = \sum_{j=1}^n \sum_{h=1}^{K-1} \tilde{k}(\boldsymbol{x}_i^{(k)}, \boldsymbol{x}_j^{(h)})\tilde{c}_j^{(h)} + \tilde{b}$, where $\tilde{k}(\cdot,\cdot)$ is the extended kernel function with the definition $\tilde{k}(\boldsymbol{x}_i^{(k)}, \boldsymbol{x}_j^{(h)}) = k(\boldsymbol{x}_i, \boldsymbol{x}_j) + e_k^T \cdot e_h$, and e_k is defined as in Section 3.1. Similar discussions were made in Ling & Lin (2006) and Cardoso & Pinto da Costa (2007). According to the newly defined extended kernel, $f(\boldsymbol{x}_i^{(k)})$ can be rewritten as $\sum_{j=1}^n k(\boldsymbol{x}_i, \boldsymbol{x}_j)c_j + b_k$, where $c_j = \sum_{h=1}^{K-1} \tilde{c}_j^{(h)}$ and $b_k = \sum_{j=1}^n \tilde{c}_j^{(k)} + \tilde{b}$. One can tell from the new $f(\boldsymbol{x}_i^{(k)})$ expression that due to the conversion of the ordinal problem into a large binary problem, the corresponding decision boundaries in the kernel-induced feature space are guaranteed not to cross with each other. Consequently, the sets $\{f(\boldsymbol{X}^{(k)}) < 0\}$ for $1 \leqslant k \leqslant K-1$ produce more flexible noncrossing boundaries for the K ordinal treatments in the original space.

Given the expression of f with respect to the kernel representation, we can follow similar Lagrange optimizer steps as before to obtain the generalized primal-dual formula. We can derive the dual problem of maximizing L_D with respect to all slack variables, where

$$\begin{aligned} L_D &= \sum_{i=1}^n \sum_{k=1}^{K-1} \alpha_i^{(k)} + \sum_{i=1}^n \sum_{k=1}^{K-1} \eta_i^{(k)} - \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^{K-1} \sum_{j=1}^n \sum_{h=1}^{K-1} \alpha_i^{(k)} \alpha_j^{(h)} a_i^{(k)} a_j^{(k)} \tilde{k}\left(\boldsymbol{x}_i^{(k)}, \boldsymbol{x}_j^{(h)}\right) \\ &- \frac{1}{2} \sum_{i=1}^n \sum_{k=1}^{K-1} \sum_{j=1}^n \sum_{h=1}^{K-1} \eta_i^{(k)} \eta_j^{(h)} a_i^{(k)} a_j^{(k)} \tilde{k}\left(\boldsymbol{x}_i^{(k)}, \boldsymbol{x}_j^{(h)}\right), \end{aligned}$$

with $0 \leqslant \alpha_i^{(k)} \leqslant \frac{C \cdot r_i^{(k)}}{P(a_i|\boldsymbol{x}_i)} I(r_i^{(k)} \geqslant 0), 0 \leqslant \eta_i^{(k)} \leqslant \frac{C \cdot r_i^{(k)}}{P(a_i|\boldsymbol{x}_i)} I(r_i^{(k)} < 0)$, and $\sum_{i=1}^n (\alpha_i^{(k)} - \eta_i^{(k)}) a_i^{(k)} =$

0. After the dual coefficients are estimated, the decision function can be written as $f(\boldsymbol{x}_i^{(k)}) = \sum_{j=1}^{n}\sum_{h=1}^{K-1}\tilde{k}(\boldsymbol{x}_i^{(k)}, \boldsymbol{x}_j^{(h)})(\hat{\alpha}_j^{(h)} a_j^{(h)} \text{sign}(r_j^{(h)} \geqslant 0))$.

To implement quadratic programming in the dual problems above, we use the open source package CVXOPT based on the Python language in practice.

3. Additional Simulation Results

3.1 *A toy example to show the limitation of the reward shifting strategy*

To investigate the performance of the reward shifting strategy as discussed in Section 2 of the main paper, we modify the first non-linear boundary example in the simulation study, and fit OWL with different ways to shift the reward. In particular, we vary the constant that is selected to shift the rewards within the range $\{5, 10, 100, 500\}$. Then we report the misclassification rates and estimated value functions, and also compare them with those produced by GOWL (see Figure 1 of the supplement). From the boxplots, both the misclassification rates and value functions differ a large amount when the reward is shifted with different constants. In the extreme case when one uses a large constant, the estimated optimal treatment tends to be very close to the actually assigned one. From this result, it appears that a smaller shifting constant may work better, and our GOWL method works the best.

[Figure 1 about here.]

3.2 *Nonlinear Boundary Examples*

For the nonlinear boundary examples, we consider the following four scenarios with $\mu(\boldsymbol{X})$ and $t(\boldsymbol{X}, A)$ defined as,

(1) **Example S1** ($K = 2$): $\mu(\boldsymbol{X}) = 1 + X_1^2 + X_2^2 - 2X_3 + 0.5X_4$ and $t(\boldsymbol{X}, A) = 4(0.7 - X_1^2 - X_2^2)(2A - 3)$, and $P(A = 2|\boldsymbol{X}) = \exp(X_1)/(\exp(X_1) + 1)$;

(2) **Example S2** ($K = 3$): $\mu(\boldsymbol{X}) = -5 + 2X_1 + X_2 + 0.5X_3$ and $t(\boldsymbol{X}, A) = 4\sum_{i=1}^{3} I(g(\boldsymbol{X}) \in (-b_{i-1}, -b_i])(2 - |A - i|)$, where $g(\boldsymbol{X}) = 3 + X_1^2 - 2\exp\{X_2\} - (X_3 - 0.6X_4)^2 - X_5^3 - \exp\{X_6^2\}$, $P(A = k|\boldsymbol{X}) = k/6$, $b_0 = \infty$, $b_1 = 0$, $b_2 = -1.3$ and $b_3 = -\infty$;

(3) **Example S3** ($K = 5$): $\mu(\boldsymbol{X}) = 2 + 2X_1 + X_2 + 0.5X_3$ and $t(\boldsymbol{X}, A) = 4\sum_{i=1}^{5} I(g(\boldsymbol{X}) \in (-b_{i-1}, -b_i])(2 - |A - i|)$, where $g(\boldsymbol{X}) = 3 + X_1^2 - 2\exp\{X_2\} - (X_3 - 0.6X_4)^2 - X_5^3 - \exp\{X_6^2\}$, $P(A|\boldsymbol{X}) = 1/5$, $b_0 = \infty$, $b_1 = 0.4$, $b_2 = -0.3$, $b_3 = -1.1$, $b_4 = -2.1$ and $b_5 = -\infty$;

(4) **Example S4** ($K = 7$): $\mu(\boldsymbol{X}) = 2 + 2X_1 + X_2 + 0.5X_3$ and $t(\boldsymbol{X}, A) = 4\sum_{i=1}^{7} I(g(\boldsymbol{X}) \in (-b_{i-1}, -b_i])(2 - |A - i|)$, where $g(\boldsymbol{X}) = 3 + X_1^2 - 2\exp\{X_2\} - (X_3 - 0.6X_4)^2 - X_5^3$, $P(A = k|\boldsymbol{X}) = 0.1 + 0.1I(k \geqslant 5)$, $b_0 = \infty$, $b_1 = 0.7$, $b_2 = 0.2$, $b_3 = -0.4$, $b_4 = -1$, $b_5 = -1.8$, $b_6 = -2.8$ and $b_7 = -\infty$.

Similar to the linear boundary cases, we have a symmetric reward-treatment curve in each scenario. We repeat the simulation 500 times with the tuning parameters ranging in the same domain. The results are provided in Table 1 of the supplement. In all the four examples, PLS-l_1 performs the worst due to its incorrect model specification. Similar to the linear boundary examples, GOWL-Gaussian with both data duplicate strategies show better performance than OWL-Gaussian in terms of both classification accuracy and value functions. The results of the two data duplicate methods do not demonstrate a significant difference. In the end, we find that none of the methods performs well when the true boundaries have complex structures such as shown in the $K = 7$ setting.

[Table 1 about here.]

3.3 *An Example with both Linear and Nonlinear Boundaries*

We would like to point out that the proposed GOWL could also work well when the parallel assumption of the true boundaries does not hold. Under these circumstances, one should consider using nonlinear learning techniques hence the estimated boundaries would

be flexible enough to approach the underlying true boundaries. To illustrate the idea with a 2-dimensional graph, we use a case with $n = 300$, $p = 2$ and $K = 3$ and follow the previous settings to simulate $\boldsymbol{X}$ and A. At this time, we have the Q-function generated by $Q(\boldsymbol{X}, A, \mathcal{D}^*(\boldsymbol{X})) = 2 + X_1 + 0.5X_2 - 2|A - \mathcal{D}^*(\boldsymbol{X})|$ where $\mathcal{D}^*(\cdot)$, the optimal treatment rule is defined as $\mathcal{D}^*(\boldsymbol{x}) = 1$ if $(X_1+1)^2+(X_2+1)^2 < 1$; $\mathcal{D}^*(\boldsymbol{X}) = 2$ if $X_1+X_2 > 2/3$; $\mathcal{D}^*(\boldsymbol{X}) = 3$ otherwise.

Different from what was discussed in the previous examples, the current boundary set consists of a straight line and one-fourth of a circle. Using GOWL-Gaussian with the same tuning range as in Section 5 of the paper, we plot the estimated boundaries (dashed curves) as well as the true boundaries (solid curves) in Figure 2 of the supplement. The results show that the estimated ITR can still capture the underlying pattern of the optimal ITR well since the RKHS with the Gaussian kernel is very flexible. We repeated the simulation 50 times and the average testing misclassification rate was 5.05%, which illustrates GOWL's competitive prediction ability under the cases of complex boundaries.

[Figure 2 about here.]

3.4 *Comparison with Chen et al. (2016)*

We have applied the method by Chen et al. (2016) for the ordinal treatment setting, despite it being originally designed for continuous dose finding. We used two examples, in which the covariates $\boldsymbol{X}$ are generated the same way as done in Section 5 of the paper, and where treatment A follows a discrete uniform distribution. The treatment has five levels in both cases with the first one following equal treatment effect distances between each pair of the adjacent levels. In this case, using a continuous approximation can be reasonable. In the second example, however, such a structure no longer holds. This implies that it can be inappropriate to approximate the ordinal levels by a continuous variable. Specifically, the two simulation settings are described as follows:

(1) $K = 5$: $\mu(\boldsymbol{X}) = 2X_1 + X_2 + 0.5X_3$ and $t(\boldsymbol{X}, A) = 4\sum_{i=1}^{5} I\left(g(\boldsymbol{X}) \in (b_{i-1}, b_i]\right)(2 - |A - i|)$, where $g(\boldsymbol{X}) = -X_1 + 2X_2 + X_3 + 0.6X_4 - 1.5(X_5 + X_6)$, $P(A|\boldsymbol{X}) = 1/5$, $b_0 = -\infty$, $b_1 = -1.9$, $b_2 = -0.5$, $b_3 = 0.5$, $b_4 = 1.7$ and $b_5 = \infty$;

(2) $K = 5$: $\mu(\boldsymbol{X}) = 2X_1 + X_2 + 0.5X_3$ and $t(X, U) = 4\sum_{i=1}^{5} I\left(g(\boldsymbol{X}) \in (b_{i-1}, b_i]\right)(2 - |U - u_i|)$, where U represents the underlying strength of the observed ordinal treatment A. In particular, we set $U = A \cdot I(A < 3) + 2A \cdot I(A > 3) + 4I(A = 3)$, $u_1 = 1$, $u_2 = 2$, $u_3 = 4$, $u_4 = 8$, and $u_5 = 10$.

