Supplementary Appendix

S1 Proof of Theorem 1

We first rewrite the expectation of the proposed estimator in Equation (3) as,

$$\mathbb{E}(\hat{\tau}_k) \;=\; K\mathbb{E}\left\{Y_i\left(f^*(\boldsymbol{X}_i, \hat{c}_k(s))\right) - Y_i\left(f^*(\boldsymbol{X}_i, \hat{c}_{k-1}(s))\right)\right\},$$

where $f^*(\boldsymbol{X}_i, c) = \mathbf{1}\{s(\boldsymbol{X}_i) < c\}$. Similarly, we can also write the estimand in Equation (2) as,

$$\tau_k \;=\; K\mathbb{E}\left\{Y_i\left(f^*(\boldsymbol{X}_i, c_k(s))\right) - Y_i\left(f^*(\boldsymbol{X}_i, c_{k-1}(s))\right)\right\}.$$

Now, define $F(c) = \mathbb{P}(s(\boldsymbol{X}_i) \leq c)$. Without loss of generality, assume $\hat{c}_k(s) > c_k(s)$ and $\hat{c}_{k-1}(s) > c_{k-1}(s)$. If this is not the case, we simply switch the upper and lower limits of the integrals in the proof below. Then, the bias of the estimator is given by,

$$\begin{aligned}
& \frac{|\mathbb{E}(\hat{\tau}_k) - \tau_k|}{K} \\
\leq\; & |\mathbb{E}\left\{Y_i\left(f^*(\boldsymbol{X}_i, \hat{c}_k(s))\right) - Y_i\left(f(\boldsymbol{X}_i, c_k(s))\right)\right\}| + |\mathbb{E}\left\{Y_i\left(f^*(\boldsymbol{X}_i, \hat{c}_{k-1}(s))\right) - Y_i\left(f^*(\boldsymbol{X}_i, c_{k-1}(s))\right)\right\}| \\
=\; & \left|\mathbb{E}_{\hat{c}_k(s)}\left[\int_{c_k(s)}^{\hat{c}_k(s)} \mathbb{E}(\tau_i \mid s(\boldsymbol{X}_i) = c)\mathrm{d}F(c)\right]\right| + \left|\mathbb{E}_{\hat{c}_{k-1}(s)}\left[\int_{c_{k-1}(s)}^{\hat{c}_{k-1}(s)} \mathbb{E}(\tau_i \mid s(\boldsymbol{X}_i) = c)\mathrm{d}F(c)\right]\right| \\
=\; & \left|\mathbb{E}_{F(\hat{c}_k(s))}\left[\int_{F(c_k(s))}^{F(\hat{c}_k(s))} \mathbb{E}(\tau_i \mid s(\boldsymbol{X}_i) = F^{-1}(x))\mathrm{d}x\right]\right| \\
& + \left|\mathbb{E}_{F(\hat{c}_{k-1}(s))}\left[\int_{F(c_{k-1}(s))}^{F(\hat{c}_{k-1}(s))} \mathbb{E}(\tau_i \mid s(\boldsymbol{X}_i) = F^{-1}(x))\mathrm{d}x\right]\right| \\
\leq\; & \mathbb{E}_{F(\hat{c}_k(s))}\left[\left|F(\hat{c}_k(s)) - \frac{k}{K}\right| \times \max_{c\in[c_k(s), \hat{c}_k(s)]} |\mathbb{E}(\tau_i \mid s(\boldsymbol{X}_i) = c)|\right] \\
& + \mathbb{E}_{F(\hat{c}_{k-1}(s))}\left[\left|F(\hat{c}_{k-1}(s)) - \frac{k-1}{K}\right| \times \max_{c\in[c_{k-1}(s), \hat{c}_{k-1}(s)]} |\mathbb{E}(\tau_i \mid s(\boldsymbol{X}_i) = c)|\right]
\end{aligned}$$

By the definition of $\hat{c}_k(s)$, $F(\hat{c}_k(s))$ is the nk/Kth order statistic of n independent uniform random variables, and thus follows the Beta distribution with the shape and scale parameters equal to nk/K and $n - nk/K + 1$, respectively. For the special case where $k - 1 = 0$, we define the 0th order statistic of n uniform random variables to be 0, and by extension also define the "beta distribution" with shape parameter ≤ 0 to be $H(x)$ where $H(x)$ is the Heaviside step function. Therefore, we have,

$$\mathbb{P}\left(|F(\hat{c}_k(s)) - \frac{k}{K}| > \epsilon\right) \;=\; 1 - B\left(\frac{k}{K} + \epsilon, \frac{nk}{K}, n - \frac{nk}{K} + 1\right) + B\left(\frac{k}{K} - \epsilon, \frac{nk}{K}, n - \frac{nk}{K} + 1\right),$$

where $B(\epsilon, \alpha, \beta) = \int_0^\epsilon t^{\alpha-1}(1-t)^{\beta-1}\mathrm{d}t$ is the incomplete beta function. Similarly, we have

$$\begin{aligned}
\mathbb{P}(|F(\hat{c}_{k-1}(s)) - \frac{k-1}{K}| > \epsilon) \;=\; & 1 - B\left(\frac{k-1}{K} + \epsilon, \frac{n(k-1)}{K}, n - \frac{n(k-1)}{K} + 1\right) \\
& + B\left(\frac{(k-1)}{K} - \epsilon, \frac{n(k-1)}{K}, n - \frac{n(k-1)}{K} + 1\right).
\end{aligned}$$

Combining the above results yields the desired bias bound expression.

To derive the exact variance, we first apply the law of total variance to Equation (3),

$$\begin{aligned}
\mathbb{V}(\hat{\tau}_k) &= \mathbb{V}\left[\mathbb{E}\left\{K\left(\frac{1}{n_1}\sum_{i=1}^{n}\hat{f}_k(\boldsymbol{X}_i)T_iY_i(1) - \frac{1}{n_0}\sum_{i=1}^{n}\hat{f}_k(\boldsymbol{X}_i)(1-T_i)Y_i(0)\right) \,\middle|\, \boldsymbol{X}, \{Y_i(1), Y_i(0)\}_{i=1}^{n}\right\}\right] \\
&\quad + \mathbb{E}\left[\mathbb{V}\left\{K\left(\frac{1}{n_1}\sum_{i=1}^{n}\hat{f}_k(\boldsymbol{X}_i)T_iY_i(1) - \frac{1}{n_0}\sum_{i=1}^{n}\hat{f}_k(\boldsymbol{X}_i)(1-T_i)Y_i(0)\right) \,\middle|\, \boldsymbol{X}, \{Y_i(1), Y_i(0)\}_{i=1}^{n}\right\}\right] \\
&= K^2\mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_{ki}(1) - Y_{ki}(0)\}\right) \\
&\quad + K^2\mathbb{E}\left[\mathbb{V}\left\{\frac{1}{n_1}\sum_{i=1}^{n}\hat{f}_k(\boldsymbol{X}_i)T_iY_i(1) - \frac{1}{n_0}\sum_{i=1}^{n}\hat{f}_k(\boldsymbol{X}_i)(1-T_i)Y_i(0) \,\middle|\, \boldsymbol{X}, \{Y_i(1), Y_i(0)\}_{i=1}^{n}\right\}\right].
\end{aligned} \tag{S1}$$

