Nonparametric Inference on Treatment-Biomarker Interaction Based on Probability Index

Zehui Wang, Yanglei Song, Wenyu Jiang and Dongsheng Tu

Queen's University

Supplementary Material

The supplementary materials include additional simulation results and proofs of all theorems and lemmas in the main text. Section S1 contains additional simulation results for unknown cutpoint problem. Section S2 presents additional simulation results for problem with prespecified cutpoint. Section S3 introduces the procedure based on the minimum p-value method. Section S4 presents additional results on the advanced colorectal cancer dataset and extensions. Section S5 establishes notations for clarity and coherence. Section S6 includes proofs of Theorems 1 and 2. Section S7 contains proofs of Theorems 3 and 4 regarding hypothesis testing with unknown cutpoint. Section S8 includes proofs of Theorems 5 and 6 regarding the limiting distribution of the cutpoint, as well as some supporting lemmas.

S1 Additional Discussions and Simulation Results for Unknown Cutpoint Problem

We describe another data-driven algorithm, referred to as Method S2, that determines m for constructing confidence intervals for the optimal cutpoint in Subsection 5.1:

1. Consider a sequence of m's of the form $m_j = \lfloor q^j n \rfloor$ for $j = 1, 2, ...,$ and $q \in (0, 1)$;

2. For each m_j-out-of-n bootstrap, compute $\hat{c}^*_{m_j,b}$ for b-th replication. Construct the bootstrap empirical cumulative distribution function

$$\hat{F}_{m_j}(x) = \frac{1}{B} \sum_{b=1}^{B} I\{m_j^{1/3}(\hat{c}^*_{m_j,b} - \hat{c}_n) \leq x\}.$$

3. The m will be selected as the value that minimizes the supremum difference between two adjacent bootstrap empirical cumulative distributions:

$$m = \text{argmin}_{m_j} \sup_x |\hat{F}_{m_j}(x) - \hat{F}_{m_{j+1}}(x)|,$$

where in the case of ties, we select the largest one.

Then, we report the empirical bias and standard error of the proposed profile estimator $\hat{c}_n$ in (4.10) for estimating c_b under alternative hypotheses. The simulation settings are identical to those used in the power analysis in

Table S1: Empirical bias and standard errors of the estimate of c_b under alternative hypothesis.

		n=300		n=500	
	c_b	Bias	SE	Bias	SE
6a	0.5	-0.0104	0.2481	0.0079	0.2325
6b	0.7	-0.1295	0.2615	-0.0846	0.242
7a	0.5	-0.0587	0.1124	-0.0398	0.0933
7b	0.7	-0.015	0.117	0.005	0.0587
8	0.5	-0.0026	0.0722	-0.0061	0.0605
9	0.52	-0.0099	0.1300	-0.0046	0.1132

Subsection 6.2, with the addition of two new cases, 6b and 7b listed in Table 1, where $c_b = 0.7$.

From Table S1, we observe that our method is accurate for estimating c_b across most cases, with small biases and standard errors. However, with a smaller sample size, such as in Case 6a and 6b, it is less accurate. This might be attributed to the small values of $|\theta_{c_b}|$ (i.e. around 0.1), as listed in Table 1, where the optimal cutpoint might be challenging to identify.

S2 Additional Simulation Results for Problem with prespecified cutpoint

In this section, additional simulation results when c is prespecified are presented. For each set of combination, simulations are repeated $R = 1000$ times. Three different sample sizes ($n = 100, 300, 500$) are considered for scenarios 1-9 described in Table 1 in Section 6. It is worth noting that the prespecified cutpoint $c_0 = 0.5$ for Cases 1-8 and $c_0 = 0.52$ for Case 9 are indeed the optimal cutpoints c_b under the alternative hypothesis in our settings. We calculate the empirical size and empirical power of the test according to Theorem 2. As summarized in Table S2, for configurations 1-5, the empirical size of the test remains close to the nominal 5% level and exhibits minimal change with increasing sample size. The empirical power exceeds 85% in most cases, and increases with the sample size and with the increase in true $|\theta_{c_b}|$, as listed in Table 1.

S3 Minimum p-value Method

In this section, we explore another method commonly used in the analysis of data from clinical trials known as the minimum p-value method. Here, we define our test statistic as the most significant one, given by

Table S2: Empirical size and power (in percentage) of the test under null hypothesis and alternative hypothesis when cutpoint c is given for significance level $\alpha = 0.05$. Here configurations 1-5 are for size, and configurations 6a-9 are for power.

n	1	2	3	4	5	6a	7a	8	9
100	4.5	4.4	4	3.7	3.5	14.3	87	100	100
300	4.9	6	5.4	5	4.3	31.2	100	100	100
500	4.5	4.5	4.2	4.2	4.4	49.4	100	100	100

$\sup_{c\in[\ell,u]} \left|\frac{\sqrt{n}\hat{\theta}_{n,c}}{\sqrt{\hat{\sigma}^2_{n,c}}}\right|$. Meanwhile, for the bootstrap method outlined in Section 4.1, we substitute the statistic $|\hat{\theta}_{n,c}|$ by $\left|\frac{\sqrt{n}\hat{\theta}_{n,c}}{\sqrt{\hat{\sigma}^2_{n,c}}}\right|$ and use it as the basis of the bootstrap to test whether the predictive effects exist or not. Upon rejecting the null hypothesis, the optimal cutpoint $\tilde{c}_n$ is determined as the value that yields the lowest among all calculated p-values by Theorem 2's method, expressed as:

$$\tilde{c}_n = \arg\min_{c\in[\ell,u]} 2\left[1 - \Phi\left(\left|\frac{\sqrt{n}\hat{\theta}_{n,c}}{\sqrt{\hat{\sigma}^2_{n,c}}}\right|\right)\right] = \arg\max_{c\in[\ell,u]} |\mu_c|, \quad \mu_c = \frac{\sqrt{n}\hat{\theta}_{n,c}}{\sqrt{\hat{\sigma}^2_{n,c}}}, \tag{S3.1}$$

where Φ is the distribution function of a standard normal variable.

Simulation studies are conducted to assess the performance of the minimum p-value method and the corresponding bootstrap test. The setups and configurations under null hypothesis and alternative hypothesis remain identical to those detailed in Section 6.

Table S3: The empirical size (in percentage) and power of minimum p-value method under null hypothesis and alternative hypothesis when the cutpoint is not given for significance level $\alpha = 0.05$. Configurations 1-5 focus on size, while configurations 6-9, where c_b is considered to be 0.5 for configurations 6-8 and 0.52 for configuration 9, are for empirical power.

n		1	2	3	4	5	6a	7a	8	9
300	PB	4.9	4.8	4.6	4.2	5.2	14.4	98.3	100	100
	WB	33.8	35.2	32	35.7	35.2	63.3	100	100	100
500	PB	4.7	4.8	4.5	6.9	5.9	21.2	100	100	100
	WB	39.7	37.5	30.4	38.2	38.4	73	100	100	100

Table S4: Empirical bias and standard errors of the estimate of c_b under alternative hypothesis.

		n=300		n=500	
	c_b	Bias	SE	Bias	SE
6a	0.5	0.0095	0.226	0.0126	0.0872
6b	0.7	-0.1143	0.233	0.0151	0.0667
7a	0.5	-0.0002	0.2035	0.001	0.0524
7b	0.7	-0.0948	0.2182	0.0143	0.044
8	0.5	-0.0188	0.1098	-0.0134	0.0938
9	0.52	-0.0093	0.2070	-0.0030	0.1934

Table S3 provides the empirical size and power of the bootstrap test (PB) based on the minimum p-value method. The results are similar to those by the profile method, where the empirical size of PB is close to the nominated significant level 5% under the null hypothesis and the empirical power increases as sample sizes increase. The minimum p-value method also yields precise estimates, as indicated in Table S4, where the empirical bias and standard errors demonstrate relatively small values for most configurations. The trend of increasing accuracy is also seen when the sample size is increasing and true $|\theta_{c_b}|$ is larger.

S4 Additional Results on Real Data and Extensions

S4.1 Additional Results on Real Data

In this section, we present additional results on the advanced colorectal cancer dataset.

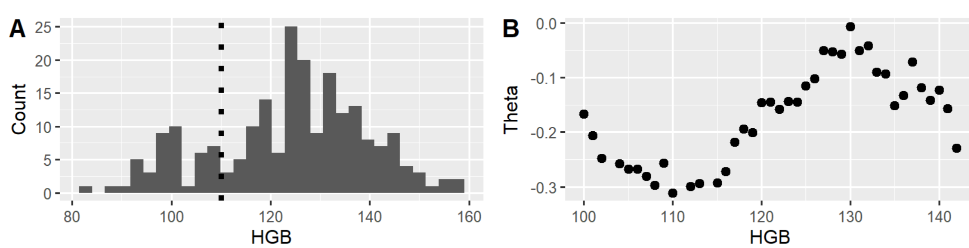

Figure 1: (A): The histogram of HGB from baseline at 8 weeks, the vertical dash line indicates the cutpoint estimated (110) by profile method and the minimum p-value method. (B): The scatter plot illustrates the trend of $\hat{\theta}_{n,c}$ at each value of HGB from baseline at 8 weeks, with the most significant absolute value observed at 110.