The estimated value functions are summarized in Table 2 of the supplement. From the table, we can see that the proposed methods have larger average value functions in both settings. The advantage becomes stronger for the second setting for unequal distances among adjacent levels of treatments. Under both situations, the fitted ITRs of GOWL are nearly optimal. Thus, GOWL works generally well for ITR estimation under ordinal treatments.

[Table 2 about here.]

4. Additional Dataset Applications

4.1 *Irritable Bowel Syndrome Dataset*

This dataset consists of a dose ranging trial that aims to develop a treatment for irritable bowel syndrome (IBS) (see Biesheuvel & Hothorn (2002) for more details). The clinical study enrolled four active treatment arms, corresponding to doses 1, 2, 3, 4 and placebo. The primary endpoint is a baseline adjusted abdominal pain score with larger values corresponding to a better treatment effect. There are 369 patients completing the study, with an almost balanced allocation across the groups of different doses. The final data set only contains three variables: patient gender, treatment, and adjusted abdominal pain score. Approximately 72% of the observed pain scores are greater than 0.

Given the small covariate dimension, we merge doses 1 and 2 together as the low dose

group and merge doses 3 and 4 together as the high dose group. The average adjusted abdominal pain scores of the total data set is 0.475, with standard deviation equal to 0.769. To estimate the optimal ITR, we apply methods including PLS-l_1, OWL-Gaussian, and GOWL-Gaussian, and modify the first two methods in the same way as in the simulation study. As to the evaluation criterion, we calculate the empirical value functions with the following cross-validation strategy. In particular, we randomly partition the dataset into 5 equal-sized parts, train the model based on 4 of them, and predict the value function using the remaining part. We repeat the partition 50 times and the corresponding means and standard deviations of the predicted value functions are 0.491(0.029), 0.503(0.004), and 0.537(0.011) for PLS-l_1, OWL-Gaussian, and GOWL-Gaussian.

The result shows that GOWL returns the highest predicted value function with a moderately low standard deviation. By reassigning the treatment, GOWL could improve the predicted value function by approximately 13%. Furthermore, as to the estimated optimal treatment assignment, PLS-l_1 suggests the optimal treatment to be either placebo or low dose. OWL assigns almost all the patients to the low dose group whereas GOWL suggests about 60% patients in high dose and 40% in low dose. Around 70% of patients are female among those recommended to be in the high dose group. This conclusion appears consistent to what Biesheuvel & Hothorn (2002) reported.

4.2 *Data Proprocessing of the T2DM study*

In this study, 634 patients satisfying aforementioned requirements are enrolled while around 5% have complete observations. To handle the missing data issue before analysis, we choose to remove observations with missing entries at any variable after certain screening steps. Thus, it is important to remove variables whose missing patterns are associated with the values of the clinical outcome significantly. To that end, a t-test is conducted to evaluate whether the missingness, i.e. the missing indicator, of each covariate is associated with the clinical

outcome (i.e. change of hba1c). Based on the test results with Bonferroni multiple-testing adjustment, we consider removing the variables that are not missing at random so that more observations can be retained in the dataset. For categorical covariates having significant test results, we relabel the missing value as a new class when encoding the covariate. We also delete the covariates with missing proportions greater than 70%.

4.3 *T2DM Dataset Analysis without the covariates HDL and LDL*

To have a larger sample size, we have tried to remove the variables high-density lipoprotein cholesterol (HDL), and low-density lipoprotein cholesterol (LDL) so that the total number of observations increased to 350. Then we refit all the methods on the new data set. The value functions are presented in Table 3 of the supplement. From the results, all of the methods produce worse value functions than those before removing the two variables, and GOWL still appears to work the best. This can be viewed as an evidence that the two cholesterol related variables can be important even though they have larger missingness proportions.

[Table 3 about here.]

5. Additional Statistical Learning Theory

We define some essential notation before getting into the details. First, in our context, a risk can be understood as a function of the decision f that averages out all the covariates and treatment information, and is commonly used to evaluate the performance of f. Correspondingly, we define the risk associated with the 0-1 loss as $\mathcal{R}(f) = \sum_{k=1}^{K-1} \mathcal{R}^{(k)}(f) = E\{\sum_{k=1}^{K-1} \frac{R^{(k)}}{P(A|\boldsymbol{X})} I(A^{(k)} \neq \text{sign}(f(\boldsymbol{X}^{(k)})))\}$, where E is taken over (R, A, X) to remove all the randomness of the data, $\mathcal{R}^{(k)}(f) = E[\frac{R^{(k)}}{P(A|\boldsymbol{X})} I(A^{(k)} \neq \text{sign}(f(\boldsymbol{X}^{(k)})))]$ for $k = 1, \cdots, K-1$, and $f(\boldsymbol{X}^{(k)})$ is an ITR associated decision function. According to the 0-1 risk above, we define its Bayes risk, the best risk one can achieve under the 0-1 loss, as $\mathcal{R}(f^*) = \inf_f \{\mathcal{R}(f) | f : \mathcal{X} \to \mathbb{R}\}$ and the corresponding optimal ITR as $\mathcal{D}^*(\boldsymbol{X}) = \sum_{k=1}^{K-1} I(f^*(\boldsymbol{X}^{(k)}) >$

$0) + 1$. Correspondingly, we define the ϕ-risk associated with the proposed surrogate loss as $\mathcal{R}_\phi(f) = \sum_{k=1}^{K-1} \mathcal{R}_\phi^{(k)}(f) = E\{\sum_{k=1}^{K-1} \frac{|R^{(k)}|}{P(A|\boldsymbol{X})}[\phi(A^{(k)} f(\boldsymbol{X}^{(k)}), R^{(k)})]\}$ where $\mathcal{R}_\phi^{(k)}(f) = E[\frac{|R^{(k)}|}{P(A|\boldsymbol{X})}\phi(A^{(k)} f(\boldsymbol{X}^{(k)}), R^{(k)})]$ and $\phi(u, r) = I(r \geqslant 0)[1-u]_+ + I(r < 0)[1+u]_+$. We also define the minimal $\phi-$risk as $\mathcal{R}_\phi(f_\phi^*) = \inf_f \{\mathcal{R}_\phi(f) | f : \mathcal{X} \to \mathbb{R}\}$ and the corresponding surrogate optimal ITR as $\mathcal{D}_\phi^*(\boldsymbol{X}) = \sum_{k=1}^{K-1} I(f_\phi^*(\boldsymbol{X}^{(k)}) > 0) + 1$. Furthermore, we assume that the number of treatment levels K is finite in the following discussions. All the details of theorem proofs are included in Section 6 of the supplement.

5.1 *Further Illustration of the assumption of Theorem 4.2*

We use a numerical illustration to explain when the assumption (i.e. $E(R|\boldsymbol{X}, A > k) > E(R|\boldsymbol{X}, A \leqslant k)$ if and only if $\mathcal{D}^*(\boldsymbol{X}) \geqslant k$) in Theorem 4.2 may fail. Suppose that the conditional expected rewards given certain covariates, i.e. $ER(A|X)$, take values $-2, 0, 8, 6, 5$ when $A = 1, 2, 3, 4, 5$ respectively. Assume that A is uniformly assigned among the five treatments. In this case, when $k = 4$, the expected reward on the left side, where the optimal level 3 is included, is $E[R|X, A \leqslant 4] = 3$, and the expected reward on the right side is $E[R|X, A > 4] = 5$, which is larger. In this way, a wrong decision of the direction can be made in the duplicate space with $k = 4$. As a consequence, the estimated optimal treatment for this subject may be incorrectly greater than the underlying truth after cumulation, i.e. level 4 instead of level 3. In contrast, the decisions of direction will always be correct if the second data duplicate strategy is applied, because comparisons are only made between each two neighboring treatments without any cumulation.

5.2 *Excess 0-1 Risk and Excess $\phi-$Risk*

The following theorem shows that for any decision function f, the excess risk of f under the 0-1 loss, $\mathcal{R}(f) - \mathcal{R}(f^*)$, can be bounded by the excess risk of f under the surrogate loss, $\mathcal{R}_\phi(f) - \mathcal{R}_\phi(f_\phi^*)$.

Theorem 4.4 For any measurable function $f : \mathcal{X} \to \mathbb{R}$ and any probability distribution of (X, A, R), we have $\mathcal{R}_\phi(f) - \mathcal{R}_\phi(f_\phi^*) \geqslant \mathcal{R}(f) - \mathcal{R}(f^*) \geqslant 0$.

Because some of our theoretic discussions are based on the $\phi-$risk, it is necessary to first show how the 0-1 loss risk $\mathcal{R}(f)$ could be controlled accordingly. The proof of Theorem 4.4 uses the idea of partitioning and integration by dividing $\mathcal{R}_\phi(f)$ into $K-1$ parts with $\mathcal{R}_\phi(f) = \sum_{k=1}^{K-1} \mathcal{R}_\phi^{(k)}(f)$. For each part $\mathcal{R}_\phi^{(k)}(f)$, we use the idea of Zhao et al. (2012) and make use of the risk bound theories in Bartlett et al. (2006) to derive the relationship between the two excess risks.

5.3 *Consistency and Convergence Rate*

Denote $\hat{f}_n$ as the sample solution for our proposed GOWL as a minimizer of (1) with $f \in \mathcal{H}$. We next discuss the consistency of $\phi-$risk with $\hat{f}_n$ in the following Theorem 4.5.

Theorem 4.5 (Consistency of $\mathcal{R}_\phi(\hat{f}_n)$) Assume the tuning parameter λ_n is selected such that $\lambda_n \to 0$ and $n\lambda_n \to \infty$. Then for any distribution of (X, A, R), we have that $\mathcal{R}_\phi(\hat{f}_n) \to \inf_{f \in \bar{\mathcal{H}}} \mathcal{R}_\phi(f)$ in probability as $n \to \infty$, where $\hat{f}_n$ is the empirical minimizer of (1) and $\bar{\mathcal{H}}$ denotes the closure of a selected space $\mathcal{H}$.

By Theorem 4.5, minimization of the $\phi-$risk depends on the selection of $\mathcal{H}$. Additionally, if f_ϕ^*, the global minimizer of $\phi-$risk, belongs to the closure of $\limsup_{n\to\infty} \mathcal{H}$, where $\mathcal{H}$ can depend on n, then we have $\inf_{f \in \bar{\mathcal{H}}} \mathcal{R}_\phi(f) = \mathcal{R}_\phi(f_\phi^*)$ and thus $\liminf_{n\to\infty} \mathcal{R}_\phi(\hat{f}_n) = \mathcal{R}_\phi(f_\phi^*)$ in probability. This result will lead to $\liminf_{n\to\infty} \mathcal{R}(\hat{f}_n) = \mathcal{R}(f^*)$ in probability by Theorem 4.4. In particular, the above conditions are met when $\mathcal{H}$ is an RKHS with the Gaussian kernel for which the bandwidth decreases to zero as $n \to \infty$ (see Zhao et al. (2012) for a related discussion).