Applying the standard result from Neyman's finite sample variance analysis to the second term shows that this term is equal to,

$$K^2\mathbb{E}\left\{\frac{1}{n}\left(\frac{n_0}{n_1}S_{k1}^2 + \frac{n_1}{n_0}S_{k0}^2 + 2S_{k01}\right)\right\}. \tag{S2}$$

where $S_{k01} = \sum_{i=1}^{n}(Y_{ki}(0) - \overline{Y_k(0)})(Y_{ki}(1) - \overline{Y_k(1)})/(n-1)$. Since $Y_{ki}(t)$ and $Y_{kj}(t)$ are correlated, we apply Lemma 1 of Nadeau and Bengio (2000) to the first term, yielding,

$$\begin{aligned}
&\mathbb{V}\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_{ki}(1) - Y_{ki}(0)\}\right) \\
&= \mathrm{Cov}(Y_{ki}(1) - Y_{ki}(0), Y_{kj}(1) - Y_{kj}(0)) + \frac{1}{n}\mathbb{E}(S_{k1}^2 + S_{k0}^2 - 2S_{k01}),
\end{aligned} \tag{S3}$$

for $i \neq j$ where

$$\begin{aligned}
&\mathrm{Cov}(Y_{ki}(1) - Y_{ki}(0), Y_{kj}(1) - Y_{kj}(0)) \\
&= \mathrm{Cov}\left(\hat{f}_k(\boldsymbol{X}_i)\tau_i, \hat{f}_k(\boldsymbol{X}_j)\tau_j\right) \\
&= \Pr(\hat{f}_k(\boldsymbol{X}_i) = \hat{f}_k(\boldsymbol{X}_j) = 1)\mathbb{E}[\tau_i\tau_j \mid \hat{f}_k(\boldsymbol{X}_i) = \hat{f}_k(\boldsymbol{X}_j) = 1] - \Pr(\hat{f}_k(\boldsymbol{X}_i) = 1)^2\mathbb{E}[\tau_i \mid \hat{f}_k(\boldsymbol{X}_i) = 1]^2 \\
&= \frac{n-K}{K^2(n-1)}\mathbb{E}[\tau_i\tau_j \mid \hat{f}_k(\boldsymbol{X}_i) = \hat{f}_k(\boldsymbol{X}_j) = 1] - \frac{1}{K^2}\mathbb{E}[\tau_i \mid \hat{f}_k(\boldsymbol{X}_i) = 1]^2 \\
&= \frac{(n-K)\kappa_{k11}}{K^2(n-1)} - \frac{\kappa_{k1}^2}{K^2}
\end{aligned}$$

Substituting Equations (S2) and (S3) into Equation S1, we obtain the desired variance expression. □

S2 Derivation of $\hat{\kappa}_{ktt}$

We first rewrite κ_{ktt} as:

$$\kappa_{ktt} = \sum_{u,v\in\{0,1\}}(-1)^{u+v}\mathbb{E}[Y_i(u)Y_j(v) \mid \hat{f}_k(\boldsymbol{X}_i) = \hat{f}_k(\boldsymbol{X}_j) = t].$$

We can estimate each conditional expectation term inside of the summation using its sample analogue:

$$\frac{\sum_{i=1}^{n}\sum_{j\neq i}\mathbf{1}\{\hat{f}_k(\boldsymbol{X}_i)=\hat{f}_k(\boldsymbol{X}_j)=t\}\{1-u+(2u-1)T_i\}\{1-v+(2v-1)T_j\}Y_iY_j}{\sum_{i=1}^{n}\sum_{j\neq i}\mathbf{1}\{\hat{f}_k(\boldsymbol{X}_i)=\hat{f}_k(\boldsymbol{X}_j)=t\}\{1-u+(2u-1)T_i\}\{1-v+(2v-1)T_j\}}$$

We can further simplify the computation by rewriting the numerator as:

$$\left[\sum_{i=1}^{n}\mathbf{1}\{\hat{f}_k(\boldsymbol{X}_i)=t\}\{1-u+(2u-1)T_i\}Y_i\right]\left[\sum_{i=1}^{n}\mathbf{1}\{\hat{f}_k(\boldsymbol{X}_i)=t\}\{1-v+(2v-1)T_i\}Y_i\right]$$
$$-\sum_{i=1}^{n}\mathbf{1}\{\hat{f}_k(\boldsymbol{X}_i)=t\}\{1-u+(2u-1)T_i\}\{1-v+(2v-1)T_i\}Y_i^2.$$

Similarly, we can rewrite the denominator as follows:

$$\left[\sum_{i=1}^{n}\mathbf{1}\{\hat{f}_k(\boldsymbol{X}_i)=t\}(1-u+(2u-1)T_i)\right]\left[\sum_{i=1}^{n}\mathbf{1}\{\hat{f}_k(\boldsymbol{X}_i)=t\}\{1-v+(2v-1)T_i\}\right]$$
$$-\sum_{i=1}^{n}\mathbf{1}\{\hat{f}_k(\boldsymbol{X}_i)=t\}\{1-u+(2u-1)T_i\}\{1-v+(2v-1)T_i\}.$$

Putting these terms together, we obtain the expression of $\hat{\kappa}_{ktt}$ given in Section 2.2.

S3 Proof of Theorem 2

Given a tuple of n samples $\{Y_i, T_i, \boldsymbol{X}_i\}_{i=1}^n$, we first reorder the sample to $(Y_{[i,n]}, T_{[i,n]}, \boldsymbol{X}_{[i,n]})$ based on the magnitude of the scoring rule, such that

$$s(\boldsymbol{X}_{[1,n]}) \leq s(\boldsymbol{X}_{[2,n]}) \leq \cdots \leq s(\boldsymbol{X}_{[n,n]})$$

Then, the proposed GATES estimator can be rewritten as

$$\hat{\tau}_k = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}\left\{\frac{(k-1)n}{K} < i \leq \frac{kn}{K}\right\}U_{[i,n]} \tag{S4}$$

where

$$U_{[i,n]} := KY_{[i,n]}\left(\frac{T_{[i,n]}}{q}-\frac{1-T_{[i,n]}}{1-q}\right), \tag{S5}$$

where $q = n_1/n$. Now, we prove the following two lemmas.

LEMMA S1 *Let $(X_1, Y_1), (X_2, Y_2), \cdots$ be a sequence of random vectors. For each $n \geq 1$, $(X_1, Y_1), \cdots, (X_n, Y_n)$ possesses a joint distribution. Let $\boldsymbol{Z}_n = ((X_1, Y_1), \cdots, (X_n, Y_n))$ and $\boldsymbol{X}_n = (X_1, \cdots, X_n)$, and let $W_n(\boldsymbol{Z}_n)$ and $S_n(\boldsymbol{X}_n)$ be measurable vector-valued functions of $\boldsymbol{Z}_n$ and $\boldsymbol{X}_n$ respectively. Suppose $S_n(\boldsymbol{X}_n)$ converges in distribution to F_S and the conditional distribution $W_n(\boldsymbol{Z}_n) \mid \boldsymbol{X}_n$ converges in distribution to F_W in probability, where F_W does not depend on $\boldsymbol{X}_n$. Then, we have that:*

$$(W_n(\boldsymbol{Z}_n), S_n(\boldsymbol{X}_n)) \to F_W F_S$$

Proof The characteristic function of the joint distribution of $(W_n(\boldsymbol{Z}_n), S_n(\boldsymbol{X}_n))$ can be written as:

$$\varphi_{W_nS_n}(t_1, t_2) = \mathbb{E}[\exp\{i(t_1W_n(\boldsymbol{Z}_n) + t_2S_n(\boldsymbol{X}_n))\}]$$