Table S5: Estimated $\hat{c}_n$ and $\tilde{c}_n$ for EREG, LDH, ALKPH and HGB by profile method and minimum p-value method respectively, and the corresponding p-values based on the bootstrap method. The bootstrap repetition number is B=2000.

EREG					LDH				
Week	$\hat{c}_n$	$p_{profile}$	$\tilde{c}_n$	p_{mini}	Week	$\hat{c}_n$	$p_{profile}$	$\tilde{c}_n$	p_{mini}
4	8.12	0.190	8.12	0.183	4	277	0.691	270	0.685
8	8.02	0.232	8.02	0.231	8	328	0.183	328	0.176
16	4.36	0.281	4.36	0.292	16	770	0.682	770	0.677
24	4.37	0.400	4.89	0.470	24	245	0.083	245	0.080
ALKPH					HGB				
Week	$\hat{c}_n$	$p_{profile}$	$\tilde{c}_n$	p_{mini}	Week	$\hat{c}_n$	$p_{profile}$	$\tilde{c}_n$	p_{mini}
4	180	0.500	180	0.500	4	112	0.192	110	0.230
8	116	0.655	116	0.626	8	110	0.007	110	0.005
16	116	0.380	116	0.370	16	119	0.662	119	0.646
24	89	0.250	91	0.299	24	118	0.087	118	0.083

Table S6: Subgroup analysis of HGB based on $\hat{c}_n$ with respect to the change score in the physical function score at 8 weeks.

		Cetuximab+BSC		BSC		
Biomarker	Value	n	Mean	n	Mean	Difference
HGB	<110	27	2.78	17	-23.53	26.31
	≥110	99	-4.21	72	-4.91	0.7

S4.2 Model Dependence, Transformation Sensitivity and Extensions for Multiple Biomarkers

The proposed index

$$\theta_c = \Pr(X_1 \leq Y_2 \mid Z_1 > c, Z_2 > c) - \Pr(X_1 \leq Y_2 \mid Z_1 \leq c, Z_2 \leq c)$$

quantifies the difference in relative treatment effects between two subgroups formed by dichotomizing the biomarker at a single threshold c. Hence, θ_c measures the degree of predictive heterogeneity that can be captured through this specific subgrouping rule. If the true treatment-biomarker interaction depends on a more complex transformation $f(Z)$ rather than directly on Z, it is indeed possible that $\theta_c = 0$ for $c \in [\ell, u]$ even when heterogeneity exists with respect to $f(Z)$. This reflects a limitation of the dichotomization scheme. One could define

$$\theta_c^{(f)} = \Pr(X_1 \leq Y_2 \mid f(Z_1) > c, f(Z_2) > c) - \Pr(X_1 \leq Y_2 \mid f(Z_1) \leq c, f(Z_2) \leq c),$$

which extends the framework to transformations or even multiple biomarkers.

Remark 1. θ_c is based on rank comparisons of biomarker values and is therefore invariant under any strictly monotone transformation of Z (e.g., logarithmic or percentile transformations) that preserves ordering. Specifically, if f is strictly increasing, then $\theta_{f(c)}^{(f)} = \theta_c$, implying that such transformations leave the null hypothesis in (1.3) unchanged. Furthermore, c^* is a maximizer of $|\theta_c|$ if and only if $f(c^*)$ is a maximizer of $|\theta_c^{(f)}|$.

In practice, the groups can be formed by multiple biomarkers, which is an area of growing interest that requires further practical investigation. Below, we outline some potential extensions for handling multiple biomarkers.

A straightforward approach is to test each biomarker individually and apply multiple testing correction methods, such as the Bonferroni correction, to control the overall Type I error rate. For example, in Section 7 of our real data application, we analyzed four biomarkers and their association with changes in the Physical Function Scale (PFS) from baseline at 8 weeks. With four hypotheses and a desired overall significance level of $\alpha = 0.05$, the Bonferroni correction adjusts the individual significance level to $\alpha/4 = 0.0125$. In this case, HGB remains a significant biomarker, as its p-value of 0.007 falls below the adjusted threshold.

Another approach involves combining multiple biomarkers into a single score, as proposed by Cai et al. (2010). This aggregated score can then be treated as a new "biomarker", and our proposed procedure can be applied to assess its association with the outcome of interest. This method simplifies the analysis while accounting for the combined effects of multiple biomarkers.

A more flexible approach is to consider a hyperplane that separates groups formed by multiple biomarkers. Specifically, we can define the model as follows:

$$\theta_c = \Pr(X_1 \leq Y_2 | g(\boldsymbol{\gamma}^T \boldsymbol{Z}_1) > c, g(\boldsymbol{\gamma}^T \boldsymbol{Z}_2) > c) - \Pr(X_1 \leq Y_2 | g(\boldsymbol{\gamma}^T \boldsymbol{Z}_1) \leq c, g(\boldsymbol{\gamma}^T \boldsymbol{Z}_2) \leq c),$$

where $\boldsymbol{Z} = (Z_1, \ldots, Z_p)$ is a p-dimensional vector of biomarker measurements, $\boldsymbol{\gamma}$ is a vector used to combine the biomarkers, c is the unknown cutpoint, and $g(\cdot)$ can be some link functions. This strategy allows for multidimensional group formation. Moreover, the interpretation is straightforward: if the combination is linear (i.e., without $g(\cdot)$), the signs of the parameters in $\boldsymbol{\gamma}$ suggest whether a biomarker contributes to forming a treatment-sensitive group or not. However, estimating the hyperplane efficiently presents challenges, requiring identifiability conditions, approximation techniques, and dimensional reduction methods; see, e.g., Fan et al. (2017); Li et al. (2021).

One could also construct a predictive tree based on the probabilistic index, where decision nodes represent splits informed by individual biomarkers. While this method provides an interpretable structure, integrating our testing procedure into the tree framework introduces challenges, particularly in maintaining the statistical properties of the test.

S4.3 Adjustment for Confounding Variables

The proposed method is developed in a predictive and associational framework. In applications such as randomized clinical trials, the treatment indicator U is independent of baseline variables by design, but the biomarker Z is a baseline characteristic and may still depend on other covariates $\boldsymbol{B}$. When $\boldsymbol{B}$ influences both Z and the outcomes under each treatment, the marginal measure

$$\theta_c = \Pr(X_1 \leq Y_2 \mid Z_1 > c, Z_2 > c) - \Pr(X_1 \leq Y_2 \mid Z_1 \leq c, Z_2 \leq c)$$

can reflect heterogeneity induced by the association between $\boldsymbol{B}$ and Z, in addition to the predictive contribution of the biomarker itself. To isolate the biomarker-specific component, one can define the conditional quantity

$$\theta_{c|\boldsymbol{B}_1,\boldsymbol{B}_2} = \Pr(X_1 \leq Y_2 \mid Z_1 > c, Z_2 > c, \boldsymbol{B}_1, \boldsymbol{B}_2) - \Pr(X_1 \leq Y_2 \mid Z_1 \leq c, Z_2 \leq c, \boldsymbol{B}_1, \boldsymbol{B}_2),$$

and its population average $\theta_c^{\mathrm{adj}} = E_{\boldsymbol{B}_1,\boldsymbol{B}_2}[\theta_{c|\boldsymbol{B}_1,\boldsymbol{B}_2}]$, which adjusts for the distribution of baseline covariates and represents the biomarker's predictive contribution after accounting for baseline imbalance. Both quantities are associational rather than causal, as the probabilistic index compares outcome distributions without invoking potential-outcome assumptions.

Directly conditioning on high-dimensional $\boldsymbol{B}$ can be challenging. A practical approach is to embed the probabilistic index within a regression framework. Following Thas et al. (2012); De Schryver and De Neve (2019), one may specify a probabilistic-index model for

$$\Pr(X_1 \le Y_2 \mid Z_1, Z_2, \boldsymbol{B}_1, \boldsymbol{B}_2) = m\big(\alpha\, I(Z_1 > c, Z_2 > c) + \boldsymbol{\beta}_1^\top \boldsymbol{B}_1 + \boldsymbol{\beta}_2^\top \boldsymbol{B}_2 + \gamma\big),$$

where $m(\cdot)$ is a user-specified link function mapping $\mathbb{R}$ to $(0,1)$ and $(\alpha, \boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \gamma)$ are regression parameters. For fixed $\boldsymbol{B}_1$ and $\boldsymbol{B}_2$, a nonzero α indicates differential probabilistic treatment effects between the subgroups defined by c, after adjusting for covariates. This regression formulation is parametric and relies on correct model specification.

Alternatively, one may estimate the conditional probability index semiparametrically or nonparametrically using flexible machine learning (ML) methods (Chernozhukov et al., 2018; Mi et al., 2021). In particular, a *double/debiased machine learning* (DML) approach could target θ_c^{adj} by combining cross-fitted ML estimates of the nuisance components, such as $\Pr(X_1 \le Y_2 \mid Z_1, Z_2, \boldsymbol{B}_1, \boldsymbol{B}_2)$ and $\Pr(Z > c \mid \boldsymbol{B})$, in an orthogonal estimating equation. Such an extension would provide a flexible adjustment framework while preserving the associational interpretation of θ_c^{adj}. We view both the regression-based and DML-based extensions as interesting directions for future research.