In the next theorem, we discuss the convergence rate of the excess 0-1 risk $\mathcal{R}(\hat{f}_n) - \mathcal{R}(f^*)$ based on the geometric noise assumption for each measure $P^{(k)}$ introduced in Steinwart & Scovel (2007). For our problem, we define the decision boundary for the optimal ITR as

$\{2\eta(\boldsymbol{X}^{(k)})-1=0\}$ in each classification subproblem between $\{1,\cdots,k\}$ and $\{k+1,\cdots,K\}$ for $k=1\cdots,K-1$, where $\eta(\boldsymbol{X}^{(k)})=\frac{E[R^{(k)}|\boldsymbol{X}^{(k)}=\boldsymbol{X}^{(k)},A^{(k)}=1]-E[R^{(k)}|\boldsymbol{X}^{(k)}=\boldsymbol{X}^{(k)},A^{(k)}=-1]}{E[R^{(k)}|\boldsymbol{X}^{(k)}=\boldsymbol{X}^{(k)},A^{(k)}=1]+E[R^{(k)}|\boldsymbol{X}^{(k)}=\boldsymbol{X}^{(k)},A^{(k)}=-1]}+\frac{1}{2}$. Furthermore, we define the purity of the corresponding data set as $\Delta(\boldsymbol{X}^{(k)})=|2\eta(\boldsymbol{X}^{(k)})-1|$. Note that $\Delta(\boldsymbol{X}^{(k)})$ can be viewed as a measure of closeness of $\boldsymbol{X}^{(k)}$ to the corresponding kth decision boundary. Using these notations, we state the geometric noise assumption in our problem for each duplicate k for $k=1,\cdots,K-1$ as follows: Let $\boldsymbol{X}^{(k)}\in\mathbb{R}^p$ be compact, we define the measure $\mathcal{P}^k$ to have geometric noise exponent $q_k>0$ if there exists a constant $C_k>0$ such that $E[|2\eta(\boldsymbol{X}^{(k)})-1|\exp(-\frac{\Delta(\boldsymbol{X}^{(k)})^2}{t})]\leqslant C_k t^{q_k p/2}$, for $t>0$. According to Steinwart & Scovel (2007), the geometric noise exponent describes the concentration and the noise level of the data generating distribution near the decision boundary. As we will discuss further, the geometric noise exponent q_k of the distribution of $(\boldsymbol{X}^{(k)},A^{(k)},R^{(k)})$ depends on how the density of the data set decreases when the point gets close to the boundary. In the extreme case, q_k can be arbitrarily large when $\eta(\boldsymbol{X}^{(k)})$ is continuous and $\Delta(\boldsymbol{X}^{(k)})>\delta>0$ for some constant $\delta>0$ (i.e., the distinctly separable case). In addition to the geometric noise condition, we also consider the RKHS associated with the Gaussian kernel as in Steinwart & Scovel (2007) in Theorem 4.6 below. We use σ_n to denote the kernel bandwidth for the Gaussian kernel.

Theorem 4.6 (Convergence Rate of the Excess Risk) Suppose that the distribution of $(\boldsymbol{X}^{(k)},A^{(k)},R^{(k)})$ satisfies the geometric noise assumption with exponent $q_k\in(0,\infty)$ for $k=1,\cdots,K-1$. Then for any $\delta>0$ and $\nu\in(0,2)$, there exists a C, which depends on ν,δ, the dimension p, and $P(A|\boldsymbol{X})$, such that $\forall\tau\geqslant 1$ and $\sigma_n=\lambda_n^{-\frac{1}{(q+1)p}}$ for the Gaussian kernel, we have $\Pr^*(\mathcal{R}(\hat{f}_n)\leqslant\mathcal{R}(f^*)+\epsilon)\geqslant 1-e^{-\tau}$, where $q=\arg\max_{q_k}\lambda_n^{q_k/(q_k+1)}$, $\Pr^*$ denotes outer probability and $\epsilon=C(\lambda_n^{-\frac{2}{2+\nu}+\frac{(2-\nu)(1+\delta)}{(2-\nu)(1+q)}}n^{-\frac{2}{2+\nu}}+\frac{\tau}{n\lambda_n}+\lambda_n^{\frac{q}{q+1}})$.

Taking a closer look at the ϵ expression in Theorem 4.6, we can find that the first two terms

can be treated as the bound for the stochastic error, whereas the last term is an error bound for the noise associated with the corresponding RKHS. There is a trade off between the two components. For example, the noise bound term will decrease and the stochastic error will inflate if the RKHS is selected to be more complex. Based on the ϵ expression, one can tell that an optimal choice of λ_n is $n^{-\frac{2(1+q)}{(4+\nu)q+2+(2-\nu)(1+\delta)}}$ and the corresponding rate of the excess risk can be expressed as $\mathcal{R}(\hat{f}_n) - \mathcal{R}(f^*) \leqslant O_p(n^{-\frac{2q}{(4+\nu)q+2+(2-\nu)(1+\delta)}})$. By the geometric noise exponent property, such a q can be sufficiently large when different optimal treatment groups are separated well enough just as in the distinctly separable case we discuss previously. In this way, the rate of convergence can be almost $O_p\left(n^{-1/2}\right)$ when we let δ and ν be small.

6. Proofs of Theorems

In this Section, we give the technical proofs for Lemma 4.1 and Theorem 4.2-4.6.

Proof of Lemma 4.1. We show the Fisher consistency property of GOWL for the binary treatment case in Lemma 4.1. Note that when $A \in \{1, 2\}$, the true optimal treatment rule $\mathcal{D}^*(\boldsymbol{x})$ can be rewritten as

$$I\left(E(R|\boldsymbol{X} = \boldsymbol{x}, A = 2) - E(R|\boldsymbol{X} = \boldsymbol{x}, A = 1) > 0\right) + 1.$$

Because X contains the intercept term, the duplicated covariate matrix $\boldsymbol{X}^{(k)}$ is degenerated into X at this time with $A^{(k)} = A^{(1)} = \text{sign}(A-1)$. Starting from the ϕ-risk $\mathcal{R}_\phi(f)$, we apply the total probability theorem to obtain

$$\begin{aligned} E\left[\frac{|R|}{P(A|\boldsymbol{X})}\phi\left(A^{(1)}f(\boldsymbol{X})\right), R|\boldsymbol{X}\right)\right] &= \sum_{a\in\{-1,1\}} E\left\{\frac{|R|}{P(A|\boldsymbol{X})}\phi(A^{(1)}f(\boldsymbol{X})), R)|\boldsymbol{X}, A^{(1)} = a\right\} P(A|\boldsymbol{X}) \\ &= E\{|R|\phi(f(\boldsymbol{X})), R)|\boldsymbol{X}, A^{(1)} = 1\} \\ &\quad + E\{|R|\phi(-f(\boldsymbol{X})), R)|\boldsymbol{X}, A^{(1)} = -1\}, \end{aligned}$$

where $\phi\left(Af(\boldsymbol{X})), R\right) = I(R \geqslant 0)\left[1 - Af(\boldsymbol{X})\right]_+ + I(R < 0)\left[1 + Af(\boldsymbol{X})\right]_+$ and the event $A^{(1)} = 1$ is equivalent to that of $A = 2$. To obtain an explicit minimizer f^*_ϕ for the objective

function above, we need to discuss its range. In particular, when $f < -1$,

$$\begin{aligned} E\left[\frac{|R|}{P(A|\boldsymbol{X})}\phi\left(A^{(1)}f(\boldsymbol{X})\right), R)\right] &= E\left[RI(R \geqslant 0)(1 - f(\boldsymbol{X}))|\boldsymbol{X}, A = 2\right] \\ &\quad -E\left[RI(R < 0)(1 - f(\boldsymbol{X}))|\boldsymbol{X}, A = 1\right] \\ &= \left\{E\left[RI(R < 0)|\boldsymbol{X}, A = 1\right] - E\left[RI(R \geqslant 0)|\boldsymbol{X}, A = 2\right]\right\} f(\boldsymbol{X}) \\ &\quad +E\left[RI(R < 0)|\boldsymbol{X}, A = 1\right] - E\left[RI(R \geqslant 0)|\boldsymbol{X}, A = 2\right]. \end{aligned}$$

In this case, one can tell that $E\left[\frac{|R|}{P(A|\boldsymbol{X})}\phi\left(A^{(1)}f(\boldsymbol{X})\right), R)\right]$ is always non-negative. This is because for $R \geqslant 0$, $E\left[\frac{|R|}{P(A|\boldsymbol{X})}\phi\left(A^{(1)}f(\boldsymbol{X})\right), R)\right] = -E\left[R|\boldsymbol{X}, A = 2\right] f(\boldsymbol{X}) - E\left[R|\boldsymbol{X}, A = 2\right] \geqslant 0$ since $f < -1$. In addition, a similar argument can be made for $R < 0$. When $f > 1$, one can show that $E\left[\frac{|R|}{P(A|\boldsymbol{X})}\phi\left(A^{(1)}f(\boldsymbol{X})\right), R)\right] \geqslant 0$ still holds based on the same derivation.

When $-1 \leqslant f \leqslant 1$, noting that $\phi\left(A^{(1)}f(\boldsymbol{X})\right), R) = I(R \geqslant 0)\left(1 - A^{(1)}f(\boldsymbol{X})\right) + I(R < 0)\left(1 + A^{(1)}f(\boldsymbol{X})\right)$, we have

$$\begin{aligned} E\left[\frac{|R|}{P(A|\boldsymbol{X})}\phi\left(A^{(1)}f(\boldsymbol{X})\right), R)\right] &= E\left[RI(R \geqslant 0)(1 - f(\boldsymbol{X})) - RI(R < 0)(1 + f(\boldsymbol{X}))|\boldsymbol{X}, A = 2\right] \\ &\quad +E\left[RI(R \geqslant 0)(1 + f(\boldsymbol{X})) - RI(R < 0)(1 - f(\boldsymbol{X}))|\boldsymbol{X}, A = 1\right] \\ &= \left\{E\left[R|A = 1\right] - E\left[R|A = 2\right]\right\} f(\boldsymbol{X}) \\ &\quad +E\left[R|A = 2\right] - E\left[R|A = 1\right]. \end{aligned}$$

The right hand side of the equation above shows that, $E\left[\frac{|R|}{P(A|\boldsymbol{X})}\phi\left(A^{(1)}f(\boldsymbol{X})\right), R)\right]$ becomes zero when $f(\boldsymbol{X}) = 1$ and takes negative values as long as $E\left[R|A = 2\right] - E\left[R|A = 1\right] < 0$. Therefore, the minimizer of the $\phi-$risk, f_ϕ^*, should be within the interval $[-1, 1]$. More specifically, f_ϕ^* should satisfy $\text{sign}(f_\phi^*) = \text{sign}(E\left[R|A = 2\right] - E\left[R|A = 1\right])$, which indicates the surrogate ITR $\mathcal{D}_\phi^*(\boldsymbol{x}) = I\left(f_\phi^*(\boldsymbol{x}) > 0\right) + 1 = \mathcal{D}^*(\boldsymbol{x})$.