Let $W \sim F_W$ and $S \sim F_S$. Then the characteristic function of F_WF_S can be written as:

$$\varphi_{WS}(t_1, t_2) = \mathbb{E}[\exp\{i(t_1W)\}]\mathbb{E}[\exp\{i(t_2S)\}]$$

We then have:

$$\begin{aligned}
&|\varphi_{W_nS_n}(t_1, t_2) - \varphi_{WS}(t_1, t_2)| \\
=& |\mathbb{E}[\exp\{i(t_1W_n(\boldsymbol{Z}_n) + t_2S_n(\boldsymbol{X}_n))\}] - \mathbb{E}[\exp\{i(t_1W)\}]\mathbb{E}[\exp\{i(t_2S)\}]| \\
\leq& |\mathbb{E}[\mathbb{E}[\exp\{i(t_1W_n(\boldsymbol{Z}_n) + t_2S_n(\boldsymbol{X}_n))\} \mid \boldsymbol{X}_n]] - \mathbb{E}[\exp\{i(t_1W)\}]\mathbb{E}[\exp\{i(t_2S_n(\boldsymbol{X}_n))\}]| \\
&+ |\mathbb{E}[\exp\{i(t_1W)\}]\mathbb{E}[\exp\{i(t_2S_n(\boldsymbol{X}_n))\}] - \mathbb{E}[\exp\{i(t_1W)\}]\mathbb{E}[\exp\{i(t_2S)\}]| \\
\leq& \mathbb{E}[|\mathbb{E}[\exp(it_1W_n(\boldsymbol{Z}_n)) \mid \boldsymbol{X}_n] - \mathbb{E}[\exp(it_1W)]|] + |\mathbb{E}[\exp\{i(t_2S_n(\boldsymbol{X}_n))\}] - \mathbb{E}[\exp\{i(t_2S)\}]|,
\end{aligned}$$

where the last inequality follows from the fact that all characteristic functions satisfy $|\varphi| \leq 1$. This expression converges to zero in probability due to the convergence of $S_n(\boldsymbol{X}_n)$ and $W_n(\boldsymbol{Z}_n) \mid \boldsymbol{X}_n$ respectively. Therefore, we have:

$$(W_n(\boldsymbol{Z}_n), S_n(\boldsymbol{X}_n)) \to F_WF_S$$

□

Lemma S2 $\lim\limits_{n\to\infty} \mathbb{E}(\hat{\tau}_k) - \tau_k = O\left(n^{-1}\right)$

Proof We bound the bias of $\mathbb{E}(\hat{\tau}_k)$ by appealing to Theorem 1 of Imai and Li (2023b), which implies,

$$\begin{aligned}
|\mathbb{E}(\hat{\tau}_k) - \tau_k| \leq& \left|K\mathbb{E}\left[\int_{F(c_k(s))}^{F(\hat{c}_k(s))} \mathbb{E}(Y_i(1) - Y_i(0) \mid s(\boldsymbol{X}_i) = F^{-1}(x))\mathrm{d}x\right]\right| \\
&+ \left|K\mathbb{E}\left[\int_{F(c_{k-1}(s))}^{F(\hat{c}_{k-1}(s))} \mathbb{E}(Y_i(1) - Y_i(0) \mid s(\boldsymbol{X}_i) = F^{-1}(x))\mathrm{d}x\right]\right|. \qquad \text{(S6)}
\end{aligned}$$

By the definition of $\hat{c}_k(s)$, $F(\hat{c}_k(s))$ is the nk/Kth order statistic of n independent uniform random variables, and therefore, follows the Beta distribution with the shape and scale parameters equal to nk/K and $n - nk/K + 1$, respectively.

Now, by Assumption 4, we can compute the first-order Taylor expansion of $\int_a^x \mathbb{E}(Y_i(1) - Y_i(0) \mid s(\boldsymbol{X}_i) = F^{-1}(x))\mathrm{d}x$:

$$\begin{aligned}
|\mathbb{E}(\hat{\tau}_k) - \tau_k| \leq& |K\mathbb{E}\left[a_0\{F(\hat{c}_k(s)) - F(c_k(s))\} + o(F(\hat{c}_k(s)) - F(c_k(s)))\right]| \\
&+ |K\mathbb{E}\left[a_1\{F(\hat{c}_{k-1}(s)) - F(c_{k-1}(s))\} + o(F(\hat{c}_{k-1}(s)) - F(c_{k-1}(s)))\right]| \\
=& |Ka_0|\left|\frac{nk}{K(n+1)} - \frac{k}{K}\right| + |Ka_1|\left|\frac{n(k-1)}{K(n+1)} - \frac{k-1}{K}\right| + o\left(n^{-1}\right) \\
=& O\left(n^{-1}\right).
\end{aligned}$$

□

Now, using these two lemmas, we prove the main result. Letting $u(s) = \mathbb{E}[U_i \mid s(\boldsymbol{X}_i) = s]$, we decompose $\hat{\tau}_k$ into two parts,

$$\hat{\tau}_k = \underbrace{\frac{1}{n} \sum_{\frac{(k-1)n}{K} < i \leq \frac{kn}{K}} U_{[i,n]} - u(s(\boldsymbol{X}_{[i,n]}))}_{\hat{\tau}_k^{(1)}} + \underbrace{\frac{1}{n} \sum_{i=1}^{n} \mathbf{1}\left\{\frac{(k-1)n}{K} < i \leq \frac{kn}{K}\right\} u(s(\boldsymbol{X}_{[i,n]}))}_{\hat{\tau}_k^{(2)}}. \tag{S7}$$

Consider the first term. By the general theory of induced order statistics presented in Bhattacharya (1974), $U_{[i,n]} - u(s(\boldsymbol{X}_{[i,n]}))$ for $i = 1, \cdots, n$ are independent of one another conditional on $\boldsymbol{X}_n = (\boldsymbol{X}_{[1,n]}, \cdots, \boldsymbol{X}_{[n,n]})$. Define the random variables $Z_{[i,n]}$ as distributed according to the joint conditional distribution $U_{[i,n]} - u(s(\boldsymbol{X}_{[i,n]})) \mid \boldsymbol{X}_n$. Then, we have

$$\hat{\tau}_k^{(1)} = \frac{1}{n} \sum_{\frac{(k-1)n}{K} < i \leq \frac{kn}{K}} Z_{[i,n]},$$

where $Z_{[i,n]}$ are conditionally independent and $\mathbb{E}[Z_{[i,n]}] = 0$ by construction. Therefore, by Assumption 5, we can utilize the Berry-Esseen Theorem. Define:

$$\sigma_1^2(n) = \frac{1}{n} \sum_{\frac{(k-1)n}{K} < i \leq \frac{kn}{K}} \mathbb{V}(Z_{[i,n]})$$

$$\rho_1(n) = \frac{1}{n} \sum_{\frac{(k-1)n}{K} < i \leq \frac{kn}{K}} \mathbb{E}(|Z_{[i,n]}|^3)$$