S4.4 Power Considerations and Potential for Hybrid Testing

The proposed procedure is built upon the probabilistic index, which naturally accommodates both continuous and ordinal outcomes. When X and Y are ordinal, their numerical means are not well-defined or meaningful, whereas the event $\{X \leq Y\}$ remains well defined. This makes it challenging, in general, to combine rank-based and mean-based approaches within a unified framework. When X and Y are numerical, however, one could consider a mean-based contrast such as

$$\tilde{\theta}_c = \mathrm{E}[X_1 - Y_2 \mid Z_1 > c, Z_2 > c] - \mathrm{E}[X_1 - Y_2 \mid Z_1 \leq c, Z_2 \leq c],$$

and develop hybrid statistics that incorporate both θ_c and $\tilde{\theta}_c$ to balance robustness (to heavy-tailed distributions) and power.

A second avenue is to consider semiparametric methods such as the density ratio model (Fokianos and Troendle, 2007; Jiang and Tu, 2012; Jiang et al., 2016), which specifies a parametric link between two distributions while retaining flexibility. Specifically, let $f_X^{c+}(x)$ and $f_Y^{c+}(y)$ denote the densities of X and Y in the biomarker-positive subgroup ($Z > c$), with $f_X^{c-}(x)$ and $f_Y^{c-}(y)$ defined analogously for the biomarker-negative subgroup.

A density ratio model assumes

$$f_X^{c+}(x) = \exp(\alpha + \boldsymbol{\beta}^\top \boldsymbol{h}(x)) f_Y^{c+}(x),$$

$$f_X^{c-}(x) = \exp(\gamma + \boldsymbol{\lambda}^\top \boldsymbol{g}(x)) f_Y^{c-}(x),$$

where $\boldsymbol{h}(x)$ and $\boldsymbol{g}(x)$ are known link functions, and $(\alpha, \boldsymbol{\beta}, \gamma, \boldsymbol{\lambda})$ are parameters. Estimation can proceed via profile likelihood methods (Qin and Zhang, 1997; Fokianos et al., 2001; Fokianos and Troendle, 2007; Jiang and Tu, 2012), giving fitted distributions $\hat{f}_X^{c+}, \hat{f}_Y^{c+}, \hat{f}_X^{c-}, \hat{f}_Y^{c-}$. Plugging these into distributional definition of θ_c in Remark 2, θ_c can then be estimated as

$$\hat{\theta}_{n,c} = \int \hat{F}_X^{c+}(x) d\hat{F}_Y^{c+}(x) - \int \hat{F}_X^{c-}(x) d\hat{F}_Y^{c-}(x).$$

This framework retains flexibility while introducing a parametric component that can improve efficiency, thus serving as a natural hybrid between distribution-based and model-based approaches. However, it applies only to numerical responses. A full theoretical development is beyond the scope of the present paper and will be pursued in future work.

Finally, when a reasonable guess about the potential cutpoint is available, we can develop a hybrid test statistic to improve power. Recall the definition of $\hat{\theta}_{n,c}$ in (2.6). Define

$$\hat{S}_{n,hybrid} := \sqrt{n} \left(\sup_{c \in [\ell, u]} (1 - \epsilon) |\hat{\theta}_{n,c}| + \epsilon |\hat{\theta}_{n,c_g}| \right),$$

where c_g is a user-supplied value and $\epsilon \in [0, 1]$ determines how much weight is given to the user-specified component. When $\epsilon = 0$, the statistic reduces to $\hat{S}_n$ for the unknown cutpoint problem in (1.3), while larger values of ϵ allow the test to focus more on θ_{c_g}. This hybridization can improve power when c_g is close to the true optimal cutpoint if it exists.

The bootstrap procedure described in Section 4.1 can be directly adapted for this hybrid test. Specifically, for each bootstrap sample, we compute the bootstrap version of the hybrid test statistic as

$$\hat{S}^*_{n,hybrid} = \sqrt{n} \left(\max_{c \in [\ell, u]} (1 - \epsilon) |\hat{\theta}^*_{n,c} - \hat{\theta}_{n,c}| + \epsilon |\hat{\theta}^*_{n,c_g} - \hat{\theta}_{n,c_g}| \right).$$

The empirical distribution of $\hat{S}^*_{n,hybrid}$ over all bootstrap replications provides an estimate of the sampling distribution of the hybrid statistic under the null. The p-value can be computed as $1 - F^*_{n,hybrid}(\hat{S}_{n,hybrid})$, where $F^*_{n,hybrid}$ denotes the distribution function of the boostrap test statistic $\hat{S}^*_{n,hybrid}$.

Table S7 summarizes the empirical size and power of the hybrid test under several configurations, which are described in Table 1 of Section 6.1. An additional Case 10 is included, in which $X \sim \mathcal{E}(0.7 + 0.2 \times I(Z > 0.5))$, and $Y \sim \mathcal{E}(0.3 - 0.2 \times I(Z > 0.5))$, with $c_b = 0.5$ and $|\theta_{c_b}| = 0.2$. Cases 1 and 3 are under null hypothesis, while Cases 6a and 10 are under alternative hypothesis.

Table S7: The empirical size and power (in percentage) of the hybrid test under the null and alternative hypotheses at a significance level of $\alpha = 0.05$. Configurations 1 and 3 correspond to empirical size, while Configurations 6a and 10 correspond to empirical power. Bootstrap replications: B= 1000; simulation rounds: 1000.

$c_g = 0.3$					$c_g = 0.5$					$c_g = 0.6$				
ϵ	1	3	6a	10	ϵ	1	3	6a	10	ϵ	1	3	6a	10
0	5.1	5.2	20.5	85.6	0	5.1	5.2	20.5	85.6	0	5.1	5.2	20.5	85.6
0.1	6.9	5.9	22.2	84.1	0.1	5.0	6.3	20.6	87.8	0.1	5.7	5.0	20.9	87.8
0.3	4.7	3.4	20.6	85.0	0.3	5.7	4.2	29.1	94.4	0.3	4.1	4.1	22.8	92.5
0.5	3.4	4.7	19.7	77.8	0.5	5.0	5.0	33.4	97.8	0.5	5.3	3.8	24.0	94.4
0.7	6.8	5.9	21.6	74.4	0.7	6.3	5.9	26.3	96.9	0.7	4.6	4.1	24.4	93.8
0.9	5.6	5.0	20.3	69.4	0.9	6.4	5.0	25.3	96.9	0.9	7.0	5.3	24.7	94.4

Each block corresponds to a different guessed cutpoint c_g, and we report results for various values of ϵ. The results indicate that the hybrid procedure maintains good size control while offering improved power for moderate values of ϵ under alternative hypothesis, particularly when the guessed c_g is close to the optimal cutpoint (i.e., 0.5). A systematic method for selecting $\epsilon \in (0, 1)$ is left for future research.

S5 Notations

We first establish certain notations for clarity and coherence. For an arbitrary index set S, let $\ell^\infty(S)$ denote the space of uniformly bounded, real-valued functions $f : S \to \mathbb{R}$ equipped with the sup norm $\|f\|_\infty = \sup_{s\in S} |f(s)|$. For a pseudometric space (S, d), $N(S, d, \epsilon)$ refers to the ϵ-covering number of (S, d), i.e., the minimum number of closed d-balls with radius at most ϵ that cover S. For a probability space $(S, \mathcal{S}, P)$ and a measurable function $f : S \to \mathbb{R}$, we use Pf to denote the integral of the function f with respect to the probability measure P, that is, $Pf = \int f dP$. For $q \in [1, \infty]$, we use the notation $\|\cdot\|_{P,q} := (P|f|^q)^{1/q} := (\int |f|^q dP)^{1/q}$ to denote the $L_q(P)$-norm. We write $X_n = O_P(1)$ (resp. $o_P(1)$) if the sequence of random variables X_n is bounded in probability (resp. converges to 0 in probability). For two sequences of positive *deterministic* numbers $\{a_n : n \geq 1\}$ and $\{b_n : n \geq 1\}$, we write $a_n = O(b_n)$ (resp. $a_n = o(b_n)$) if $\limsup_{n\to\infty} a_n/b_n < \infty$ (resp. $\lim_{n\to\infty} a_n/b_n = 0$).

Let $\mathcal{H}$ be a class of symmetric measurable functions $h : S^r \to \mathbb{R}$, and define the associated U-process of order r as follows:

$$U_n^{(r)}(h) = \frac{1}{\binom{n}{r}} \sum_{(i_1,...,i_r)\in I_{n,r}} h\left(X_{i_1}, ..., X_{i_r}\right), \quad h \in \mathcal{H}, \tag{S5.1}$$

where $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ denotes the number of r-combinations, and $I_{n,r} =$

$\{(i_1, ..., i_r) : 1 \leq i_1 < i_2 < ... < i_r \leq n\}$. For each $k = 1, ..., r$, let $P^{r-k}h$ denote the function on S^k defined by $P^{r-k}h(x_1, ...x_k) = E\left[h(x_1, ..., x_k, X_{k+1}, ..., X_r)\right]$. For a distribution Q on S^r, define $\|Q\|_{\mathcal{H}} = \sup_{h\in\mathcal{H}} |Qh|$.