Proof of Theorem 4.2. We show that the Fisher consistency property of GOWL still holds for the ordinal treatment case in Theorem 4.2. By Lemma 4.1, for each $\mathcal{R}_\phi^{(k)}$ for which $k = 1, \cdots, K-1$, we note that the minimizer f_ϕ^* has a universal formula for all k so

that $\text{sign}(f^*(\boldsymbol{X}^{(k)})) = \text{sign}\Big(E(R|\boldsymbol{X}^{(k)} = \boldsymbol{X}^{(k)}, A^{(k)} = 1) - E(R|\boldsymbol{X}^{(k)} = \boldsymbol{X}^{(k)}, A^{(k)} = -1)\Big)$. Therefore, by definition, the surrogate ITR of x is

$$\begin{aligned}
\mathcal{D}^*_\phi(\boldsymbol{x}) &= \sum_{k=1}^{K-1} I(E(R^{(k)}|\boldsymbol{X}^{(k)} = \boldsymbol{X}^{(k)}, A^{(k)} = 1) > E(R|\boldsymbol{X}^{(k)} = \boldsymbol{X}^{(k)}, A^{(k)} = -1)) + 1 \\
&= \sum_{k=1}^{K-1} I\left(E(R|\boldsymbol{X} = \boldsymbol{x}, A > k) > E(R|\boldsymbol{X} = \boldsymbol{x}, A \leqslant k)\right) + 1 \\
&= \sum_{k=1}^{K-1} I(\mathcal{D}^*(\boldsymbol{x}) > k) + 1 \\
&= \mathcal{D}^*(\boldsymbol{x}).
\end{aligned}$$

The second equality holds due to the definition of a duplicated data set wherein $\boldsymbol{X}^{(k)} = (\boldsymbol{X}, k)$ and $A^{(k)} = \text{sign}(A - k)$. The third equality holds due to the reward distribution assumption in Theorem 4.2. Note that the second and third equality always hold under the modified duplicate method introduced in Section 4 in which $r_i^{(k)} = r_i$ if $a_i \in \{k, k+1\}$.

Proof of Theorem 4.3. Recall the discussion in Section 3.3 that the $\phi-$risk can be rewritten as

$$\mathcal{R}_\phi = E\left\{\sum_{k=1}^{K-1}\left[\frac{|R|}{P(A|\boldsymbol{X})}\left(I(R \geqslant 0)\phi_1\left(A^{(k)}(g(\boldsymbol{X}) + b_k)\right) + I(R < 0)\phi_2\left(A^{(k)}(g(\boldsymbol{X}) + b_k)\right)\right)\right]\right\},$$

where $\phi_1(u) = [1-u]_+$ and $\phi_2(u) = [1+u]_+$. Without loss of generality, we only need to show that under $E(R|A = k) > 0$ for $k = 1, \cdots, K$, the $\phi-$risk will not be decreased by swapping any two neighbors in the intercept vector $\boldsymbol{b}$ under $b_k > b_{k+1}$ for $k = 1, \cdots, K-2$. Suppose that we swap b_m and b_{m+1} for any $m \in \{1, \cdots, K-2\}$, then the new $\phi-$risk based on the swapped $\boldsymbol{b}$ can be written as

$$\begin{aligned}
\mathcal{R}^s_\phi &= E\left\{\sum_{k\neq m, m+1}\left[\frac{|R|}{P(A|\boldsymbol{X})}\left(I(R \geqslant 0)\phi_1\left(A^{(k)}(g(\boldsymbol{X}) + b_k)\right) + I(R < 0)\phi_2\left(A^{(k)}(g(\boldsymbol{X}) + b_k)\right)\right)\right]\right\} \\
&+ E\left[\frac{|R|}{P(A|\boldsymbol{X})}\left(I(R \geqslant 0)\phi_1\left(A^{(m)}(g(\boldsymbol{X}) + b_{m+1})\right) + I(R < 0)\phi_2\left(A^{(m)}(g(\boldsymbol{X}) + b_{m+1})\right)\right)\right] \\
&+ E\left[\frac{|R|}{P(A|\boldsymbol{X})}\left(I(R \geqslant 0)\phi_1\left(A^{(m+1)}(g(\boldsymbol{X}) + b_m)\right) + I(R < 0)\phi_2\left(A^{(m+1)}(g(\boldsymbol{X}) + b_m)\right)\right)\right].
\end{aligned}$$

Now we discuss how the two risks above are different based on the values of $A^{(m)}$ and $A^{(m+1)}$. One can note that $A^{(m)} \geqslant A^{(m+1)}$ always holds for any m by the fact that $A^{(m)} = \text{sign}\,(A > m)$. In this way, there are three possible situations for the values of $\left(A^{(m)}, A^{(m+1)}\right)$ to take: $(1,-1)$, $(-1,-1)$, and $(1,1)$. We discuss each situation as follows.

First, given the event $\mathcal{A}_{10}^{(m)} = \left\{A^{(m)} = 1, A^{(m+1)} = -1\right\}$, we have that the difference between the swapped risk and original $\phi-$risk is

$$\begin{aligned}
E_{\mathcal{A}_{10}^{(m)}}\left(\mathcal{R}_\phi^s - \mathcal{R}_\phi\right) &= E_{\mathcal{A}_{10}^{(m)}}\left[\frac{|R|}{P(m+1|\boldsymbol{X})} I(R \geqslant 0)\left(\phi_1\left(g(\boldsymbol{X}) + b_{m+1}\right) - \phi_1\left(g(\boldsymbol{X}) + b_m\right)\right)\right] \\
&+ E_{\mathcal{A}_{10}^{(m)}}\left[\frac{|R|}{P(m+1|\boldsymbol{X})} I(R \geqslant 0)\left(\phi_1\left(-\left(g(\boldsymbol{X}) + b_m\right)\right) - \phi_1\left(-\left(g(\boldsymbol{X}) + b_{m+1}\right)\right)\right)\right] \\
&+ E_{\mathcal{A}_{10}^{(m)}}\left[\frac{|R|}{P(m+1|\boldsymbol{X})} I(R < 0)\left(\phi_2\left(g(\boldsymbol{X}) + b_{m+1}\right) - \phi_2\left(g(\boldsymbol{X}) + b_m\right)\right)\right] \\
&+ E_{\mathcal{A}_{10}^{(m)}}\left[\frac{|R|}{P(m+1|\boldsymbol{X})} I(R < 0)\left(\phi_2\left(-(g(\boldsymbol{X}) + b_m)\right) - \phi_2\left(-(g(\boldsymbol{X}) + b_{m+1})\right)\right)\right] \\
&= E\left[RI(R \geqslant 0) \cdot \psi_1(\boldsymbol{x}, \mathbf{b}) | \mathcal{A}_{10}^{(m)}\right] + E\left[RI(R < 0) \cdot \psi_2(\boldsymbol{x}, \mathbf{b}) | \mathcal{A}_{10}^{(m)}\right],
\end{aligned}$$

where $E_{\mathcal{A}_{10}^{(m)}}\left(\mathcal{R}_\phi^s - \mathcal{R}_\phi\right)$ denotes the difference of the two risks under the event $\mathcal{A}_{10}^{(m)} = \left\{A^{(m)} = 1, A^{(m+1)} = -1\right\}$, $\psi_1(\boldsymbol{x}, \mathbf{b}) = \phi_1(g(\boldsymbol{X}) + b_{m+1}) - \phi_1(g(\boldsymbol{X}) + b_m) + \phi_1(-(g(\boldsymbol{X}) + b_m)) - \phi_1(-(g(\boldsymbol{X}) + b_{m+1}))$, and $\psi_2(\boldsymbol{x}, \mathbf{b}) = \phi_2(g(\boldsymbol{X}) + b_m) - \phi_2(g(\boldsymbol{X}) + b_{m+1}) + \phi_2(-(g(\boldsymbol{X}) + b_{m+1})) - \phi_2(-(g(\boldsymbol{X}) + b_m))$. The difference of such conditional expected rewards depends on whether $g(\boldsymbol{X}) + b_m$ and $g(\boldsymbol{X}) + b_{m+1}$ are greater than -1 or not. We summarize the result of each scenario in Table 4. One can find that $\psi_1(\boldsymbol{x}, \mathbf{b})$ is always equal to $\psi_2(\boldsymbol{x}, \mathbf{b})$ and they are always non-negative. In this way, $E_{\mathcal{A}_{10}^{(m)}}\left(\mathcal{R}_\phi^s - \mathcal{R}_\phi\right) = \psi_1(\boldsymbol{x}, \mathbf{b})\{E[RI(R \geqslant 0)|\mathcal{A}_{10}^{(m)}] + E[RI(R < 0)|\mathcal{A}_{10}^{(m)}]\} = \psi_1(\boldsymbol{x}, \mathbf{b}) E(R|\mathcal{A}_{10}^{(m)})$. Thus, one can see that when $b_m > b_{m+1}$, $E_{\mathcal{A}_{10}^{(m)}}(\mathcal{R}_\phi^s - \mathcal{R}_\phi) > 0$ will hold for arbitrary $m = 1, \cdots, K-2$ under the assumption that $E(R|\mathcal{A}_{10}^{(m)}) > 0$ where $\mathcal{A}_{10}^{(m)} = \{A^{(m)} = 1, A^{(m+1)} = -1\} = \{A = m+1\}$.

[Table 4 about here.]

Under the second situation when the event $\mathcal{A}_{11}^{(m)} = \left\{A^{(m)} = 1, A^{(m+1)} = 1\right\}$ holds, the conditional difference between the two risks can be expressed as

$$\begin{aligned}
E_{\mathcal{A}_{11}^{(m)}}\left(\mathcal{R}_\phi^s - \mathcal{R}_\phi\right) &= E_{\mathcal{A}_{11}^{(m)}}\left[\frac{|R|}{P(A > m+1|\boldsymbol{X})} I(R \geqslant 0)\left(\phi_1\left(g(\boldsymbol{X}) + b_{m+1}\right) - \phi_1\left(g(\boldsymbol{X}) + b_m\right)\right)\right] \\
&+ E_{\mathcal{A}_{11}^{(m)}}\left[\frac{|R|}{P(A > m+1|\boldsymbol{X})} I(R \geqslant 0)\left(\phi_1\left(g(\boldsymbol{X}) + b_m\right) - \phi_1\left(g(\boldsymbol{X}) + b_{m+1}\right)\right)\right] \\
&+ E_{\mathcal{A}_{11}^{(m)}}\left[\frac{|R|}{P(A > m+1|\boldsymbol{X})} I(R < 0)\left(\phi_2\left(g(\boldsymbol{X}) + b_{m+1}\right) - \phi_2\left(g(\boldsymbol{X}) + b_m\right)\right)\right] \\
&+ E_{\mathcal{A}_{11}^{(m)}}\left[\frac{|R|}{P(A > m+1|\boldsymbol{X})} I(R < 0)\left(\phi_2\left(g(\boldsymbol{X}) + b_m\right) - \phi_2\left(g(\boldsymbol{X}) + b_{m+1}\right)\right)\right] \\
&= 0.
\end{aligned}$$

Lastly, when the event $\mathcal{A}_{00}^{(m)} = \left\{A^{(m)} = -1, A^{(m+1)} = -1\right\}$ holds, $E_{\mathcal{A}_{11}^{(m)}}\left(\mathcal{R}_\phi^s - \mathcal{R}_\phi\right) = 0$ and the deductions are the same as that in the second scenario. Therefore, when $E_{\mathcal{A}_{10}^{(k)}}[R] > 0$, we will have $E\left(\mathcal{R}_\phi^s - \mathcal{R}_\phi\right) \geqslant 0$, which means $b_k > b_{k+1}$ always holds for $k = 1, \cdots, K-1$. The same deduction can be made for $b_k < b_{k+1}$ when the assumption $E_{\mathcal{A}_{10}^{(k)}}[R] < 0$ holds.