Then the Berry-Esseen Theorem states that for $W \sim N(0,1)$, we have:

$$d\left(\frac{\sqrt{n}\hat{\tau}_k^{(1)}}{\sqrt{\sigma_1^2(n)}}, W\right) \leq \frac{C_0}{\sqrt{n}} \left(\sigma_1^2(n)\right)^{-3/2} \rho_1(n)$$

where $d(\cdot,\cdot)$ is the Kolmogorov distance. Now define the asymptotic variance and third moment by:

$$\sigma_1^2 = \lim_{n\to\infty} \sigma_1^2(n)$$

$$\rho_1 = \lim_{n\to\infty} \rho_1(n)$$

Both quantities exist by the strong law of large numbers for functions of order statistics (see Theorem 4 of Wellner (1977)). Specifically, by the strong law, σ_1^2 and ρ_1 does not depend on $\boldsymbol{X}_n$ for all but at most a measure zero set of $\boldsymbol{X}_n$. Therefore, the Berry-Esseen theorem implies that:

$$\sqrt{n}\hat{\tau}_k^{(1)} \mid \boldsymbol{X}_n \xrightarrow{d} N(0, \sigma_1^2) \text{ with probability 1} \tag{S8}$$

Next, consider the second term of Equation (S7). To prove the convergence of this summation of a function of order statistics, we utilize Theorem 1 and Example 1 from Shorack (1972), which we restate in our notation below:

Theorem S1 (Shorack (1972)) *Consider an independently and identically distributed random sample $X_1, \cdots, X_n$ of size n from a cumulative distribution function F, and a function of bounded variation g such that $\mathbb{E}[g(X)^3] < \infty$. Define:*

$$T_n = \frac{1}{n}\sum_{i=1}^{n} J\left(\frac{i}{n}\right) g(X_{[i,n]})$$

where $X_{[i,n]}$ is the ith order statistics of the sample, and J is a function that is continuous except at a finite number of points at which $g(F^{-1})$ is continuous. Suppose that there exists $\delta > 0$ such that:

$$|J(t)| \leq M(t(1-t))^{-\frac{1}{6}+\delta} \;\; \forall 0 < t < 1$$

Then, we have:

$$\sqrt{n}(T_n - \mathbb{E}[T_n]) \xrightarrow{d} N(0, \sigma^2)$$

where $\sigma^2 = \lim_{n\to\infty} n\mathbb{V}(T_n) < \infty$.

Now, set $X_i = s(\boldsymbol{X}_i)$, $g(\cdot) = u(\cdot)$, and $J(t) = \mathbf{1}\{(k-1)n/K < tn \leq kn/K\}$. Then, we have $T_n = \hat{\tau}_k^2$. Assumption 5 guarantees $\mathbb{E}[g(X)^3] < \infty$. The function $J(t)$ is discontinuous only at the quantile points $t = k/K$ and $t = \frac{k-1}{K}$, and Assumption 4 guarantees the continuity of $g(F^{-1})$ at those points. The function J clearly satisfies the bounding condition with $\delta = 1/6$ and $M = 1$. Therefore, define the asymptotic variance as $\sigma_2^2 = \lim_{n\to\infty} n\mathbb{V}(\hat{\tau}_k^{(2)})$, and we can utilize Theorem S1 to show the following convergence:

$$\sqrt{n}(\hat{\tau}_k^{(2)} - \mathbb{E}(\hat{\tau}_k^{(2)})) \xrightarrow{d} N(0, \sigma_2^2) \tag{S9}$$

Now, we aim to combine the results given in Equations (S8) and (S9). Using Lemma S2, we can replace τ_k with $\mathbb{E}(\hat{\tau}_k)$ by adding a small bias term. Then, we have

$$\begin{aligned}\sqrt{n}(\hat{\tau}_k - \tau_k) &= \sqrt{n}(\hat{\tau}_k^{(1)} - \mathbb{E}(\hat{\tau}_k^{(1)})) + \sqrt{n}(\hat{\tau}_k^{(2)} - \mathbb{E}(\hat{\tau}_k^{(2)})) + O\left(n^{-1/2}\right) \\ &\xrightarrow{d} N(0, \sigma_1^2 + \sigma_2^2) \qquad\qquad \text{(S10)}\end{aligned}$$

where the last line follows from the application of Lemma S1 to the convergence results given in Equations (S8) and (S9).

Equivalently, we can write Equation (S10) as,

$$\sqrt{n}\frac{\hat{\tau}_k - \tau_k}{\sqrt{\sigma_1^2 + \sigma_2^2}} \xrightarrow{d} N(0, 1)$$

Now, note that by the law of total variance, we have that

$$\begin{aligned}n\mathbb{V}(\hat{\tau}_k) &= n\mathbb{E}[\mathbb{V}(\hat{\tau}_k^{(1)} + \hat{\tau}_k^{(2)} \mid \boldsymbol{X}_n)] + n\mathbb{V}[\mathbb{E}(\hat{\tau}_k^{(1)} + \hat{\tau}_k^{(2)} \mid \boldsymbol{X}_n)] \\ &= n\mathbb{E}[\mathbb{V}(\hat{\tau}_k^{(1)} \mid \boldsymbol{X}_n)] + n\mathbb{V}(\hat{\tau}_k^{(2)}) \\ &\to \sigma_1^2 + \sigma_2^2 \qquad\qquad \text{(S11)}\end{aligned}$$

Therefore, by Slutsky's lemma, we have that:

$$\frac{\hat{\tau}_k - \tau_k}{\sqrt{\mathbb{V}(\hat{\tau}_k)}} \xrightarrow{d} N(0, 1)$$

□

S4 Proof of Proposition 1

We prove this proposition by finding an example that satisfies it. Define $t(x) = \mathbb{E}(Y_i(1) - Y_i(0) \mid s(\boldsymbol{X}_i) = F^{-1}(x))$. Then, consider a scoring function s and a population such that:

$$t(x) = \begin{cases} 2 & x \geq F(c_k(s)) \\ 1 & x < F(c_k(s)) \end{cases}$$

Note that $t(x)$ is bounded everywhere but has a discontinuity. By definition of $\hat{c}_k(s)$, $F(\hat{c}_k(s))$ follows the Beta distribution with the shape and scale parameters equal to nk/K and $n - nk/K + 1$, respectively. Therefore, we have the following normal approximation:

$$\sqrt{n+1}\left(F(\hat{c}_k(s)) - \frac{nk}{K(n+1)}\right) \xrightarrow{d} N\left(0, \frac{k}{K}\left(1 - \frac{k}{K}\right)\right)$$

In particular, as $n \to \infty$, $F(\hat{c}_k(s))$ is distributed approximately symmetric around $F(c_k(s)) = \frac{k}{K}$ with an error of $O(n^{-1})$ and has a standard deviation of $O(n^{-1/2})$. Thus, we have,

$$\begin{aligned}
\mathbb{E}(\hat{\tau}_k) - \tau_k &= K\mathbb{E}\left[\int_{F(c_k(s))}^{F(\hat{c}_k(s))} f(x)\mathrm{d}x\right] + K\mathbb{E}\left[\int_{F(c_{k-1}(s))}^{F(\hat{c}_{k-1}(s))} f(x)\mathrm{d}x\right] \\
&= (2-1)O\left(n^{-1/2}\right) + (1-1)O\left(n^{-1/2}\right) + O\left(n^{-1}\right) \\
&= O\left(n^{-1/2}\right)
\end{aligned}$$

We can now conclude $\sqrt{n}(\mathbb{E}(\hat{\tau}_k) - \tau_k) \not\to 0$. □

S5 Proof of Theorem 3

We wish to prove that for $\hat{\boldsymbol{\tau}} = (\hat{\tau}_1, \cdots, \hat{\tau}_K)$, $\boldsymbol{\tau} = (\tau_1, \cdots, \tau_K)$, and $\boldsymbol{\Sigma}_n = \mathbb{V}(\hat{\boldsymbol{\tau}})$, we have:

$$\boldsymbol{\Sigma}_n^{-1/2}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}) \xrightarrow{d} N(0, \boldsymbol{I})$$

where $\boldsymbol{I}$ is the $K \times K$ identity matrix.