Define

$$\mathcal{S} := \{(t, u, z) \in \mathbb{R} \times \{0, 1\} \times \mathbb{R}\}. \tag{S5.2}$$

Recall that $\boldsymbol{\mathcal{D}}_i = (T_i, U_i, Z_i), i \in [n]$ are independently and identically distributed observations taking values in $\mathcal{S}$. For $\ell \leq c \leq u$, we define

$$W_{c,n}^{(1)} = \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} w_c^{(1)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j), \quad W_{c,n}^{(2)} = \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} w_c^{(2)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j),$$
$$M_{c,n}^{(1)} = \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} m_c^{(1)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j), \quad M_{c,n}^{(2)} = \frac{2}{n(n-1)} \sum_{1\leq i<j\leq n} m_c^{(2)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j), \tag{S5.3}$$

which are U statistics with the symmetric kernels $w_c^{(1)}(\cdot,\cdot)$, $w_c^{(2)}(\cdot,\cdot)$, $m_c^{(1)}(\cdot,\cdot)$ and $m_c^{(2)}(\cdot,\cdot)$ defined respectively as

$$\begin{aligned}
w_c^{(1)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j) &:= T_{i,j}(1-U_i)U_j Z_i^{c+} Z_j^{c+} + T_{j,i}(1-U_j)U_i Z_i^{c+} Z_j^{c+}, \\
w_c^{(2)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j) &:= T_{i,j}(1-U_i)U_j Z_i^{c-} Z_j^{c-} + T_{j,i}(1-U_j)U_i Z_i^{c-} Z_j^{c-}, \\
m_c^{(1)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j) &:= (1-U_i)U_j Z_i^{c+} Z_j^{c+} + (1-U_j)U_i Z_i^{c+} Z_j^{c+}, \\
m_c^{(2)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j) &:= (1-U_i)U_j Z_i^{c-} Z_j^{c-} + (1-U_j)U_i Z_i^{c-} Z_j^{c-}.
\end{aligned} \tag{S5.4}$$

We define the following functions: for $1 \leq i \neq j \leq n$ and $c \in [\ell, u]$,

$$
\begin{aligned}
\bar{w}_c^{(1)}(\boldsymbol{\mathcal{D}}_i) &:= E\left[w_c^{(1)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j)|\boldsymbol{\mathcal{D}}_i\right] \\
&= \left[(1 - G^{(+)}(T_{i-}|c, 1))(1 - U_i)\lambda + G^{(+)}(T_i|c, 0)U_i(1 - \lambda)\right] Z_i^{c+}(1 - F(c)), \\
\bar{w}_c^{(2)}(\boldsymbol{\mathcal{D}}_i) &:= E\left[w_c^{(2)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j)|\boldsymbol{\mathcal{D}}_i\right] \\
&= \left[(1 - G^{(-)}(T_{i-}|c, 1))(1 - U_i)\lambda + G^{(-)}(T_i|c, 0)U_i(1 - \lambda)\right] Z_i^{c-}F(c), \\
\bar{m}_c^{(1)}(\boldsymbol{\mathcal{D}}_i) &:= E\left[m_c^{(1)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j)|\boldsymbol{\mathcal{D}}_i\right] = \left[(1 - U_i)\lambda + U_i(1 - \lambda)\right] Z_i^{c+}(1 - F(c)), \\
\bar{m}_c^{(2)}(\boldsymbol{\mathcal{D}}_i) &:= E\left[m_c^{(2)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j)|\boldsymbol{\mathcal{D}}_i\right] = \left[(1 - U_i)\lambda + U_i(1 - \lambda)\right] Z_i^{c-}F(c),
\end{aligned} \tag{S5.5}
$$

In the above, we recall that $G^{(+)}(t|c, k) = \Pr(T \leq t|Z > c, U = k)$, $G^{(-)}(t|c, k) = \Pr(T \leq t|Z \leq c, U = k), k = 0, 1$ is defined in (2.5).

Further, note that for $1 \leq i \neq j \leq n$, and $c \in [\ell, u]$,

$$
\begin{aligned}
&E\left[w_c^{(1)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j)\right] = 2\lambda(1 - \lambda)\Pr(X_1 \leq Y_2|Z_1 > c, Z_2 > c)(1 - F(c))^2 = E\left[W_{c,n}^{(1)}\right], \\
&E\left[w_c^{(2)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j)\right] = 2\lambda(1 - \lambda)\Pr(X_1 \leq Y_2|Z_1 \leq c, Z_2 \leq c)F^2(c) = E\left[W_{c,n}^{(2)}\right], \\
&E\left[m_c^{(1)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j)\right] = 2\lambda(1 - \lambda)(1 - F(c))^2 = E\left[M_{c_0,n}^{(1)}\right], \\
&E\left[m_c^{(2)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j)\right] = 2\lambda(1 - \lambda)F^2(c) = E\left[M_{c_0,n}^{(2)}\right].
\end{aligned} \tag{S5.6}
$$

Recall the definition of $\mathcal{S}$ in (S5.2). Define the following constant function on $\mathcal{S}^2$: $F^{(1)}(\boldsymbol{d}_1, \boldsymbol{d}_2) = 1$ for $\boldsymbol{d}_1, \boldsymbol{d}_2 \in \mathcal{S}$. Consider the function classes

$\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$ and $\mathcal{F}_4$ on $\mathcal{S}^2$ defined as follows:

$$\begin{aligned}
\mathcal{F}_1 &= \{\mathcal{S}^2 \ni (\boldsymbol{d}_1, \boldsymbol{d}_2) \mapsto t_{1,2}(1-u_1)u_2 z_1^{c+} z_2^{c+} + t_{2,1}(1-u_2)u_1 z_1^{c+} z_2^{c+} : c \in [\ell, u]\},\\
\mathcal{F}_2 &= \{\mathcal{S}^2 \ni (\boldsymbol{d}_1, \boldsymbol{d}_2) \mapsto t_{1,2}(1-u_1)u_2 z_1^{c-} z_2^{c-} + t_{2,1}(1-u_2)u_1 z_1^{c-} z_2^{c-} : c \in [\ell, u]\},\\
\mathcal{F}_3 &= \{\mathcal{S}^2 \ni (\boldsymbol{d}_1, \boldsymbol{d}_2) \mapsto (1-u_1)u_2 z_1^{c+} z_2^{c+} + (1-u_2)u_1 z_1^{c+} z_2^{c+} : c \in [\ell, u]\},\\
\mathcal{F}_4 &= \{\mathcal{S}^2 \ni (\boldsymbol{d}_1, \boldsymbol{d}_2) \mapsto (1-u_1)u_2 z_1^{c-} z_2^{c-} + (1-u_2)u_1 z_1^{c-} z_2^{c-} : c \in [\ell, u]\},
\end{aligned} \tag{S5.7}$$

where $d_i = (t_i, u_i, z_i)$ and $t_{i,j} = I(t_i \leq t_j)$ for $i, j \in \{1, 2\}$. It is clear that $F^{(1)}(\cdot)$ is an envelope function for these function classes. Further, due to Example 2.6.1 in Van der Vaart and Wellner (1996) and the permanence property (Van der Vaart and Wellner, 1996, Lemma 2.6.18), $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$ and $\mathcal{F}_4$ are VC-subgraph classes (Van der Vaart and Wellner, 1996, Section 2.6.2).

Note that $\{W_{c,n}^{(1)} : c \in [\ell, u]\}$, $\{W_{c,n}^{(2)} : c \in [\ell, u]\}$ $\{M_{c,n}^{(1)} : c \in [\ell, u]\}$ and $\{M_{c,n}^{(2)} : c \in [\ell, u]\}$ are respectively the U-processes indexed by $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$ and $\mathcal{F}_4$.

S6 Proofs in the Case with prespecified cutpoint

Proof of Theorem 1. Recall that F is the distribution function of Z with $0 < F(\ell) < F(u) < 1$ and $c_0 \in [\ell, u]$, and that U and Z are independent.

We note the following decomposition:

$$\sqrt{n}\left(\hat{\theta}_{n,c_0}-\theta_{c_0}\right)=\sqrt{n}\left(\frac{W^{(1)}_{c_0,n}}{M^{(1)}_{c_0,n}}-\frac{W^{(2)}_{c_0,n}}{M^{(2)}_{c_0,n}}-\left(\frac{E\left[W^{(1)}_{c_0,n}\right]}{E\left[M^{(1)}_{c_0,n}\right]}-\frac{E\left[W^{(2)}_{c_0,n}\right]}{E\left[M^{(2)}_{c_0,n}\right]}\right)\right),$$

where $W^{(1)}_{c_0,n}$, $W^{(2)}_{c_0,n}$, $M^{(1)}_{c_0,n}$, $M^{(2)}_{c_0,n}$ and their kernels $w^{(1)}_{c_0}(\cdot,\cdot)$, $w^{(2)}_{c_0}(\cdot,\cdot)$, $m^{(1)}_{c_0}(\cdot,\cdot)$, $m^{(2)}_{c_0}(\cdot,\cdot)$ are defined in (S5.3) and (S5.4) with the substitution of c by c_0. Further, the expectations above are given in (S5.6).