Proof of Theorem 4.4. We first decompose the 0-1 risk based on its definition,

$$\mathcal{R}(f) = \sum_{k=1}^{K-1} \mathcal{R}^{(k)}(f) = \sum_{k=1}^{K-1} E\left[\frac{R}{P(A|\boldsymbol{X})} I\left(A^{(k)} \neq \operatorname{sign}\left(f(\boldsymbol{X}^{(k)})\right)\right)\right],$$

where $\mathcal{R}^{(k)}(f) = E\left[\frac{R}{P(A|\boldsymbol{X})} I\left(A^{(k)} \neq \operatorname{sign}\left(f(\boldsymbol{X}^{(k)})\right)\right)\right]$. Similarly, the $\phi-$risk can be decomposed as,

$$\mathcal{R}_\phi(f) = \sum_{k=1}^{K-1} \mathcal{R}_\phi^{(k)}(f) = \sum_{k=1}^{K-1} E\left[\frac{|R|}{P(A|\boldsymbol{X})} \phi\left(A^{(k)} f(\boldsymbol{X}^{(k)}), R\right)\right],$$

where $\mathcal{R}_\phi^{(k)}(f) = E\left[\frac{|R|}{P(A|\boldsymbol{X})} \phi\left(A^{(k)} f(\boldsymbol{X}^{(k)}), R\right)\right]$ and the $\phi(\cdot)$ function has the same definition as before. Next, we discuss the property of each $\mathcal{R}^{(k)}(f)$ piece following a similar idea found in Zhao et al. (2012) and then combine them to draw the final conclusion.

Without loss of generality, we consider the case where the reward is a discrete variable and the derivation for the continuous case is analogous. To simplify notation, we let $\eta_r(\boldsymbol{x}) =$

$\Pr(A^{(k)} = 1|R = r, \boldsymbol{X}^{(k)} = \boldsymbol{x})$ and $q_r(\boldsymbol{x}) = |r|\Pr(R = r|\boldsymbol{X}^{(k)} = \boldsymbol{x})$ for certain k. When the reward is discrete, the kth component of the Bayes risk is

$$\begin{aligned}\mathcal{R}^{(k)}(f) =& E\left[\sum_r |r| \Pr\left(R = r|\boldsymbol{X}^{(k)}\right) E\left(\frac{I\left(A^{(k)} \neq \text{sign}\left(f(\boldsymbol{X}^{(k)})\right)\right)}{P(A|\boldsymbol{X})}|R = r, \boldsymbol{X}^{(k)}\right)\right] \\ =& E\left[\sum_r q_r(\boldsymbol{X}^{(k)}) E\left(\frac{\eta_r(\boldsymbol{X}^{(k)})}{P(A > k|\boldsymbol{X})} I\left(1 \neq \text{sign}(f(\boldsymbol{X}^{(k)}))\right)\right.\right. \\ &\left.\left. + \frac{1 - \eta_r(\boldsymbol{X}^{(k)})}{P(A \leqslant k|\boldsymbol{X})} I(-1 \neq \text{sign}(f(\boldsymbol{X}^{(k)})))\right)\right]. \end{aligned} \tag{3}$$

To further simplify the expression, we define $h(\boldsymbol{x})$ and $\psi(\boldsymbol{x})$ such that given r and x, the following equations are satisfied:

$$\begin{aligned} h(\boldsymbol{x})\psi(x) &= \sum_r q_r(\boldsymbol{x}) \frac{\eta_r(\boldsymbol{x})}{P(A > k|\boldsymbol{X})} \\ h(\boldsymbol{x})\left(1 - \psi(x)\right) &= \sum_r q_r(\boldsymbol{x}) \frac{1 - \eta_r(\boldsymbol{x})}{P(A \leqslant k|\boldsymbol{X})}. \end{aligned}$$

It can be shown that $h(\boldsymbol{x}) = \sum_r q_r(\boldsymbol{x})[\frac{\eta_r(\boldsymbol{x})}{P(A>k|\boldsymbol{X})} + \frac{1-\eta_r(\boldsymbol{x})}{P(A\leqslant k|\boldsymbol{X})}] > 0$ and $\psi(x) = [\sum_r q_r(\boldsymbol{x})[\frac{\eta_r(\boldsymbol{x})}{P(A>k|\boldsymbol{X})} + \frac{1-\eta_r(\boldsymbol{x})}{P(A\leqslant k|\boldsymbol{X})}]]^{-1}[\sum_r q_r(\boldsymbol{x})\frac{\eta_r(\boldsymbol{x})}{P(A>k|\boldsymbol{X})}]$. Therefore, (3) becomes

$$\mathcal{R}^{(k)}(f) = E\left\{h(\boldsymbol{X}^{(k)})\left[\psi(\boldsymbol{X}^{(k)}) I\left(\text{sign}\left(f(\boldsymbol{X}^{(k)})\right) \neq 1\right) + (1 - \psi(\boldsymbol{X}^{(k)})) I\left(\text{sign}\left(f(\boldsymbol{X}^{(k)})\right) \neq 1\right)\right]\right\}. \tag{4}$$

We follow the same steps above and obtain that the kth component of the $\phi-$risk is

$$\mathcal{R}_\phi^{(k)}(f) = E\left\{h(\boldsymbol{X}^{(k)})\left[\psi(\boldsymbol{X}^{(k)})\phi\left(f(\boldsymbol{X}^{(k)})\right) + \left(1 - \psi(\boldsymbol{X}^{(k)})\right)\phi\left(-f(\boldsymbol{X}^{(k)})\right)\right]\right\}.$$

We define the new function $C(\psi, \alpha) = \psi\phi(\alpha) + (1 - \psi)\phi(-\alpha)$ to rewrite the optimal $\phi-$risk as

$$\mathcal{R}_\phi(f^*) = \sum_{k=1}^{K-1} \mathcal{R}_\phi^{(k)}(f^*) = \inf_{\alpha\in\mathbb{R}} \sum_{k=1}^{K-1} E\left[h\left(\boldsymbol{X}^{(k)}\right) C\left(\psi(\boldsymbol{X}^{(k)}), \alpha\right)\right].$$

Then the excess $\phi-$risk is

$$\mathcal{R}_\phi(f) - \mathcal{R}_\phi(f_\phi^*) = \sum_{k=1}^{K-1} E\left[C\left(\psi(\boldsymbol{X}^{(k)}), f(\boldsymbol{X}^{(k)})\right) h(\boldsymbol{X}^{(k)}) - \inf_{\alpha\in\mathbb{R}} C\left(\psi(\boldsymbol{X}^{(k)}), \alpha\right) h(\boldsymbol{X}^{(k)})\right].$$

According to the result of Bartlett et al. (2006) and the convexity of the loss $\phi(\boldsymbol{x})$, we have for an arbitrary element x in the duplicated sample space $\mathcal{X}^{(k)}$,

$$h(\boldsymbol{x})\,(2\psi-1) = \inf_{\alpha:\alpha(2\psi-1)\leqslant 0} C\,(\psi,\alpha)\,h(\boldsymbol{x}) - \inf_{\alpha\in\mathbb{R}} C\,(\psi,\alpha)\,h(\boldsymbol{x}). \tag{5}$$

In this way, according to (4) and (5), we have for each $k = 1, \cdots, K-1$,

$$\begin{aligned}
\mathcal{R}^{(k)}(f) - \mathcal{R}^{(k)}(f^*) &\leqslant E\left\{ I\left[\operatorname{sign}\left(f(\boldsymbol{X}^{(k)})\right) \neq \operatorname{sign}\left(h(\boldsymbol{X}^{(k)})\left(\psi(\boldsymbol{X}^{(k)}) - \frac{1}{2}\right)\right)\right]\right. \\
&\quad \left.\times \left| h(\boldsymbol{X}^{(k)})\left(2\psi\left(\boldsymbol{X}^{(k)}\right) - 1\right)\right|\right\} \\
&= E\left\{ I\left[\operatorname{sign}\left(f(\boldsymbol{X}^{(k)})\right) \neq \operatorname{sign}\left(h(\boldsymbol{X}^{(k)})(\psi(\boldsymbol{X}^{(k)}) - \frac{1}{2})\right)\right]\right. \\
&\quad \left.\times \left| \inf_{\alpha:\alpha(2\psi-1)\leqslant 0} C\left(\psi(\boldsymbol{X}^{(k)}),\alpha\right) h(\boldsymbol{X}^{(k)}) - \inf_{\alpha\in\mathbb{R}} C\left(\psi(\boldsymbol{X}^{(k)}),\alpha\right) h(\boldsymbol{X}^{(k)})\right|\right\}.
\end{aligned}$$

Since $C(\psi(\boldsymbol{X}^{(k)}), f(\boldsymbol{X}^{(k)}))h(\boldsymbol{X}^{(k)}) \geqslant \inf_{\alpha:\alpha(2\psi-1)\leqslant 0} C(\psi(\boldsymbol{X}^{(k)}),\alpha)h(\boldsymbol{X}^{(k)})$ holds when $\operatorname{sign}(f(\boldsymbol{X}^{(k)})) \neq \operatorname{sign}(h(\boldsymbol{X}^{(k)})(\psi(\boldsymbol{X}^{(k)}) - \frac{1}{2}))$, the second term on the right side of the equal sign above is bounded by $C(\psi(\boldsymbol{X}^{(k)}), f(\boldsymbol{X}^{(k)}))h(\boldsymbol{X}^{(k)}) - \inf_{\alpha\in\mathbb{R}} C(\psi(\boldsymbol{X}^{(k)}),\alpha)h(\boldsymbol{X}^{(k)})$. Therefore, when we sum the inequality through $k = 1, \cdots, K-1$, we have

$$\begin{aligned}
\mathcal{R}(f) - \mathcal{R}(f^*) &= \sum_{k=1}^{K-1} \left\{\mathcal{R}^{(k)}(f) - \mathcal{R}^{(k)}(f^*)\right\} \\
&\leqslant \sum_{k=1}^{K-1} E\left[C\left(\psi(\boldsymbol{X}^{(k)}), f(\boldsymbol{X}^{(k)})\right) h(\boldsymbol{X}^{(k)}) - \inf_{\alpha\in\mathbb{R}} C\left(\psi(\boldsymbol{X}^{(k)}),\alpha\right) h(\boldsymbol{X}^{(k)})\right] \\
&= \mathcal{R}_\phi(f) - \mathcal{R}_\phi(f_\phi^*).
\end{aligned}$$

Proof of Theorem 4.5. We consider the same decomposition idea used in the proof of Theorem 4.4 and express the $\phi-$risk as

$$\mathcal{R}_\phi(f) = \sum_{k=1}^{K-1} \mathcal{R}_\phi^{(k)}(f) = \sum_{k=1}^{K-1} E\left[\frac{|R|}{P(A|\boldsymbol{X})}\phi\left(A^{(k)} f(\boldsymbol{X}^{(k)}), R\right)\right].$$