By Equation (S10) in the proof of Theorem 2, for all $k = 1, \cdots, k$ we have,

$$\sqrt{n}(\hat{\tau}_k - \tau_k) \xrightarrow{d} N(0, \sigma_k^2),$$

where $\sigma_k^2 = \lim_{n\to\infty} n\mathbb{V}(\hat{\tau}_k)$. To prove the multi-dimensional result, we utilize the Cramer-Wold device, which we restate below:

Theorem S2 (Cramér and Wold (1936)) *Let $\boldsymbol{X}_n = (X_{n1}, \cdots, X_{nk})$ and $\boldsymbol{X} = (X_1, \cdots, X_k)$ be k-dimensional random vectors. Then $\boldsymbol{X}_n \to \boldsymbol{X}$ if and only if for all $(t_1, \cdots, t_k) \in \mathbb{R}^k$, we have:*

$$\sum_{i=1}^{k} t_i X_{ni} \xrightarrow{d} \sum_{i=1}^{k} t_i X_i$$

Now, consider $\boldsymbol{t} = (t_1, \cdots, t_K) \in \mathbb{R}^K$ and $\hat{\tau}_{\boldsymbol{t}} = \sum_{k=1}^{K} t_k \hat{\tau}_k$. Then, we can write $\hat{\tau}_{\boldsymbol{t}}$ as:

$$\hat{\tau}_{\boldsymbol{t}} = \frac{1}{n}\sum_{i=1}^{n}\left(\sum_{k=1}^{K} t_k \mathbf{1}\left\{\frac{(k-1)n}{K} < i \leq \frac{kn}{K}\right\}\right) U_{[i,n]}$$

where $U_{[i,n]}$ is defined in Equation (S5). We use the same proof strategy as the one used to prove Theorem 1. We define $u(s) = \mathbb{E}[U_i \mid s(\boldsymbol{X}_i) = s]$ and write:

$$\hat{\tau}_{\boldsymbol{t}} = \underbrace{\frac{1}{n}\sum_{k=1}^{K} t_k \sum_{\frac{(k-1)n}{K} < i \leq \frac{kn}{K}} U_{[i,n]} - u(s(\boldsymbol{X}_{[i,n]}))}_{\hat{\tau}_{\boldsymbol{t}}^{(1)}}$$
$$+ \underbrace{\frac{1}{n}\sum_{i=1}^{n}\left(\sum_{k=1}^{K} t_k \mathbf{1}\left\{\frac{(k-1)n}{K} < i \leq \frac{kn}{K}\right\}\right) u(s(\boldsymbol{X}_{[i,n]}))}_{\hat{\tau}_{\boldsymbol{t}}^{(2)}}$$

Using Lindberg's Central Limit Theorem for the conditional distribution of $\hat{\tau}_{\boldsymbol{t}}^{(1)}$ given $\boldsymbol{X}_n$ and applying Theorem S1 to $\hat{\tau}_{\boldsymbol{t}}^{(2)}$ yield,

$$\sqrt{n}(\hat{\tau}_{\boldsymbol{t}}^{(1)} - \mathbb{E}(\hat{\tau}_{\boldsymbol{t}}^{(1)})) \mid \boldsymbol{X}_n \xrightarrow{d} N(0, \sigma_1^2)$$
$$\sqrt{n}(\hat{\tau}_{\boldsymbol{t}}^{(2)} - \mathbb{E}(\hat{\tau}_{\boldsymbol{t}}^{(2)})) \xrightarrow{d} N(0, \sigma_2^2)$$

where $\sigma_1^2 = \lim_{n\to\infty} n\mathbb{V}(\hat{\tau}_{\boldsymbol{t}}^{(1)} \mid \boldsymbol{X}_n)$ and $\sigma_2^2 = \lim_{n\to\infty} n\mathbb{V}(\hat{\tau}_{\boldsymbol{t}}^{(2)})$. This implies,

$$\sqrt{n}(\hat{\tau}_{\boldsymbol{t}} - \tau_{\boldsymbol{t}}) = \sqrt{n}\left(\hat{\tau}_{\boldsymbol{t}}^{(1)} - \mathbb{E}[\hat{\tau}_{\boldsymbol{t}}^{(1)}]\right) + \sqrt{n}\left(\hat{\tau}_{\boldsymbol{t}}^{(2)} - \mathbb{E}[\hat{\tau}_{\boldsymbol{t}}^{(2)}]\right) + O\left(n^{-1/2}\right) \xrightarrow{d} N(0, \sigma_1^2 + \sigma_2^2)$$

where the result follows from Lemma S1. Since we can easily show $n\mathbb{V}(\hat{\tau}_t) \to \sigma_1^2 + \sigma_2^2$, we have: $\sqrt{n}(\hat{\tau}_{\boldsymbol{t}} - \tau_{\boldsymbol{t}}) \to N(0, \lim_{n\to\infty} n\mathbb{V}(\hat{\tau}_{\boldsymbol{t}}))$. Therefore, by the Cramer-Wold device (Theorem S2), we have $\sqrt{n}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}) \to N(0, \lim_{n\to\infty} n\boldsymbol{\Sigma}_n)$. Finally, Slutsky's Lemma implies the desired result,

$$\boldsymbol{\Sigma}_n^{-1/2}(\hat{\boldsymbol{\tau}} - \boldsymbol{\tau}) \to N(0, \boldsymbol{I}).$$

To derive the expression for the covariance matrix $\boldsymbol{\Sigma}_{kk'}$, we utilize the same approach as the one used in the proof of Theorem 1. We first apply the law of total covariance to obtain:

$$\begin{aligned}
&\operatorname{Cov}(\hat{\tau}_k, \hat{\tau}_{k'}) \\
&= K^2 \operatorname{Cov}\left(\frac{1}{n}\sum_{i=1}^{n}\{Y_{ki}^*(1) - Y_{ki}^*(0)\}, \frac{1}{n}\sum_{i=1}^{n}\{Y_{k'i}^*(1) - Y_{k'i}^*(0)\}\right) \\
&\quad + K^2\mathbb{E}\left[\operatorname{Cov}\left\{\frac{1}{n_1}\sum_{i=1}^{n}\left(\hat{f}_k(\boldsymbol{X}_i) - \frac{1}{K}\right)T_iY_i(1) - \frac{1}{n_0}\sum_{i=1}^{n}\left(\hat{f}_k(\boldsymbol{X}_i) - \frac{1}{K}\right)(1-T_i)Y_i(0),\right.\right. \\
&\qquad \left.\left.\frac{1}{n_1}\sum_{i=1}^{n}\left(\hat{f}_{k'}(\boldsymbol{X}_i) - \frac{1}{K}\right)T_iY_i(1) - \frac{1}{n_0}\sum_{i=1}^{n}\left(\hat{f}_{k'}(\boldsymbol{X}_i) - \frac{1}{K}\right)(1-T_i)Y_i(0) \;\middle|\; \boldsymbol{X}, \{Y_i(1), Y_i(0)\}_{i=1}^{n}\right\}\right].
\end{aligned} \tag{S12}$$