As the kernels $w^{(1)}_{c_0}(\cdot,\cdot)$, $w^{(2)}_{c_0}(\cdot,\cdot)$, $m^{(1)}_{c_0}(\cdot,\cdot)$, $m^{(2)}_{c_0}(\cdot,\cdot)$ are all bounded, by the central limit theorem for U-statistics (Van der Vaart, 2007, Theorem 12.3), we have

$$\sqrt{n}\begin{pmatrix} W^{(1)}_{c_0,n}-E\left[W^{(1)}_{c_0,n}\right] \\ W^{(2)}_{c_0,n}-E\left[W^{(2)}_{c_0,n}\right] \\ M^{(1)}_{c_0,n}-E\left[M^{(1)}_{c_0,n}\right] \\ M^{(2)}_{c_0,n}-E\left[M^{(1)}_{c_0,n}\right]\end{pmatrix}=\frac{2}{\sqrt{n}}\sum_{i=1}^{n}\begin{pmatrix}\bar{w}^{(1)}_{c_0}(\boldsymbol{\mathcal{D}}_i)-E\left[\bar{w}^{(1)}_{c_0}(\boldsymbol{\mathcal{D}}_i)\right] \\ \bar{w}^{(2)}_{c_0}(\boldsymbol{\mathcal{D}}_i)-E\left[\bar{w}^{(2)}_{c_0}(\boldsymbol{\mathcal{D}}_i)\right] \\ \bar{m}^{(1)}_{c_0}(\boldsymbol{\mathcal{D}}_i)-E\left[\bar{m}^{(1)}_{c_0}(\boldsymbol{\mathcal{D}}_i)\right] \\ \bar{m}^{(2)}_{c_0}(\boldsymbol{\mathcal{D}}_i)-E\left[\bar{m}^{(2)}_{c_0}(\boldsymbol{\mathcal{D}}_i)\right]\end{pmatrix}+o_P(1)\rightsquigarrow\mathcal{N}(\mathbf{0},4\boldsymbol{\Sigma}),$$

where $\bar{w}^{(1)}_{c_0}(\cdot),\bar{w}^{(2)}_{c_0}(\cdot),\bar{m}^{(1)}_{c_0}(\cdot),\bar{m}^{(2)}_{c_0}(\cdot)$ are defined in (S5.5), $\mathbf{0}=(0,0,0,0)^\top$, and

$$\boldsymbol{\Sigma}=\operatorname{cov}\left(\bar{w}^{(1)}_{c_0}(\boldsymbol{\mathcal{D}}_i),\bar{w}^{(2)}_{c_0}(\boldsymbol{\mathcal{D}}_i),\bar{m}^{(1)}_{c_0}(\boldsymbol{\mathcal{D}}_i),\bar{m}^{(2)}_{c_0}(\boldsymbol{\mathcal{D}}_i)\right) \tag{S6.1}$$

is the 4×4 covariance matrix. Due to condition (3.7), Σ has a full rank.

Consider the function $g(x,y,z,\gamma)=\frac{x}{z}-\frac{y}{\gamma}$ for $x,y,z,\gamma\in\mathbb{R}$ and $z,\gamma\neq 0$. By assumptions, the map g is differentiable at the point $\beta=(E[W^{(1)}_{c_0,n}],E[W^{(2)}_{c_0,n}],E[M^{(1)}_{c_0,n}],E[M^{(2)}_{c_0,n}])$, with derivative

$$\boldsymbol{g}_\beta' = \left(\frac{1}{E[M_{c_0,n}^{(1)}]}, -\frac{1}{E[M_{c_0,n}^{(2)}]}, -\frac{E[W_{c_0,n}^{(1)}]}{\left(E[M_{c_0,n}^{(1)}]\right)^2}, \frac{E[W_{c_0,n}^{(2)}]}{\left(E[M_{c_0,n}^{(2)}]\right)^2} \right). \tag{S6.2}$$

Applying the Delta method (Van der Vaart, 2007, Theorem 3.1), we then obtain

$$\sqrt{n}\left(\hat{\theta}_{n,c_0} - \theta_{c_0}\right) = \frac{2}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{g}_\beta' \begin{pmatrix} \bar{w}_{c_0}^{(1)}(\boldsymbol{\mathcal{D}}_i) - E\left[\bar{w}_{c_0}^{(1)}(\boldsymbol{\mathcal{D}}_i)\right] \\ \bar{w}_{c_0}^{(2)}(\boldsymbol{\mathcal{D}}_i) - E\left[\bar{w}_{c_0}^{(2)}(\boldsymbol{\mathcal{D}}_i)\right] \\ \bar{m}_{c_0}^{(1)}(\boldsymbol{\mathcal{D}}_i) - E\left[\bar{m}_{c_0}^{(1)}(\boldsymbol{\mathcal{D}}_i)\right] \\ \bar{m}_{c_0}^{(2)}(\boldsymbol{\mathcal{D}}_i) - E\left[\bar{m}_{c_0}^{(2)}(\boldsymbol{\mathcal{D}}_i)\right] \end{pmatrix} + o_P(1) \rightsquigarrow \mathcal{N}(0, \sigma_{c_0}^2),$$

where

$$\sigma_{c_0}^2 = 4\boldsymbol{g}_\beta' \boldsymbol{\Sigma} \boldsymbol{g}_\beta'^{\top}. \tag{S6.3}$$

Since Σ has a full rank and $\boldsymbol{g}_\beta' \neq \boldsymbol{0}$, we have $\sigma_{c_0}^2$ is positive. □

Proof of Theorem 2. We first define the jackknife pseudo-values for $1 \leq k \leq n$ and their average as follows:

$$\tilde{\theta}_{n,k,T_k} := (n-1)\hat{\theta}_{n,c_0} - (n-1)\hat{\theta}_{n,c_0}^{-k}, \qquad \bar{\theta}_n := \frac{1}{n}\sum_{k=1}^{n} \tilde{\theta}_{n,k,T_k}, \tag{S6.4}$$

where $\hat{\theta}_{n,c_0}^{-k}$ is the estimate of θ_c based on the sample with the k-th observation left out. By elementary calculation, we have that $\hat{\sigma}_{n,c_0}^2$ can be computed using jackknife pseudo-values:

$$\hat{\sigma}_{n,c_0}^2 = \frac{1}{n-1}\sum_{k=1}^{n}(\tilde{\theta}_{n,k,T_k} - \bar{\theta}_n)^2 = \frac{1}{n-1}\sum_{k=1}^{n}\left(\tilde{\theta}_{n,k,T_k} - \theta_{c_0}\right)^2 - \frac{n}{n-1}\left(\bar{\theta}_n - \theta_{c_0}\right)^2. \tag{S6.5}$$

For each $x \in \mathbb{R}$ and $1 \leq k \leq n$, define

$$\tilde{\theta}_{n,k,x} := U_k Z_k^{c_0+} \tilde{\theta}_{n,k,x}^{(1)} + (1-U_k) Z_k^{c_0+} \tilde{\theta}_{n,k,x}^{(2)} + U_k Z_k^{c_0-} \tilde{\theta}_{n,k,x}^{(3)} + (1-U_k) Z_k^{c_0-} \tilde{\theta}_{n,k,x}^{(4)}, \tag{S6.6}$$

where, $\tilde{\theta}_{n,k,x}^{(j)}$ for $j = 1, 2, 3, 4$ are defined as follows:

$$\begin{aligned}
\tilde{\theta}_{n,k,x}^{(1)} :=& \frac{(n-1) \sum_{i \neq k} I(T_i \leq x) (1-U_i) Z_i^{c_0+}}{\left[\sum_{i \neq k} (1-U_i) Z_i^{c_0+}\right] \left[\sum_{i \neq k} U_i Z_i^{c_0+} + 1\right]} \\
&- \frac{(n-1) \sum_{i \neq k} \sum_{j \neq k} T_{i,j} (1-U_i) U_j Z_i^{c_0+} Z_j^{c_0+}}{\left[\sum_{i \neq k} (1-U_i) Z_i^{c_0+}\right] \left[\sum_{i \neq k} U_i Z_i^{c_0+} + 1\right] \left[\sum_{i \neq k} U_i Z_i^{c_0+}\right]}, \\
\tilde{\theta}_{n,k,x}^{(2)} :=& \frac{(n-1) \sum_{i \neq k} I(x \leq T_i) U_i Z_i^{c_0+}}{\left[\sum_{i \neq k} (1-U_i) Z_i^{c_0+} + 1\right] \left[\sum_{i \neq k} U_i Z_i^{c_0+}\right]} \\
&- \frac{(n-1) \sum_{i \neq k} \sum_{j \neq k} T_{i,j} (1-U_i) U_j Z_i^{c_0+} Z_j^{c_0+}}{\left[\sum_{i \neq k} (1-U_i) Z_i^{c_0+} + 1\right] \left[\sum_{i \neq k} U_i Z_i^{c_0+}\right] \left[\sum_{i \neq k} (1-U_i) Z_i^{c_0+}\right]}, \\
\tilde{\theta}_{n,k,x}^{(3)} :=& -\frac{(n-1) \sum_{i \neq k} I(T_i \leq x) (1-U_i) Z_i^{c_0-}}{\left[\sum_{i \neq k} (1-U_i) Z_i^{c_0-}\right] \left[\sum_{i \neq k} U_i Z_i^{c_0-} + 1\right]} \\
&+ \frac{(n-1) \sum_{i \neq k} \sum_{j \neq k} T_{i,j} (1-U_i) U_j Z_i^{c_0-} Z_j^{c_0-}}{\left[\sum_{i \neq k} (1-U_i) Z_i^{c_0-}\right] \left[\sum_{i \neq k} U_i Z_i^{c_0-} + 1\right] \left[\sum_{i \neq k} U_i Z_i^{c_0-}\right]}, \\
\tilde{\theta}_{n,k,x}^{(4)} :=& -\frac{(n-1) \sum_{i \neq k} I(x \leq T_i) U_i Z_i^{c_0-}}{\left[\sum_{i \neq k} (1-U_i) Z_i^{c_0-} + 1\right] \left[\sum_{i \neq k} U_i Z_i^{c_0-}\right]} \\
&+ \frac{(n-1) \sum_{i \neq k} \sum_{j \neq k} T_{i,j} (1-U_i) U_j Z_i^{c_0-} Z_j^{c_0-}}{\left[\sum_{i \neq k} (1-U_i) Z_i^{c_0-} + 1\right] \left[\sum_{i \neq k} U_i Z_i^{c_0-}\right] \left[\sum_{i \neq k} (1-U_i) Z_i^{c_0-}\right]},
\end{aligned}$$

where $\sum_{i \neq k}$ represents $\sum_{i=1, i \neq k}^{n}$. Then by definition, the k-th pseudo-value $\tilde{\theta}_{n,k,T_k}$ is equal to $\tilde{\theta}_{n,k,x}$ with x substituted by T_k.

Further, for each $x \in \mathbb{R}$ and $1 \leq k \leq n$, define

$$\tilde{\Psi}_{k,x} = U_k Z_k^{c_0+} \Psi_x^{(1)} + (1-U_k) Z_k^{c_0+} \Psi_x^{(2)} + U_k Z_k^{c_0-} \Psi_x^{(3)} + (1-U_k) Z_k^{c_0-} \Psi_x^{(4)}, \tag{S6.7}$$

where

$$\begin{aligned}
\Psi_x^{(1)} &= \frac{E\left[I(T_i \leq x)\,(1-U_i)\,Z_i^{c_0+}\right]}{\lambda(1-\lambda)(1-F(c_0))^2} - \frac{\Pr(X_1 \leq Y_2 | Z_1 > c_0, Z_2 > c_0)}{\lambda(1-F(c_0))}, \\
\Psi_x^{(2)} &= \frac{E\left[I(x \leq T_i) U_i Z_i^{c_0+}\right]}{\lambda(1-\lambda)(1-F(c_0))^2} - \frac{\Pr(X_1 \leq Y_2 | Z_1 > c_0, Z_2 > c_0)}{(1-\lambda)(1-F(c_0))}, \\
\Psi_x^{(3)} &= -\frac{E\left[I(T_i \leq x)\,(1-U_i)\,Z_i^{c_0-}\right]}{\lambda(1-\lambda)F^2(c_0)} + \frac{\Pr(X_1 \leq Y_2 | Z_1 \leq c_0, Z_2 \leq c_0)}{\lambda F(c_0)}, \\
\Psi_x^{(4)} &= -\frac{E\left[I(x \leq T_i) U_i Z_i^{c_0-}\right]}{\lambda(1-\lambda)F^2(c_0)} + \frac{\Pr(X_1 \leq Y_2 | Z_1 \leq c_0, Z_2 \leq c_0)}{(1-\lambda)F(c_0)},
\end{aligned}$$

In Lemma 1, we show that $\tilde{\theta}_{n,k,x}$ is approximated by $\tilde{\Psi}_{k,x}$ uniformly over $1 \leq k \leq n$ and $x \in \mathbb{R}^n$ as $n \to \infty$, which implies that if we define

$$R_{n,k} := \tilde{\theta}_{n,k,T_k} - \tilde{\Psi}_{k,T_k},$$

then $M_n := \max_{1 \leq k \leq n} |R_{n,k}| = o_P(1)$. Further, due to (S6.5) and by definition,

$$\begin{aligned}
\hat{\sigma}^2_{n,c_0} =& \frac{1}{n-1} \sum_{k=1}^n \left(\tilde{\Psi}_{k,T_k} - \theta_{c_0}\right)^2 + \frac{2}{n-1} \sum_{k=1}^n \left(\tilde{\Psi}_{k,T_k} - \theta_{c_0}\right) R_{n,k} + \frac{1}{n-1} \sum_{k=1}^n R_{n,k}^2 \\
&+ \frac{n}{n-1} \left(\frac{1}{n} \sum_{k=1}^n \tilde{\Psi}_{k,T_k} - \theta_{c_0} + \frac{1}{n} \sum_{k=1}^n R_{n,k}\right)^2.
\end{aligned} \tag{S6.8}$$

Note that

$$\frac{1}{n}\sum_{k=1}^{n}|R_{n,k}| \leq M_n = o_P(1), \qquad \frac{1}{n-1}\sum_{k=1}^{n}R_{n,k}^2 \leq \frac{n}{n-1}M_n^2 = o_P(1)$$
$$\left|\frac{1}{n-1}\sum_{k=1}^{n}\left(\tilde{\Psi}_{k,T_k}-\theta_{c_0}\right)R_{n,k}\right| \leq \left(\frac{1}{n-1}\sum_{k=1}^{n}\left|\tilde{\Psi}_{k,T_k}-\theta_{c_0}\right|\right)M_n = o_P(1),$$

where the last equality is due to the law of large numbers, since $\{\tilde{\Psi}_{k,T_k} : k \in [n]\}$ are i.i.d.

Finally, we note that for $1 \leq k \leq n$, $E\left[\tilde{\Psi}_{k,T_k}\right] = \theta_{c_0}$, and that by definition,

$$\tilde{\Psi}_{k,T_k} = 2\boldsymbol{g}'_\beta\left(\bar{w}_{c_0}^{(1)}(\boldsymbol{\mathcal{D}}_k), \bar{w}_{c_0}^{(2)}(\boldsymbol{\mathcal{D}}_k), \bar{m}_{c_0}^{(1)}(\boldsymbol{\mathcal{D}}_k), \bar{m}_{c_0}^{(2)}(\boldsymbol{\mathcal{D}}_k)\right)^\top,$$

where we recall $\boldsymbol{g}'_\beta$ in (S6.2) and $\bar{w}_{c_0}^{(1)}(\cdot), \bar{w}_{c_0}^{(2)}(\cdot), \bar{m}_{c_0}^{(1)}(\cdot), \bar{m}_{c_0}^{(2)}(\cdot)$ are defined in (S5.5),

As a result, due to the definition of Σ in (S6.1) and $\sigma_{c_0}^2$ in (S6.3), we have

$$\mathrm{Var}(\tilde{\Psi}_{k,T_k}) = 4\boldsymbol{g}'_\beta\boldsymbol{\Sigma}\boldsymbol{g}'^{\top}_\beta = \sigma_{c_0}^2.$$

Then the proof is complete by the law of large numbers and the central limit theorem. □

Recall the definition of $\tilde{\theta}_{n,k,x}$ and $\tilde{\Psi}_{k,x}$ in (S6.6) and (S6.7) respectively.

Lemma 1. *Suppose that condition (C.0) holds. Then we have*

$$\max_{1\leq k\leq n}\sup_{x\in\mathbb{R}}\left|\tilde{\theta}_{n,k,x} - \tilde{\Psi}_{k,x}\right| = o_P(1).$$

Proof. By the definition of $\tilde{\theta}_{n,k,x}$ and $\tilde{\Psi}_{k,x}$ in (S6.6) and (S6.7) respectively, it suffices to show that for $j = 1, \ldots, 4$,

$$\max_{1\le k\le n} \sup_{x\in\mathbb{R}} \left|\tilde{\theta}^{(j)}_{n,k,x} - \Psi^{(j)}_x\right| = o_P(1).$$

By the union bound, it suffices to show that for any $\epsilon > 0$ and for $j = 1, \ldots, 4$,

$$\lim_{n\to\infty} n \Pr\left(\sup_{x\in\mathbb{R}} \left|\tilde{\theta}^{(j)}_{n,1,x} - \tilde{\Psi}^{(j)}_x\right| > \epsilon\right) = 0. \tag{S6.9}$$

We only present details for the case $j = 1$, noting that similar arguments apply to $j = 2, 3, 4$.