For the kth component of the $\phi-$risk, we define the loss part as $L_\phi^{(k)}(f) = \frac{|R|}{P(A|\boldsymbol{X})}\phi(A^{(k)}f, R)$. Then for any $f \in \mathcal{H}$ and any minimizer $\hat{f}_n$ of $\mathbb{P}_n\left(\sum_{k=1}^{K-1} L_\phi(f) + \lambda_n||f||^2\right)$, where $\mathbb{P}_n$ denotes

the empirical mean, we have

$$\begin{aligned}\mathbb{P}_n\left(\sum_{k=1}^{K-1} L_\phi^{(k)}(\hat{f}_n)\right) &\leqslant \mathbb{P}_n\left(\sum_{k=1}^{K-1} L_\phi^{(k)}(\hat{f}_n) + \lambda_n||\hat{f}_n||^2\right) \\ &\leqslant \mathbb{P}_n\left(\sum_{k=1}^{K-1} L_\phi^{(k)}(f) + \lambda_n||f||^2\right). \qquad (6)\end{aligned}$$

The second inequality holds because $\hat{f}_n$ minimizes $\mathbb{P}_n\left(\sum_{k=1}^{K-1} L_\phi(f) + \lambda_n||f||^2\right)$ given λ_n. By taking the limit superior on both sides of (6), we have that the following inequality holds for any $f \in \mathcal{H}$:

$$\limsup_{n\to\infty}\mathbb{P}_n\left(\sum_{k=1}^{K-1} L_\phi^{(k)}(\hat{f}_n)\right) \leqslant \limsup_{n\to\infty}\mathbb{P}_n\left(\sum_{k=1}^{K-1} L_\phi^{(k)}(f) + \lambda_n||f||^2\right) = \mathbb{P}\left(\sum_{k=1}^{K-1} L_\phi^{(k)}(f)\right).$$

This yields the fact that

$$\limsup_{n\to\infty}\mathbb{P}_n\left(\sum_{k=1}^{K-1} L_\phi^{(k)}(\hat{f}_n)\right) \leqslant \inf_{f\in\mathcal{H}}\mathbb{P}\left(\sum_{k=1}^{K-1} L_\phi^{(k)}(f)\right).$$

Furthermore, since $\lambda_n \to 0$ when $n \to \infty$, Theorem 4.5 will be proved if we can show $\mathbb{P}_n\left(\sum_{k=1}^{K-1} L_\phi^{(k)}(\hat{f}_n)\right) - \mathbb{P}\left(\sum_{k=1}^{K-1} L_\phi^{(k)}(f)\right) \to 0$ in probability.

To show the convergence condition above, we first prove that $||\hat{f}_n||^2$ can be bounded by some constant depending on n. By (6), if we let $f = 0$ then the inequality becomes

$$\mathbb{P}_n\left(\sum_{k=1}^{K-1} L_\phi^{(k)}(\hat{f}_n)\right) + \lambda_n||\hat{f}_n||^2 \leqslant \mathbb{P}_n\left(\sum_{k=1}^{K-1} L_\phi^{(k)}(0)\right) = \mathbb{P}_n\left(\sum_{k=1}^{K-1}\frac{|R|}{P(A|\boldsymbol{X})}\phi\left(0,R\right)\right).$$

Based on the fact that $\mathbb{P}_n\left(\sum_{k=1}^{K-1} L_\phi^{(k)}(\hat{f}_n)\right) \geqslant 0$ and $\phi(0,R)$ is bounded by 2, if we denote $\pi_0 = \min\{P(a_i|\boldsymbol{x}_i)\}$ for $i = 1,\cdots,n$ (i.e. the smallest prior probability among the K treatments), then

$$||\hat{f}_n||^2 \leqslant \frac{(K-1)\,\phi\left(0\right)}{\pi_0\lambda_n}\sum_{i=1}^{n}\frac{|r_i|}{n} \leqslant \frac{2\,(K-1)}{\pi_0\lambda_n}\sum_{i=1}^{n}\frac{|r_i|}{n}.$$

Due to the existence of $E|R|$, $\exists N \in \mathbb{N}^+$ so that $\forall n > N$, there is an upper bound M such that $||\hat{f}_n||^2 < M$. Furthermore, since the class $\{\sqrt{\lambda_n}f : ||\lambda_n f|| \leqslant \sqrt{M}\}$ is included in a Donsker class and $\sum_{k=1}^{K-1} L_\phi^{(k)}(f)$ is Lipschitz continuous with respect to f, then $\{\sqrt{\lambda_n}\sum_{k=1}^{K-1} L_\phi^{(k)}(f) : ||\lambda_n f|| \leqslant \sqrt{M}\}$ is also a P-Donsker class. In this way, if we denote $\mathbb{P}$ as the population mean

operator, then

$$\sqrt{n}\,(\mathbb{P}_n - \mathbb{P}) \sum_{k=1}^{K-1} L_\phi^{(k)}(\hat{f}_n) = \sqrt{\lambda_n^{-1}}\sqrt{n}\,(\mathbb{P}_n - \mathbb{P}) \left(\sqrt{\lambda_n} \sum_{k=1}^{K-1} L_\phi^{(k)}(\hat{f}_n) \right) = O_p(\sqrt{\lambda_n^{-1}}).$$

Finally, moving the $\sqrt{n}$ from the left hand side to the right hand side in the equation above and taking limits in probability on both sides, we have $\lim_{n\to\infty} (\mathbb{P}_n - \mathbb{P}) \sum_{k=1}^{K-1} L_\phi^{(k)}(\hat{f}_n) = \lim_{n\to\infty} O_p\left(\sqrt{(n\lambda_n)^{-1}}\right) = 0$ in probability as $n\lambda_n \to \infty$.

Proof of Theorem 4.6. We apply the same technique used in Vert & Vert (2006), Steinwart & Scovel (2007) and Zhao et al. (2012) to show the risk convergence property presented in Theorem 4.6. According to Theorem 4.4, Theorem 4.6 will be obtained immediately if we can show the same convergence results for $\mathcal{R}_\phi(\hat{f}_n) - \mathcal{R}_\phi(f_\phi^*)$. We decompose the upper bound of $\mathcal{R}_\phi(\hat{f}_n) - \mathcal{R}_\phi(f_\phi^*)$ using the decomposition idea discussed before, then

$$\begin{aligned}
\mathcal{R}_\phi(\hat{f}_n) - \mathcal{R}_\phi(f_\phi^*) &\leqslant \sum_{k=1}^{K-1} \left[\lambda_n ||\hat{f}_n||^2 + \mathcal{R}_\phi^{(k)}(\hat{f}_n) - \mathcal{R}_\phi^{(k)}(f_\phi^*) \right] \\
&= \sum_{k=1}^{K-1} \left[\lambda_n ||\hat{f}_n||^2 + \mathcal{R}_\phi^{(k)}(\hat{f}_n) - \inf_{f\in\mathcal{H}} \left(\lambda_n ||f||^2 + \mathcal{R}_\phi^{(k)}(f) \right) \right] \\
&\quad + \sum_{k=1}^{K-1} \left[\inf_{f\in\mathcal{H}} \left(\lambda_n ||f||^2 + \mathcal{R}_\phi^{(k)}(f) \right) - \mathcal{R}_\phi^{(k)}(f_\phi^*) \right] \\
&= \sum_{k=1}^{K-1} \left[\lambda_n ||\hat{f}_n||^2 + \mathcal{R}_\phi^{(k)}(\hat{f}_n) - \inf_{f\in\mathcal{H}} \left(\lambda_n ||f||^2 + \mathcal{R}_\phi^{(k)}(f) \right) \right] \\
&\quad + \sum_{k=1}^{K-1} \left[\inf_{f\in\mathcal{H}} \left(\lambda_n ||f||^2 + \mathcal{R}_\phi^{(k)}(f) - \mathcal{R}_\phi^{(k)}(f_\phi^*) \right) \right]. \qquad (7)
\end{aligned}$$

Now, we are to bound each of the $K-1$ pieces on the right hand side of (7) under the new loss function.

We first discuss how to bound the second term in (7). Because the distribution $\mathcal{P}_k$ has geometric noise exponent $0 < q_k < \infty$ with constant C_k for each $k = 1, \cdots, K-1$, then we can find $K-1$ pairs of q_k and C_k such that the following inequality holds for all k

$$E\left[\exp\left(-\frac{\Delta(\boldsymbol{X}^{(k)})^2}{t} \right) \left| 2\eta\left(\boldsymbol{X}^{(k)}\right) - 1 \right| \right] \leqslant C_k t^{q_k p/2}, t > 0.$$

By Theorem 2.7 in Steinwart & Scovel (2007), we can show that there exists $K-1$ constants

$c_{p,k}$ such that for arbitrary $\lambda_n > 0$, we have

$$\sum_{k=1}^{K-1}\left[\inf_{f\in\mathcal{H}}\left(\lambda_n||f||^2+\mathcal{R}_\phi^{(k)}(f)-\mathcal{R}_\phi^{(k)}(f_\phi^*)\right)\right] \leqslant \sum_{k=1}^{K-1} c_{p,k}\left(\sigma_n^p\lambda_n + C_k\left(2p\right)^{q_kp/2}\sigma_n^{-q_kp}\right). \quad (8)$$

Noting that the kth item in the summation of (8) can be considered as $O(\lambda_n^{q_k/(q_k+1)})$ for $k = 1,\cdots,K-1$, we can further bound the summation as follows by defining $q = \arg\max_{q_k}\lambda_n^{q_k/(q_k+1)}$ and thus,

$$\sum_{k=1}^{K-1}\left[\inf_{f\in\mathcal{H}}\left(\lambda_n||f||^2+\mathcal{R}_\phi^{(k)}(f)-\mathcal{R}_\phi^{(k)}(f_\phi^*)\right)\right] \leqslant O(\lambda_n^{q/(q+1)}).$$

As to bounding the first term in (7), we choose to apply Theorem 5.6 of Steinwart & Scovel (2007). To meet the assumptions, we first need to define the corresponding $\mathcal{F}$, Z, T, $\mathcal{G}$, $f_{T,\mathcal{F}}$ and $f_{P,\mathcal{F}}$ in Theorem 5.6 of Steinwart & Scovel (2007) in our new framework.

We define Z as our sample space $\mathcal{X}$ in Section 2, T as the empirical measure P_n, $\mathcal{F}$ as $B_{\mathcal{H}}(\sqrt{\frac{M}{\lambda_n}})$, the subspace of $\mathcal{H}$ which is a ball of $\mathcal{H}$ of radius $\sqrt{\frac{M}{\lambda_n}}$ (where M is the upper bound of $||\lambda_n f||$ according to the proof of Theorem 4.5), $f_{P,\mathcal{F}}$ as the minimizer of the regularized $\phi-$risk under $\mathcal{F}$ and $f_{T,\mathcal{F}}$ as the empirical minimizer $\hat{f}_n$, i.e.,

$$f_{P,\mathcal{F}} = \arg\min_{f\in B_{\mathcal{H}}(\sqrt{\frac{M}{\lambda_n}})}\left(\sum_{k=1}^{K-1}\mathcal{R}_\phi^{(k)}(f)+\lambda_n||f||^2\right).$$

We define $\mathcal{G}$ as the function space considering the loss $L_\phi(f)+\lambda_n||f||^2$ where $\mathcal{R}_\phi(f) = EL_\phi(f)$. That is to say,

$$\mathcal{G}_{\phi,\lambda_n} = \left\{\sum_{k=1}^{K-1}L_\phi^{(k)}(f)+\lambda_n||f||^2-\sum_{k=1}^{K-1}L_\phi^{(k)}(f_{P,\mathcal{F}})-\lambda_n||f_{P,\mathcal{F}}||^2 : f\in B_{\mathcal{H}}(\sqrt{\frac{M}{\lambda_n}})\right\}.$$

Then the remaining work is to show the two conditions in Theorem 5.6 of Steinwart & Scovel (2007): First, $\exists c \geqslant 0$, $0<\alpha\leqslant 1$ and $B>0$ such that $||g||_\infty \leqslant B$ and $E_P\left(g^2\right) \leqslant c\left(E_P g\right)^\alpha$ for $\forall g\in\mathcal{G}_{\phi,\lambda_n}$. Second, $\exists a\geqslant 1$ and $0<b<2$ such that $\sup_{P_n\in\mathcal{X}}\log\mathcal{N}\left(B^{-1}\mathcal{G}_{\phi,\lambda_n},\epsilon,L_2\left(P_n\right)\right)\leqslant a\epsilon^{-b}$ for $\forall\epsilon>0$.