Applying Neyman's finite sample variance analysis to the second term shows that this term is equal to:

$$K^2\mathbb{E}\left\{\frac{1}{n}\left(\frac{n_0}{n_1}S_{kk'1}^{*2} + \frac{n_1}{n_0}S_{kk'0}^{*2} + 2S_{kk'01}^*\right)\right\}, \tag{S13}$$

where $S^*_{kk'01} = \sum_{i=1}^n (Y^*_{ki}(0) - \overline{Y^*_k(0)})(Y^*_{k'i}(1) - \overline{Y^*_{k'}(1)})/(n-1)$. Since $Y^*_{ki}(t)$ and $Y^*_{k'j}(t)$ are correlated, we have:

$$\begin{aligned}
&\operatorname{Cov}\left(\frac{1}{n}\sum_{i=1}^n (Y^*_{ki}(1) - Y^*_{ki}(0)), \frac{1}{n}\sum_{i=1}^n (Y^*_{k'i}(1) - Y^*_{k'i}(0))\right) \\
&= \operatorname{Cov}(Y^*_{ki}(1) - Y^*_{ki}(0), Y^*_{k'j}(1) - Y^*_{k'j}(0)) + \frac{1}{n}\mathbb{E}(S^{*2}_{kk'1} + S^{*2}_{kk'0} - 2S^*_{kk'01}). \qquad \text{(S14)}
\end{aligned}$$

For $k \neq k'$, we have:

$$\begin{aligned}
&\operatorname{Cov}(Y^*_{ki}(1) - Y_{ki}(0), Y^*_{k'j}(1) - Y_{k'j}(0)) \\
= &\operatorname{Cov}\left(\left(\hat{f}_k(\boldsymbol{X}_i) - \frac{1}{K}\right)\tau_i, \left(\hat{f}_{k'}(\boldsymbol{X}_j) - \frac{1}{K}\right)\tau_j\right) \\
= &\left(1 - \frac{2}{K}\right)\operatorname{Cov}(\hat{f}_k(\boldsymbol{X}_i)\tau_i, \hat{f}_{k'}(\boldsymbol{X}_j)\tau_j) - \frac{1}{K}\operatorname{Cov}((1 - \hat{f}_k(\boldsymbol{X}_i))\tau_i, \hat{f}_{k'}(\boldsymbol{X}_j)\tau_j) \\
&- \frac{1}{K}\operatorname{Cov}((1 - \hat{f}_{k'}(\boldsymbol{X}_i))\tau_i, \hat{f}_k(\boldsymbol{X}_j)\tau_j) \\
= &\left(1 - \frac{2}{K}\right)\left(\frac{1}{K^2}\kappa_{kk'11} - \frac{1}{K^2}\kappa_{k1}\kappa_{k'1}\right) \\
&- \frac{1}{K}\left\{\frac{n/K(n - n/K - 1)}{n(n-1)}\kappa_{kk'01} + \frac{n/K(n - n/K - 1)}{n(n-1)}\kappa_{kk'10} - \frac{1}{K^2}\kappa_{k1}\kappa_{k'0} - \frac{1}{K^2}\kappa_{k0}\kappa_{k'1}\right\} \\
= &\frac{1}{K^3}\left\{(K-2)\left(\kappa_{kk'11} - \kappa_{k1}\kappa_{k'1}\right) - \frac{Kn - n - 1}{n-1}\left(\kappa_{kk'10} + \kappa_{kk'01}\right) + \kappa_{k1}\kappa_{k'0} + \kappa_{k0}\kappa_{k'1}\right\}.
\end{aligned}$$

For $k = k'$, we have

$$\begin{aligned}
&\operatorname{Cov}(Y^*_{ki}(1) - Y^*_{ki}(0), Y^*_{kj}(1) - Y^*_{kj}(0)) \\
= &\operatorname{Cov}\left(\left(\hat{f}_k(\boldsymbol{X}_i) - \frac{1}{K}\right)\tau_i, \left(\hat{f}_k(\boldsymbol{X}_j) - \frac{1}{K}\right)\tau_j\right) \\
= &\left(1 - \frac{2}{K}\right)\operatorname{Cov}(\hat{f}_k(\boldsymbol{X}_i)\tau_i, \hat{f}_k(\boldsymbol{X}_j)\tau_j) - \frac{2}{K}\operatorname{Cov}((1 - \hat{f}_k(\boldsymbol{X}_i))\tau_i, \hat{f}_k(\boldsymbol{X}_j)\tau_j) \\
= &\left(1 - \frac{2}{K}\right)\left\{\frac{n-K}{K^2(n-1)}\kappa_{k11} - \frac{1}{K^2}\kappa_{k1}^2\right\} - \frac{2}{K}\left\{\frac{n(K-1)}{K^2(n-1)}\kappa_{k01} - \frac{1}{K^2}\kappa_{k1}\kappa_{k0}\right\} \\
= &\frac{1}{K^3}\left\{(K-2)\left(\frac{n-K}{n-1}\kappa_{k11} - \kappa_{k1}^2\right) - \frac{2n(K-1)}{(n-1)}\kappa_{k01} + 2\kappa_{k1}\kappa_{k0}\right\}.
\end{aligned}$$

Substituting Equations (S13) and (S14) into Equation S12, we obtain the desired covariance expression.

□

S6 Proof of Theorem 4

The proof of Theorem 3 above establishes that $\boldsymbol{\Sigma}^{-1/2}\hat{\boldsymbol{\tau}}$ is asymptotically normally distributed with the identity variance matrix $\boldsymbol{I}$. For simplicity, throughout this proof, we will assume that $\boldsymbol{\Sigma}^{-1/2}\hat{\boldsymbol{\tau}}$ is exactly normally distributed with unknown mean $\boldsymbol{\theta} = (\tau_1 - \tau, \cdots, \tau_K - \tau)$, i.e., $\boldsymbol{\Sigma}^{-1/2}\hat{\boldsymbol{\tau}} \sim N(\boldsymbol{\theta}, \boldsymbol{I})$.

Let the likelihood of the data $\hat{\boldsymbol{\tau}}$ under the null and alternative hypotheses as $L_{\hat{\boldsymbol{\tau}}}(H_0^C)$ and $L_{\hat{\boldsymbol{\tau}}}(H_1^C)$. Under the asymptotic normal assumption, the likelihood ratio is given by:

$$\frac{L_{\hat{\boldsymbol{\tau}}}(H_0^C)}{L_{\hat{\boldsymbol{\tau}}}(H_1^C)} = \begin{cases} \exp\left\{(\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}_1(\hat{\boldsymbol{\tau}}))^\top \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}_1(\hat{\boldsymbol{\tau}}))\right\} & \boldsymbol{\theta} \in \Theta_0 \\ \exp\left\{-(\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}_0(\hat{\boldsymbol{\tau}}))^\top \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}_0(\hat{\boldsymbol{\tau}}))\right\} & \boldsymbol{\theta} \in \Theta_1 \end{cases}$$

Where $\boldsymbol{\mu}_i(\hat{\boldsymbol{\tau}})$ are the optimal mean vectors given data $\hat{\boldsymbol{\tau}}$ for region j of the hypothesis test, and is the solution to the following optimization problems for $j \in \{0, 1\}$:

$$\boldsymbol{\mu}_j(\hat{\boldsymbol{\tau}}) = \underset{\boldsymbol{\mu} \in \Theta_j}{\operatorname{argmin}} \|\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}\|^2$$

We can identify the optimal means $(\boldsymbol{\mu}_1, \boldsymbol{\mu}_0)$ for each region of the hypothesis test through this optimization problem because the multivariate normal distribution is spherical and symmetric.