Due to monotonicity, it suffices to show that (S6.9) holds for all sufficiently small $\epsilon > 0$. Now, we fix some $\epsilon > 0$ such that

$$\epsilon \le \min\Big\{(1-\lambda)(1-F(c_0)),\ \lambda(1-F(c_0)),\ (1-\lambda)F(c_0),\ \lambda F(c_0)\Big\}. \tag{S6.10}$$

Define the following event

$$\begin{aligned} B_n := \Bigg\{ &\left|\frac{1}{n-1}\sum_{i\neq 1} (1-U_i)\, Z_i^{c_0+} - (1-\lambda)(1-F(c_0))\right| \le \epsilon/2, \\ &\left|\frac{1}{n-1}\sum_{i\neq 1} U_i Z_i^{c_0+} - \lambda(1-F(c_0))\right| \le \epsilon/2, \\ &\left|\frac{1}{n-1}\sum_{i\neq 1} (1-U_i)\, Z_i^{c_0-} - (1-\lambda)F(c_0)\right| \le \epsilon/2, \\ &\left|\frac{1}{n-1}\sum_{i\neq 1} U_i Z_i^{c_0-} - \lambda F(c_0)\right| \le \epsilon/2 \Bigg\}. \end{aligned}$$

Due to (S6.10), on the event B_n, we have

$$\min\left\{\frac{1}{n-1}\sum_{i\neq 1}(1-U_i)\,Z_i^{c_0+},\ \ \frac{1}{n-1}\sum_{i\neq 1}U_iZ_i^{c_0+}\right\}\geq \epsilon/2$$
$$\min\left\{\frac{1}{n-1}\sum_{i\neq 1}(1-U_i)\,Z_i^{c_0-},\ \ \frac{1}{n-1}\sum_{i\neq 1}U_iZ_i^{c_0-}\right\}\geq \epsilon/2.$$

Further, we define the following event:

$$A_{n,1}:=\left\{\sup_{x\in\mathbb{R}}\left|\frac{1}{n-1}\sum_{i\neq 1}I(T_i\leq x)\,(1-U_i)\,Z_i^{c_0+}-E\left[I(T_i\leq x)\,(1-U_i)\,Z_i^{c_0+}\right]\right|\leq\frac{\epsilon^3}{16}\right\},$$
$$A_{n,2}:=\left\{\left|\frac{1}{(n-1)^2}\left[\sum_{i\neq 1}(1-U_i)\,Z_i^{c_0+}\right]\left[\sum_{i\neq 1}U_iZ_i^{c_0+}+1\right]-\lambda(1-\lambda)(1-F(c_0))^2\right|\leq\frac{\epsilon^5}{16}\right\},$$
$$A_{n,3}:=\left\{\left|\frac{2}{(n-1)(n-2)}\sum_{2\leq i<j\leq n}w_{c_0,n}^{(1)}(\boldsymbol{\mathcal{D}}_i,\boldsymbol{\mathcal{D}}_j)-E\left[W_{c_0,n}^{(1)}\right]\right|\leq\frac{\epsilon^4}{32}\right\},$$
$$A_{n,4}:=\left\{\left|\frac{1}{(n-1)^3}\left[\sum_{i\neq 1}(1-U_i)\,Z_i^{c_0+}\right]\left[\sum_{i\neq 1}U_iZ_i^{c_0+}+1\right]\left[\sum_{i\neq 1}U_iZ_i^{c_0+}\right]-\lambda^2(1-\lambda)(1-F(c_0))^3\right|\leq\frac{\epsilon^7}{32}\right\}$$

Since $E\left[I(T_1\leq x)\,(1-U_1)\,Z_1^{c_0+}\right]\leq 1$ and $E\left[W_{c_0,n}^{(1)}\right]\leq 1$, by definition, we have

$$B_n\ \cap\ A_{n,1}\ \cap\ A_{n,2}\ \cap\ A_{n,3}\ \cap\ A_{n,4}\ \ \subset\ \ \left\{\left|\tilde{\theta}_{n,1,x}^{(j)}-\tilde{\Psi}_x^{(j)}\right|\leq\epsilon.\right\}$$

By the union bound, we have

$$\Pr\left(\sup_{x\in\mathbb{R}}\left|\tilde{\theta}_{n,1,x}^{(j)}-\tilde{\Psi}_x^{(j)}\right|>\epsilon\right)\leq\Pr(B_n^{\complement})+\Pr(A_{n,1}^{\complement})+\Pr(A_{n,2}^{\complement})+\Pr(A_{n,3}^{\complement})+\Pr(A_{n,4}^{\complement}).$$

Since U_i, $T_{i,j}$ and $Z_{c_0}^+$ are all bounded by 1, by McDiarmid's inequality (McDiarmid, 1989), there exists some constant $C>0$ such that for $n\geq 1$

$$\Pr(B_n^{\complement})+\Pr(A_{n,2}^{\complement})+\Pr(A_{n,3}^{\complement})+\Pr(A_{n,4}^{\complement})\leq Ce^{-n\epsilon^{14}/C}. \qquad \text{(S6.11)}$$

Finally, we bound the probability of the event $A_{n,1}^{\complement}$. Define

$$Z_n := \sup_{x\in\mathbb{R}} \left| \frac{1}{n-1} \sum_{i\neq 1} I(T_i \leq x)\,(1-U_i)\,Z_i^{c_0+} - E\left[I(T_i \leq x)\,(1-U_i)\,Z_i^{c_0+}\right] \right|.$$

By the bounded differences inequality (Giné and Nickl, 2021, Theorem 3.3.14) (see in particular case (b) in (Giné and Nickl, 2021, Example 3.3.13)), there exists some constant $C > 0$ such that

$$\Pr(Z_n \geq E[Z_n] + t) \leq e^{-2nt^2}, \text{ for } t > 0.$$

Further, by the maximal inequality (Van der Vaart and Wellner, 1996, Corollary 2.2.8), for some constant $C > 0$,

$$E[Z_n] \leq C/\sqrt{n}.$$

Combining two parts, for some constant $C > 0$ and large enough n, we have

$$\Pr(A_{n,1}^{\complement}) = \Pr(Z_n \geq \epsilon^3/16) \leq Ce^{-n\epsilon^6/C}.$$

Thus, together with (S6.11), we prove (S6.9). The proof is complete. □

S7 Proofs regarding hypothesis testing with unknown cut-point

Proof of Theorem 3. Under H_0, $\theta_c = 0$ for all $c \in [\ell, u]$. Note that for any $c \in [\ell, u]$, $\sqrt{n}\hat{\theta}_{n,c}$ can be expressed as

$$\begin{aligned}\sqrt{n}\hat{\theta}_{n,c} &= \sqrt{n}\left(\frac{W_{c,n}^{(1)}}{M_{c,n}^{(1)}} - \frac{E\left[W_{c,n}^{(1)}\right]}{E\left[M_{c,n}^{(1)}\right]}\right) - \sqrt{n}\left(\frac{W_{c,n}^{(2)}}{M_{c,n}^{(2)}} - \frac{E\left[W_{c,n}^{(2)}\right]}{E\left[M_{c,n}^{(2)}\right]}\right) \\ &= \frac{\sqrt{n}\left(W_{c,n}^{(1)} - E\left[W_{c,n}^{(1)}\right]\right)}{M_{c,n}^{(1)}} + E\left[W_{c,n}^{(1)}\right]\sqrt{n}\left(\frac{1}{M_{c,n}^{(1)}} - \frac{1}{E\left[M_{c,n}^{(1)}\right]}\right) \\ &\quad - \frac{\sqrt{n}\left(W_{c,n}^{(2)} - E\left[W_{c,n}^{(2)}\right]\right)}{M_{c,n}^{(2)}} - E\left[W_{c,n}^{(2)}\right]\sqrt{n}\left(\frac{1}{M_{c,n}^{(2)}} - \frac{1}{E\left[M_{c,n}^{(2)}\right]}\right),\end{aligned} \tag{S7.1}$$

where recall that $W_{c,n}^{(1)}$, $W_{c,n}^{(2)}$, $M_{c,n}^{(1)}$ and $M_{c,n}^{(2)}$ are defined in (S5.3), and $\left\{W_{c,n}^{(1)} : c \in [\ell, u]\right\}$, $\left\{W_{c,n}^{(2)} : c \in [\ell, u]\right\}$, $\left\{M_{c,n}^{(1)} : c \in [\ell, u]\right\}$ and $\left\{M_{c,n}^{(2)} : c \in [\ell, u]\right\}$ are U processes indexed by $c \in [\ell, u]$.