For the first condition, because the new $\phi-$loss function in $L_\phi^{(k)}(f)$ is Lipschitz continuous

for $k = 1, \cdots, K-1$ as discussed early, there exists constants C_k such that $\left|L_\phi^{(k)}(f) - L_\phi^{(k)}(f_{P,\mathcal{F}})\right| \leqslant C_k \left|f - f_{P,\mathcal{F}}\right|$, therefore

$$|L_\phi(f) - L_\phi(f_{P,\mathcal{F}})| \leqslant \sum_{k=1}^{K-1} \left|L_\phi^{(k)}(f) - L_\phi^{(k)}(f_{P,\mathcal{F}})\right| \leqslant C||f - f_{P,\mathcal{F}}||,$$

where $C = \sum_{k=1}^{K-1} C_k$. In this way,

$$\begin{aligned} |g| &\leqslant \left|\sum_{k=1}^{K-1} L_\phi^{(k)}(f) - \sum_{k=1}^{K-1} L_\phi^{(k)}(f_{P,\mathcal{F}})\right| + \left|\lambda_n||f||^2 - \lambda_n||f_{P,\mathcal{F}}||^2\right| \\ &\leqslant C||f - f_{P,\mathcal{F}}|| + \lambda_n||f||^2 - \lambda_n||f_{P,\mathcal{F}}||^2 \qquad (9) \\ &\leqslant C||f - f_{P,\mathcal{F}}|| + \lambda_n||f||^2. \end{aligned}$$

Because we have $f \in B_{\mathcal{H}}(\sqrt{\frac{M}{\lambda_n}})$, then both f and $f_{P,\mathcal{F}}$ are bounded by $\sqrt{\frac{M}{\lambda_n}}$ so that

$$||g||_\infty \leqslant 2C\sqrt{\frac{M}{\lambda_n}} + M.$$

In other words, the constant B in the first condition can be taken to be $2C\sqrt{\frac{M}{\lambda_n}} + M$. To reach the second part of the first condition, we take the second moment of (9) on both sides and obtain

$$\begin{aligned} E\left(g^2\right) &\leqslant E\left(C||f - f_{P,\mathcal{F}}|| + \lambda_n||f||^2 - \lambda_n||f_{P,\mathcal{F}}||^2\right)^2 \\ &\leqslant E\left(C||f - f_{P,\mathcal{F}}|| + \lambda_n \left|||f + f_{P,\mathcal{F}}|| \cdot ||f - f_{P,\mathcal{F}}|||\right)^2 \right. \\ &\leqslant \left(C + 2\lambda_n\sqrt{\frac{M}{\lambda_n}}\right)^2 ||f - f_{P,\mathcal{F}}||^2. \qquad (10) \end{aligned}$$

To show $E_P\left(g^2\right) \leqslant c\left(E_P g\right)^\alpha$, we need to prove that the right hand side of (10) can be upper bounded by $c\left(E_P g\right)^\alpha$. Due to the convexity of $L_\phi(f)$, we have

$$\begin{aligned} \frac{1}{2}\left[L_\phi(f) + L_\phi(f_{P,\mathcal{F}}) + \lambda_n||f||^2 + \lambda_n||f_{P,\mathcal{F}}||^2\right] &\geqslant L_\phi\left(\frac{f + f_{P,\mathcal{F}}}{2}\right) + \frac{1}{2}\left[\lambda_n||f||^2 + \lambda_n||f_{P,\mathcal{F}}||^2\right] \\ &= L_\phi\left(\frac{f + f_{P,\mathcal{F}}}{2}\right) + \lambda_n||\frac{f + f_{P,\mathcal{F}}}{2}||^2 + \lambda_n||\frac{f - f_{P,\mathcal{F}}}{2}||^2. \end{aligned}$$

Taking expectation on both sides and by the definition of $f_{P,\mathcal{F}}$ we have

$$\begin{aligned} \frac{1}{2}\left[\mathcal{R}_\phi(f) + \mathcal{R}_\phi(f_{P,\mathcal{F}}) + \lambda_n||f||^2 + \lambda_n||f_{P,\mathcal{F}}||^2\right] &\geqslant \mathcal{R}_\phi\left(\frac{f + f_{P,\mathcal{F}}}{2}\right) + \lambda_n||\frac{f + f_{P,\mathcal{F}}}{2}||^2 + \lambda_n||\frac{f - f_{P,\mathcal{F}}}{2}||^2 \\ &\geqslant \mathcal{R}_\phi\left(f_{P,\mathcal{F}}\right) + \lambda_n||f_{P,\mathcal{F}}||^2 + \lambda_n||\frac{f - f_{P,\mathcal{F}}}{2}||^2. \end{aligned}$$

Adjusting the inequality slightly, we obtain

$$\frac{1}{2}E_P g = \frac{1}{2}\left[\mathcal{R}_\phi(f) - \mathcal{R}_\phi(f_{P,\mathcal{F}}) + \lambda_n||f||^2 - \lambda_n||f_{P,\mathcal{F}}||^2\right] \geqslant \lambda_n||\frac{f - f_{P,\mathcal{F}}}{2}||^2. \tag{11}$$

Then combining (10) and (11), we have

$$E\left(g^2\right) \leqslant 2\left(C + 2\sqrt{\lambda_n M}\right)^2 \lambda_n^{-1} E_P g.$$

Thus, $E_P\left(g^2\right) \leqslant c\left(E_P g\right)^\alpha$ holds when $\alpha = 1$ and $c = 2\left(C + 2\sqrt{\lambda_n M}\right)^2 \lambda_n^{-1}$. The proof for the first condition is now completed. For the second condition, the entropy we are concerned about can be decomposed by the subadditivity property,

$$\begin{aligned}
\log\mathcal{N}\left(B^{-1}\mathcal{G}_{\phi,\lambda_n}, \epsilon, L_2\left(P_n\right)\right) &= \log\mathcal{N}\left(B^{-1}\{\sum_{k=1}^{K-1} L_\phi^{(k)}(f) + \lambda_n||f||^2\} : f \in B_\mathcal{H}\left(\sqrt{\frac{M}{\lambda_n}}\right), \epsilon, L_2\left(P_n\right)\right) \\
&\leqslant \log\mathcal{N}\left(B^{-1}L_\phi\left(f\right) : f \in B_\mathcal{H}\left(\sqrt{\frac{M}{\lambda_n}}\right), \epsilon, L_2\left(P_n\right)\right) \\
&\quad + \log\mathcal{N}\left(B^{-1}\lambda_n||f||^2 : f \in B_\mathcal{H}\left(\sqrt{\frac{M}{\lambda_n}}\right), \epsilon, L_2\left(P_n\right)\right).
\end{aligned} \tag{12}$$

Since we have$|L_\phi(f_1) - L_\phi(f_2)| \leqslant C||f_1 - f_2||$ for any f_1 and f_2, the corresponding $b_1 = B^{-1}L_\phi\left(f_1\right)$ and $b_2 = B^{-1}L_\phi\left(f_2\right)$ in $\left\{B^{-1}L_\phi\left(f\right) : f \in B_\mathcal{H}\left(\sqrt{\frac{M}{\lambda_n}}\right)\right\}$ must also satisfy $||b_1 - b_2|| \leqslant B^{-1}C||f_1 - f_2||$. In this way, the first term in (12) can be bounded as

$$\begin{aligned}
\log\mathcal{N}\left(B^{-1}L_\phi\left(f\right) : f \in B_\mathcal{H}\left(\sqrt{\frac{M}{\lambda_n}}\right), \epsilon, L_2\left(P_n\right)\right) &\leqslant \log\mathcal{N}\left(B_\mathcal{H}\left(\sqrt{\frac{M}{\lambda_n}}\right), \frac{B}{C}\epsilon, L_2\left(P_n\right)\right) \\
&\leqslant \log\mathcal{N}\left(B_\mathcal{H}\left(1\right), \frac{B}{C}\left[\sqrt{\frac{M}{\lambda_n}}\right]^{-1}\epsilon, L_2\left(P_n\right)\right).
\end{aligned}$$

If we apply Theorem 2.1 in Steinwart & Scovel (2007), because $\frac{B}{C}\left[\sqrt{\frac{M}{\lambda_n}}\right]^{-1}$ is a constant, then $\forall 0 < \nu \leqslant 2$ and $\forall \delta > 0$, there exists a constant c_1 such that for $\epsilon > 0$:

$$\log\mathcal{N}\left(B^{-1}L_\phi\left(f\right) : f \in B_\mathcal{H}\left(\sqrt{\frac{M}{\lambda_n}}\right), \epsilon, L_2\left(P_n\right)\right) \leqslant c_1\sigma_n^{(1-\nu/2)(1+\delta)p}\epsilon^{-\nu}.$$

In this way, there exists a constant c_2 such that

$$\begin{aligned}
\log\mathcal{N}\left(B^{-1}\mathcal{G}_{\phi,\lambda_n}, \epsilon, L_2\left(P_n\right)\right) &\leqslant c_1\sigma_n^{(1-\nu/2)(1+\delta)p}\epsilon^{-\nu} + \log\mathcal{N}\left(B^{-1}\lambda_n||f||^2 : f \in B_\mathcal{H}\left(\sqrt{\frac{M}{\lambda_n}}\right), \epsilon, L_2\left(P_n\right)\right) \\
&\leqslant c_1\sigma_n^{(1-\nu/2)(1+\delta)p}\epsilon^{-\nu} + \log\left(\frac{M}{B\epsilon}\right) \\
&\leqslant c_2\sigma_n^{(1-\nu/2)(1+\delta)p}\epsilon^{-\nu}.
\end{aligned}$$

The proof for the second condition is accomplished. Having established the two conditions

above, we can apply Theorem 5.6 in Steinwart & Scovel (2007) directly and reach the conclusion that there exists a $c_\nu > 0$ depending only on ν such that $\forall n \geqslant 1$ and $\forall \tau \geqslant 1$,

$$\Pr^* \left(\mathcal{R}_\phi \left(\hat{f}_n \right) + \lambda_n ||\hat{f}_n||^2 > \mathcal{R}_\phi \left(f_\phi^* \right) + \lambda_n ||f_\phi^*||^2 + c_\nu \epsilon \left(n, a, B, c, \delta, x \right) \right) \leqslant e^{-\tau},$$

where

$$\begin{aligned} \epsilon \left(n, a, B, c, \delta, x \right) &= B^{2\nu/(4-2\alpha+\alpha\nu)} c^{(2-\nu)/(4-2\alpha+\alpha\nu)} \left(\frac{a}{n} \right)^{2/(4-2\alpha+\alpha\nu)} + B^{\nu/2} \delta^{(2-\nu)/4} \left(\frac{a}{n} \right)^{1/2} \\ &\quad + B \left(\frac{a}{n} \right)^{2/(2+\nu)} + \sqrt{\frac{\delta\tau}{n}} + \left(\frac{c\tau}{n} \right)^{1/(2-\alpha)} + \frac{B\tau}{n}, \end{aligned}$$

and $\alpha = 1$, $c = c_2 \sigma_n^{(1-\nu/2)(1+\delta)p}$, $\sigma_n = \lambda_n^{-1/(q+1)p}$. Once we obtain the convergence rate results of the surrogate risk, the same conclusion can be reached for the 0-1 loss immediately by applying our Theorem 4.4.