We use $(\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}_0(\hat{\boldsymbol{\tau}}))^\top \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}_0(\hat{\boldsymbol{\tau}}))$ as our test statistic. Note that when $\hat{\boldsymbol{\tau}} \in \Theta_0$, the statistic is always 0, so the null hypothesis is never rejected and thus we are consistent. Given that we have a composite test, we are interested in finding the uniformly most powerful test. This requires calculating the size of a test α, as a function of the critical value $C(\alpha)$:

$$\alpha = \sup_{\boldsymbol{\theta} \in \Theta_0} \Pr((\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}_0(\hat{\boldsymbol{\tau}}))^\top \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}_0(\hat{\boldsymbol{\tau}})) > C(\alpha) \mid \boldsymbol{\theta})$$

Since the supremum must occur at the boundary $\partial \boldsymbol{\Theta}_0$ of the polytope $\boldsymbol{\Theta}_0$ the set Θ_0, the probability of exceeding $C(\alpha)$ is maximized when the solid angle of the $\boldsymbol{\Theta}_0$ region is minimized. By considering the shape of the polytope $\boldsymbol{\Theta}_0$, we recognize that the boundary points, which minimize the solid angle, are precisely those on the boundary when all constraints are active:

$$\alpha = \sup_{t} \Pr((\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}_0(\hat{\boldsymbol{\tau}}))^\top \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}_0(\hat{\boldsymbol{\tau}})) > C(\alpha) \mid \tau_1 - \tau = \cdots = \tau_K - \tau = t).$$

We now note that we have translational invariance on this boundary, i.e., all points along $\tau_1 - \tau = \cdots = \tau_K - \tau$ have the same probability, yielding,

$$\alpha = \Pr((\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}_0(\hat{\boldsymbol{\tau}}))^\top \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}_0(\hat{\boldsymbol{\tau}})) > C(\alpha) \mid \tau_1 - \tau = \cdots = \tau_K - \tau = 0)$$

Therefore, to identify the value of α, we just need the CDF of the statistic $(\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}_0(\hat{\boldsymbol{\tau}}))^\top \boldsymbol{\Sigma}^{-1}(\hat{\boldsymbol{\tau}} - \boldsymbol{\mu}_0(\hat{\boldsymbol{\tau}}))$ when $\hat{\tau} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$. This can be easily estimated using Monte Carlo simulation. □

S7 Proof of Theorem 5

The derivation of bias is essentially identical to that given in Supplementary Appendix S1 and thus is omitted. To derive the variance, we first introduce the following useful lemma, adapted from Nadeau and Bengio (2000).

Lemma S3

$$\begin{aligned} \mathbb{E}(S_{Fk}^2) &= \mathbb{V}(\hat{\tau}_k^\ell) - \operatorname{Cov}(\hat{\tau}_k^\ell, \hat{\tau}_k^{\ell'}), \\ \mathbb{V}(\hat{\tau}_k(F, n-m)) &= \frac{\mathbb{V}(\hat{\tau}_k^\ell)}{L} + \frac{L-1}{L} \operatorname{Cov}(\hat{\tau}_k^\ell, \hat{\tau}_k^{\ell'}). \end{aligned}$$

where $\ell \neq \ell'$.

The lemma implies,

$$\mathbb{V}(\hat{\tau}_k(F, n-m)) \;=\; \mathbb{V}(\hat{\tau}_k^\ell) - \frac{L-1}{L}\mathbb{E}(S^2_{Fk}).$$

We then follow the same process of derivation as in Appendix S1 for the first term. The only difference occurs in the derivation of the covariance term:

$$\begin{aligned}
&\mathrm{Cov}(Y^\ell_{ki}(1) - Y^\ell_{ki}(0), Y^\ell_{kj}(1) - Y^\ell_{kj}(0)) \\
=&\mathbb{E}_\ell\left[\mathrm{Cov}_{\boldsymbol{X},Y}\left(\hat{f}^\ell_k(\boldsymbol{X}_i^{(\ell)})\tau_i, \hat{f}^\ell_k(\boldsymbol{X}_j^{(\ell)})\tau_j\right)\right] + \mathrm{Cov}_\ell\left[\mathbb{E}_{\boldsymbol{X},Y}[\hat{f}^\ell_k(\boldsymbol{X}_i^{(\ell)})\tau_i], \mathbb{E}_{\boldsymbol{X},Y}[\hat{f}^\ell_k(\boldsymbol{X}_j^{(\ell)})\tau_j]\right] \\
=&\mathbb{E}_\ell\left[\frac{(n-K)\kappa^\ell_{k11}}{K^2(n-1)} - \frac{1}{K^2}(\kappa^\ell_{k1})^2\right] + \frac{1}{K^2}\mathbb{V}_\ell\left(\kappa^\ell_{k1}\right).
\end{aligned}$$

□

S8 Proof of Consistency of $\widehat{\mathbb{E}(S^2_{Fk})}$

We show that $\widehat{\mathbb{E}(S^2_{Fk})}$ is consistent as L approaches infinity under the assumption that the fourth moments $\mathbb{E}(Y_i(t)^4) < \infty$ for $t = 0, 1$ and a sufficiently large value of m. Theorem 5 implies,

$$\begin{aligned}
\mathbb{V}(\hat{\tau}_k(F, n-m)) =\;& K^2\left(\frac{\mathbb{E}(S^2_{Fk1})}{m_1} + \frac{\mathbb{E}(S^2_{Fk0})}{m_0}\right) \\
&+ \mathbb{E}_\ell\left[\frac{(n-K)\kappa^\ell_{k11}}{K^2(n-1)} - \frac{1}{K^2}(\kappa^\ell_{k1})^2\right] + \mathbb{V}\left(\kappa^\ell_{k1}\right) - \frac{L-1}{L}\mathbb{E}(S^2_{Fk}),
\end{aligned}$$

Now, define:

$$\mathbb{V}(\hat{\tau}_k(\widehat{F, n-m})) = K^2\left(\frac{\widehat{\mathbb{E}(S^2_{Fk1})}}{m_1} + \frac{\widehat{\mathbb{E}(S^2_{Fk0})}}{m_0}\right) + \frac{(n-K)\widehat{\mathbb{E}[\kappa^\ell_{k11}]}}{K^2(n-1)} - \frac{1}{K^2}(\widehat{\mathbb{E}[\kappa^\ell_{k1}]})^2 + \widehat{\mathbb{V}\left(\kappa^\ell_{k1}\right)}$$

By construction, we have that as $m \to \infty$:

$$\frac{\mathbb{V}(\hat{\tau}_k(\widehat{F, n-m}))}{\mathbb{V}(\hat{\tau}_k(F, n-m)) + \frac{L-1}{L}\mathbb{E}(S^2_{Fk})} \xrightarrow{p} 1$$

Applying Lemma 1 from Nadeau and Bengio (2000) to $\hat{\tau}_k(F, n-m)$ gives

$$\mathbb{V}(\hat{\tau}_k(F, n-m)) \geq \; \mathbb{E}(S^2_{Fk}).$$

Therefore, we have:

$$\lim_{m\to\infty} \frac{\mathbb{V}(\hat{\tau}_k(\widehat{F, n-m}))}{\mathbb{E}(S^2_{Fk}) + \frac{L-1}{L}\mathbb{E}(S^2_{Fk})} \geq 1$$

Note that we can write $\widehat{\mathbb{E}(S^2_{Fk})}$ as:

$$\widehat{\mathbb{E}(S^2_{Fk})} \;=\; \min\left(S^2_{Fk}, \mathbb{V}(\hat{\tau}_k(\widehat{F, n-m}))\right).$$

By definition of S^2_{Fk}, if the fourth moments of $Y_i(t)$ exist, we have $\mathbb{V}(S^2_{Fk}) = O(L^{-1})$, and thus as $L \to \infty$:

$$\frac{S^2_{Fk}}{\mathbb{E}(S^2_{Fk})} \xrightarrow{p} 1$$

Let $\epsilon > 0$. There exists L_0 such that for all $L > L_0$, $|\frac{S^2_{Fk}}{\mathbb{E}(S^2_{Fk})} - 1| < \epsilon$. Similarly, there exists m_0 such that for all $\mathbb{V}(\hat{\tau}_k(\widehat{F, n-m})) > (1-\epsilon)\mathbb{E}(S^2_{Fk})\ \forall m > m_0$. Therefore, for all $m > m_0$, we have that:

$$\lim_{L\to\infty} \frac{\widehat{\mathbb{E}(S^2_{Fk})}}{\mathbb{E}(S^2_{Fk})} = \lim_{L\to\infty} \frac{\min\left(S^2_{Fk}, \mathbb{V}(\hat{\tau}_k(\widehat{F, n-m}))\right)}{\mathbb{E}(S^2_{Fk})} = \lim_{L\to\infty} \frac{S^2_{Fk}}{\mathbb{E}(S^2_{Fk})} \xrightarrow{p} 1$$

□

S9 Proof of Theorem 6

We first shows the bias is small.

Lemma S4 $\lim_{n\to\infty} |\hat{\tau}_k(F, n-m) - \tau_k(F, n-m)| = O\left(m^{-1}\right)$

Proof

$$\begin{aligned}
|\hat{\tau}_k(F, n-m) - \tau_k(F, n-m)| &\leq \frac{1}{L}\sum_{\ell=1}^{L} |\mathbb{E}(\hat{\tau}^\ell_k(F, n-m)) - \tau^\ell_k(F, n-m)| \\
&= \frac{1}{L}\sum_{\ell=1}^{L} \mathbb{E}_{\mathcal{Z}^{-\ell}}\left[O\left(m^{-1}\right)\right] \\
&= O\left(m^{-1}\right).
\end{aligned}$$

The first equality follows because the estimator for each fold $\hat{\tau}^\ell_k(F, n-m)$ is equivalent to the non-cross-fitting estimator under m samples and so Lemma S2 is applicable. The second equality follows from Assumption 6. □

We first write:

$$\hat{\tau}_k(F, n-m) = \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}\left\{\frac{(k-1)m}{K} < i \leq \frac{km}{K}\right\} U_{[i,m]}$$

where $U_{[i,m]} \in \mathbb{R}$ is defined as,

$$U_{[i,m]} := \frac{1}{L}\sum_{\ell=1}^{L} K\hat{f}^\ell_k(\boldsymbol{X}^{(\ell)}_{[i,m]})Y^{(\ell)}_{[i,m]}\left(\frac{T^{(\ell)}_{[i,m]}}{q} - \frac{1-T^{(\ell)}_{[i,m]}}{1-q}\right).$$

and $(Y^{(\ell)}_{[i,m]}, T^{(\ell)}_{[i,m]}, \boldsymbol{X}^{(\ell)}_{[i,m]})$ are ordered separately for each split ℓ such that:

$$s^\ell(\boldsymbol{X}^{(\ell)}_{[i,m]}) \leq s^\ell(\boldsymbol{X}^{(\ell)}_{[i,m]}) \leq \cdots \leq s^\ell(\boldsymbol{X}^{(\ell)}_{[i,m]})$$

Now by Assumption 7, there exists a fixed scoring rule $s(\boldsymbol{X})$ and corresponding treatment rule $f_k(\boldsymbol{X}_i) = \mathbf{1}\{s(\boldsymbol{X}_i) > c_{k-1}(s)\} - \mathbf{1}\{s(\boldsymbol{X}_i) > c_k(s)\}$ such that we can write:

$$U_{[i,m]} = \tilde{U}_{[i,m]} + \epsilon_{[i,m]}$$

$$\tilde{U}_{[i,m]} := \frac{1}{L}\sum_{\ell=1}^{L} K f_k(\boldsymbol{X}_{[i,m]}^{(\ell)}) Y_{[i,m]}^{(\ell)} \left(\frac{T_{[i,m]}^{(\ell)}}{q} - \frac{1 - T_{[i,m]}^{(\ell)}}{1-q} \right)$$
$$\tilde{\tau}_k(F, n-m) = \frac{1}{m}\sum_{i=1}^{m} \mathbf{1}\left\{ \frac{(k-1)m}{K} < i \leq \frac{km}{K} \right\} \tilde{U}_{[i,m]}$$

where

$$\begin{aligned}
\mathbb{E}[\epsilon_{[i,m]}] &= \mathbb{E}\left[\frac{1}{L}\sum_{l=1}^{L} K(f_k(\boldsymbol{X}_{[i,m]}^{(\ell)}) - \hat{f}_k^{\ell}(\boldsymbol{X}_{[i,m]}^{(\ell)})) Y_{[i,m]}^{(\ell)} \left(\frac{T_{[i,m]}^{(\ell)}}{q} - \frac{1 - T_{[i,m]}^{(\ell)}}{1-q} \right) \right] \\
&\leq \sqrt{\mathbb{E}\left[\left(\frac{1}{L}\sum_{l=1}^{L} K(f_k(\boldsymbol{X}_{[i,m]}^{(\ell)}) - \hat{f}_k^{\ell}(\boldsymbol{X}_{[i,m]}^{(\ell)})) Y_{[i,m]}^{(\ell)} \left(\frac{T_{[i,m]}^{(\ell)}}{q} - \frac{1 - T_{[i,m]}^{(\ell)}}{1-q} \right) \right)^2 \right]} \\
&= o\left(m^{-1/2}\right)
\end{aligned}$$

Then, we can apply the proof of Theorem 2 on $\tilde{U}_{[i,m]}$ as f_k is fixed, which gives:

$$\frac{\tilde{\tau}_k(F, n-m) - \mathbb{E}[\tilde{\tau}_k(F, n-m)]}{\sqrt{\mathbb{V}(\tilde{\tau}_k(F, n-m))}} \to N(0, 1)$$

Since $\mathbb{V}(\tilde{\tau}_k(F, n-m)) = \mathbb{V}(\hat{\tau}_k(F, n-m)) + o(m^{-1})$ and $\hat{\tau}_k(F, n-m) = \tilde{\tau}_k(F, n-m) + o(m^{-1/2})$, we have:

$$\frac{\hat{\tau}_k(F, n-m) - \tau_k(F, n-m)}{\sqrt{\mathbb{V}(\hat{\tau}_k(F, n-m))}} \to N(0, 1)$$

$\square$

S10 Proof of Theorem 7

The proof follows identically to the proof of Theorem 3 by applying the Cramer-Wold Device in Theorem S2 to the sequence $\sum_{k=1}^{K} t_k \hat{\tau}_k(F, n-m)$.