7. Software and Sample Code

The following Python files are available from the Biometrics website on Wiley Online Library:

- `gowl_ord.py` is the key class that contains the functions ti implement our method as well as all the other methods involved in the numerical studies.
- `lin_example.py` and `nonlin_example.py` are two examples (linear and non-linear) consisting of data generation, method implementation, and output saving.

References

Bartlett, P. L., Jordan, M. I., & McAuliffe, J. D. (2006). Convexity, Classification, and Risk Bounds. *Journal of the American Statistical Association*, *101*(473), 138–156.

Biesheuvel, E., & Hothorn, L. A. (2002). Many-to-one Comparisons in Stratified Designs. *Biometrical Journal*, *44*(1), 101–116.

Cardoso, J. S., & Pinto da Costa, J. F. (2007). Learning to Classify Ordinal Data: The Data Replication Method. *J. Mach. Learn. Res.*, *8*, 1393–1429.

Chen, G., Zeng, D., & Kosorok, M. R. (2016). Personalized dose finding using outcome weighted learning (with discussions and rejoinder). *Journal of the American Statistical Association*, *111*(516), 1509–1547.

Hastie, T., Tibshirani, R., & Friedman, J. (2011). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York, NY: Springer.

Kimeldorf, G. S., & Wahba, G. (1970). A Correspondence Between Bayesian Estimation on Stochastic Processes and Smoothing by Splines. *The Annals of Mathematical Statistics*, *41*(2), 495–502.

Ling, L., & Lin, H.-T. (2006). Ordinal regression by extended binary classification. *Advances in neural information processing systems*, (pp. 865–872).

Qian, M., & Murphy, S. A. (2011). Performance guarantees for individualized treatment rules. *The Annals of Statistics*, *39*(2), 1180–1210.

Steinwart, I., & Scovel, C. (2007). Fast rates for support vector machines using Gaussian kernels. *The Annals of Statistics*, *35*(2), 575–607.

Vazirani, V. V. (2013). *Approximation Algorithms*. Springer Science & Business Media.

Vert, R., & Vert, J.-P. (2006). Consistency and Convergence Rates of One-Class SVMs and Related Algorithms. *J. Mach. Learn. Res.*, *7*, 817–854.

Zhao, Y., Zeng, D., Rush, A. J., & Kosorok, M. R. (2012). Estimating Individualized Treatment Rules Using Outcome Weighted Learning. *Journal of the American Statistical Association*, *107*(449), 1106–1118.

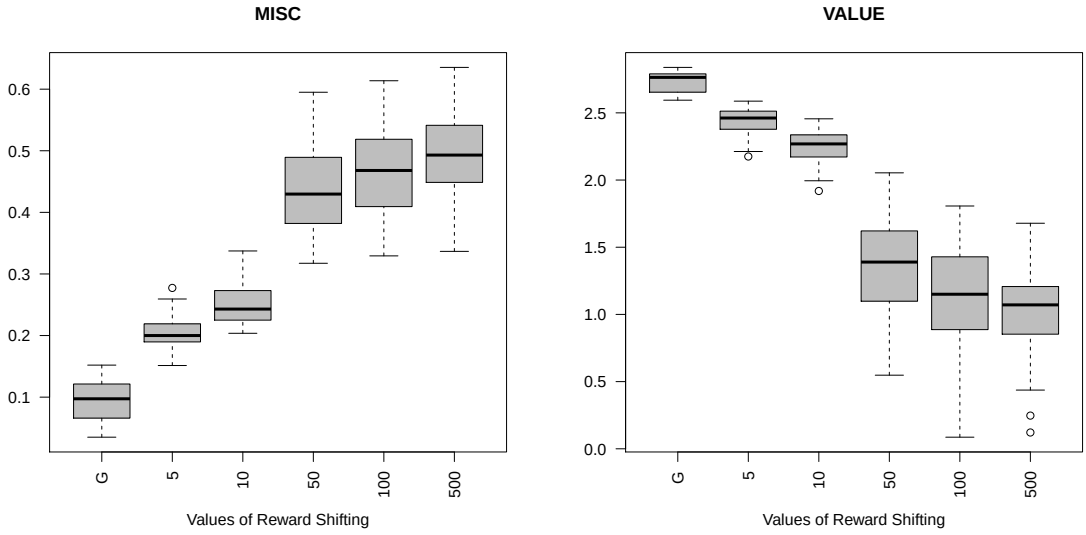

Figure 1. Misclassification rates (MISC) and value functions (VALUE) comparison between GOWL (G) and OWL with different constants $(5, 10, 100, 500)$ used for reward shifting. The simulations are repeated 500 times.

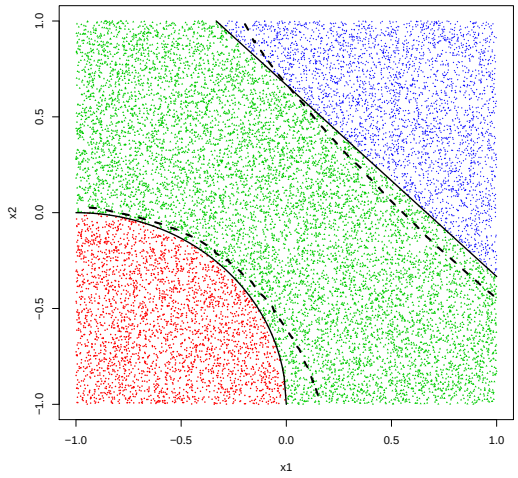

Figure 2. Illustrating plot for the example with the true boundaries containing a linear line and a nonlinear curve. The solid curves indicate the true boundaries and the dashed curves represent the estimated boundaries by GOWL-Gaussian in one simulation. The points correspond to the observations in the test set with the color representing the optimal treatment: red-1, green-2 and blue-3.

Methods	(K, n)	(2,300)	(3,300)	(5,500)	(7,500)
PLS-l_1	MISC	0.399 (0.110)	0.493 (0.189)	0.552 (0.099)	0.746 (0.377)
	VMSE	1.327 (0.427)	1.275 (0.564)	1.601 (0.325)	1.502 (0.894)
OWL-Lin	MISC	0.204 (0.016)	0.381 (0.055)	0.460 (0.089)	0.686 (0.224)
	VMSE	0.489 (0.060)	0.912 (0.429)	1.780 (0.448)	1.659 (0.532)
OWL-Gau	MISC	0.177 (0.027)	0.351 (0.155)	0.373 (0.249)	0.683 (0.222)
	VMSE	0.097 (0.064)	0.712 (0.429)	1.311 (0.471)	1.694 (0.557)
GOWL1-Lin	MISC	0.217 (0.009)	0.362 (0.042)	0.387 (0.022)	0.641 (0.168)
	VMSE	0.222 (0.049)	0.453 (0.130)	0.681 (0.146)	1.085 (0.371)
GOWL1-Gau	MISC	**0.092** (0.034)	0.152 (0.047)	**0.251** (0.081)	0.529 (0.124)
	VMSE	**0.023** (0.012)	**0.112** (0.064)	**0.189** (0.042)	0.622 (0.112)
GOWL2-Lin	MISC	0.217 (0.009)	0.386 (0.038)	0.392 (0.018)	0.660 (0.214)
	VMSE	0.222 (0.049)	0.462 (0.276)	0.634 (0.127)	1.280 (0.279)
GOWL2-Gau	MISC	**0.092** (0.034)	**0.144** (0.036)	0.264 (0.125)	**0.514** (0.135)
	VMSE	**0.023** (0.012)	0.177 (0.068)	0.231 (0.077)	**0.448** (0.154)

Table 1

Results of nonlinear boundary examples: K represents the number of treatment levels; n represents the training set size; the MISC rows show the means and standard deviations (in parenthesis) of the misclassification rates; and the VMSE rows show the means and standard deviations (in parenthesis) of the value function MSEs. PLS$-l_1$ represents penalized least squares including covariate-treatment interactions with l_1 penalty (Qian & Murphy, 2011); OWL represents the outcome weighted learning, and GOWL1 and GOWL2 represent the proposed generalized outcome weighted learning with the first and second data duplication methods. In each scenario, the model producing the best criterion is in bold.

Methods	Dose-Lin	Dose-Gau	GOWL1-Lin	GOWL1-Gau	GOWL2-Lin	GOWL2-Gau	Optimal
1	3.591	3.650	3.869	**3.877**	3.850	3.714	4.019
	(0.109)	(0.055)	(0.105)	(0.085)	(0.116)	(0.158)	(0.112)
2	5.006	5.619	6.228	**6.300**	6.180	6.280	6.876
	(0.279)	(0.145)	(0.269)	(0.313)	(0.317)	(0.327)	(0.213)

Table 2

Results of the means and standard deviations (in parenthesis) of the value functions between GOWL and the dose finding method (Chen et al. (2016)). Dose-Lin and Dose-Gau represent the method by Chen et al. (2016) under linear and Gaussian kernels, GOWL1 and GOWL2 represent the proposed methods with the two data duplicate techniques, and the Optimal column refers to the value function when the optimal ITR is used.

Model	Training	Testing
PLS-l_1	1.804 (0.001)	1.791 (0.014)
OWL-Linear	1.886 (0.010)	1.888 (0.362)
OWL-Gaussian	2.024 (0.005)	1.996 (0.031)
GOWL1-Linear	2.126 (0.042)	2.101 (0.060)
GOWL2-Linear	2.123 (0.044)	2.103 (0.062)
GOWL1-Gaussian	**2.445** (0.005)	**2.339** (0.046)
GOWL2-Gaussian	2.429 (0.004)	2.328 (0.044)

Table 3

Analysis Results for the T2DM Dataset without HDL and LDL. Empirical Value Function Results using 5-fold Cross-Validation with 500 Replications are reported. The abbreviations of methods are the same as those in the tables of simulation studies.

When	$\psi_1(\boldsymbol{x}, \mathbf{b})$, $\psi_2(\boldsymbol{x}, \mathbf{b})$
$g(\boldsymbol{X}) + b_{m+1} < g(\boldsymbol{X}) + b_m \leqslant -1$	$b_m - b_{m+1} > 0$
$g(\boldsymbol{X}) + b_{m+1} \leqslant -1 < g(\boldsymbol{X}) + b_m$	$2 + b_m - b_{m+1} - \phi_1(g(\boldsymbol{X}) + b_m) > 0$
$-1 < g(\boldsymbol{X}) + b_{m+1} < g(\boldsymbol{X}) + b_m$	$b_m - b_{m+1} + \phi_1(g(\boldsymbol{X}) + b_{m+1}) - \phi_1(g(\boldsymbol{X}) + b_m) > 0$

Table 4
All possible $\psi_1(\boldsymbol{x}, \boldsymbol{b})$ and $\psi_2(\boldsymbol{x}, \boldsymbol{b})$ results