Define two index sets T and T' as follows:

$$T = [\ell, u] \times \{1, \ldots, 4\}, \quad T' = [\ell, u] \times \{1, 2\}.$$

Since $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$ and $\mathcal{F}_4$ defined in (S5.7) are VC-subgraph classes (Van der Vaart and Wellner, 1996, Section 2.6.2), by the central limit theorem for U process (Peña and Giné, 1999, Theorem 5.3.3; Arcones and Gine, 1993, Theorem 4.1), there exists a tight centered Gaussian process $\{G_{1,c}, G_{2,c}, G_{3,c}, G_{4,c} :$

$c \in [\ell, u]\}$ such that

$$\begin{Bmatrix} \sqrt{n}\left(W_{c,n}^{(1)} - E\left[W_{c,n}^{(1)}\right]\right) \\ \sqrt{n}\left(M_{c,n}^{(1)} - E\left[M_{c,n}^{(1)}\right]\right) \\ \sqrt{n}\left(W_{c,n}^{(2)} - E\left[W_{c,n}^{(2)}\right]\right) \\ \sqrt{n}\left(M_{c,n}^{(2)} - E\left[M_{c,n}^{(2)}\right]\right) \end{Bmatrix} : c \in [\ell, u] \rightsquigarrow \begin{Bmatrix} G_{1,c} \\ G_{2,c} \\ G_{3,c} \\ G_{4,c} \end{Bmatrix} : c \in [\ell, u] \quad \text{in} \quad \ell^\infty(T).$$

Let $\mathbb{D}_\phi = D[\ell, u]$ be the space of cadlag functions on $[\ell, u]$ equipped with the sup norm. Now, define the pointwise inverse map $\phi(\cdot) : \mathbb{D}_\phi \mapsto \mathbb{D}_\phi$ as follows: for $\{D_c : c \in [\ell, u]\} \in \mathbb{D}_\phi$,

$$\phi\left(\{D_c : c \in [\ell, u]\}\right) = \begin{cases} \{0 : c \in [\ell, u]\}, & \text{if } \inf_{c \in [\ell, u]} |D_c| = 0 \\ \{{D_c}^{-1} : c \in [\ell, u]\}, & \text{otherwise} \end{cases} . \qquad \text{(S7.2)}$$

By Section 2.2.4 and Lemma 12.2 of Kosorok (2008), due to condition (*C.0*), $\phi(\cdot)$ is Hadamard-differentiable, tangentially to $\mathbb{D}_\phi$, at $\{E[M_{c,n}^{(1)}] : c \in [\ell, u]\}$ and $\{E[M_{c,n}^{(2)}] : c \in [\ell, u]\}$ with the derivative map $\phi'_{\{E[M_{c,n}^{(1)}]\}}(\{D_c\}) = \{-D_c/(E[M_{c,n}^{(1)}])^2\}$ and $\phi'_{\{E[M_{c,n}^{(2)}]\}}(\{D_c\}) = \{-D_c/(E[M_{c,n}^{(2)}])^2\}$ respectively for any $\{D_c\} \in \mathbb{D}_\phi$, where in the above $c \in [\ell, u]$. Then by the functional delta method (Kosorok, 2008, Theorem 2.8), we have

$$\left\{\begin{matrix} \sqrt{n}\left(W_{c,n}^{(1)} - E\left[W_{c,n}^{(1)}\right]\right) \\ \sqrt{n}\left({M_{c,n}^{(1)}}^{-1} - E\left[M_{c,n}^{(1)}\right]^{-1}\right) \\ \sqrt{n}\left(W_{c,n}^{(2)} - E\left[W_{c,n}^{(2)}\right]\right) \\ \sqrt{n}\left({M_{c,n}^{(2)}}^{-1} - E\left[M_{c,n}^{(2)}\right]^{-1}\right) \end{matrix} : c \in [\ell, u]\right\} \rightsquigarrow \left\{\begin{matrix} G_{1,c} \\ -G_{2,c}/(E[M_{c,n}^{(1)}])^2 \\ G_{3,c} \\ -G_{4,c}/(E[M_{c,n}^{(2)}])^2 \end{matrix} : c \in [\ell, u]\right\} \quad \text{in} \quad \ell^\infty(T).$$

Moreover, by applying the law of large numbers for U processes (Peña and Giné, 1999, Corollary 5.2.3) and the continuous mapping theorem (Van der Vaart and Wellner, 1996, Theorem 1.3.6), we can conclude that almost surely,

$$\left\{\begin{matrix} {M_{c,n}^{(1)}}^{-1} \\ {M_{c,n}^{(2)}}^{-1} \end{matrix} : c \in [\ell, u]\right\} \to \left\{\begin{matrix} E\left[M_{c,n}^{(1)}\right]^{-1} \\ E\left[M_{c,n}^{(2)}\right]^{-1} \end{matrix} : c \in [\ell, u]\right\} \quad \text{in} \quad \ell^\infty(T').$$

Then by Slutsky's theorem (Kosorok, 2008, Theorem 7.15) and continuous mapping (Van der Vaart and Wellner, 1996, Theorem 1.3.6) again, we obtain that under H_0, $\{\sqrt{n}\hat{\theta}_{n,c} : c \in [\ell, u]\}$ converges weakly to a tight centered Gaussian process $\{G_c : c \in [\ell, u]\}$ in $\ell^\infty([\ell, u])$, where

$$G_c := E\left[M_{c,n}^{(1)}\right]^{-1} G_{1,c} - \frac{E\left[W_{c,n}^{(1)}\right]}{E\left[M_{c,n}^{(1)}\right]^2} G_{2,c} - E\left[M_{c,n}^{(2)}\right]^{-1} G_{3,c} + \frac{E\left[W_{c,n}^{(2)}\right]}{E\left[M_{c,n}^{(2)}\right]^2} G_{4,c}. \tag{S7.3}$$

The last statement is again due to the continuous mapping theorem (Van der Vaart and Wellner, 1996, Theorem 1.3.6). The proof is complete. □

S7.1 Bootstrap consistency for hypothesis testing with unknown cut-point

Proof of Theorem 4. For each c, $\sqrt{n}(\hat{\theta}^*_{n,c} - \hat{\theta}_{n,c})$ can be written as:

$$\begin{aligned}\sqrt{n}(\hat{\theta}^*_{n,c} - \hat{\theta}_{n,c}) = & \frac{\sqrt{n}\left(W^{(1)^*}_{c,n} - W^{(1)}_{c,n}\right)}{M^{(1)^*}_{c,n}} + W^{(1)}_{c,n}\sqrt{n}\left(\frac{1}{M^{(1)^*}_{c,n}} - \frac{1}{M^{(1)}_{c,n}}\right) \\ & - \frac{\sqrt{n}\left(W^{(2)^*}_{c,n} - W^{(2)}_{c,n}\right)}{M^{(2)^*}_{c,n}} - W^{(2)}_{c,n}\sqrt{n}\left(\frac{1}{M^{(2)^*}_{c,n}} - \frac{1}{M^{(2)}_{c,n}}\right),\end{aligned}$$

where $W^{(1)^*}_{c,n}$, $W^{(2)^*}_{c,n}$, $M^{(1)^*}_{c,n}$ and $M^{(2)^*}_{c,n}$ are defined by replacing original data by the bootstrap data in (S5.3). We will apply the central limit theorem for bootstrapped U-processes (Arcones and Giné, 1994, Theorem 2.1). First, we verify the two required conditions.

Recall that the function classes $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$ and $\mathcal{F}_4$ defined in (S5.7) are VC-subgraph classes with the constant envelope function $F^{(1)}(\cdot,\cdot) = 1$ as mentioned in the proof of Theorem 3. Due to Example 2.6.1 and Theorem 2.6.7 in Van der Vaart and Wellner (1996), for $j = 1, \ldots, 4$,

$$\int_0^\infty \left(\sup_Q \log N\left(\mathcal{F}_j, \|\cdot\|_{Q,2}, \epsilon\|F^{(1)}\|_{Q,2}\right)\right)^{1/2} d\epsilon < \infty,$$

where the supremum is taken over all probability measures Q on the space $\mathcal{S}^2$. Thus condition (i) in (Arcones and Giné, 1994, Theorem 2.1) is satisfied. Further, since the envelope function $F^{(1)}$ is constant, the condition (ii) in (Arcones and Giné, 1994, Theorem 2.1) is also trivially met.

Therefore, we can apply the central limit theorem for bootstrapped U-processes (Arcones and Giné, 1994, Theorem 2.1). Recall the definitions of $G_{1,c}, G_{2,c}, G_{3,c}, G_{4,c}$ in Theorem 3. As $n \to \infty$, conditional on $(T_i, U_i, Z_i), i \geq 1$, for almost every sequence $(T_i, U_i, Z_i), i \geq 1$, we have:

$$\left\{\begin{array}{l}\sqrt{n}\left(W_{c,n}^{(1)^*} - V_{c,w,n}^{(1)}\right)\\ \sqrt{n}\left(M_{c,n}^{(1)^*} - V_{c,m,n}^{(1)}\right)\\ \sqrt{n}\left(W_{c,n}^{(2)^*} - V_{c,w,n}^{(2)}\right)\\ \sqrt{n}\left(M_{c,n}^{(2)^*} - V_{c,m,n}^{(2)}\right)\end{array} : c \in [\ell, u]\right\} \rightsquigarrow \left\{\begin{array}{l}G_{1,c}\\ G_{2,c}\\ G_{3,c}\\ G_{4,c}\end{array} : c \in [\ell, u]\right\} \quad \text{in} \quad \ell^\infty(T),$$

where $\{V_{c,w,n}^{(1)} : c \in [\ell, u]\}$, $\{V_{c,m,n}^{(1)} : c \in [\ell, u]\}$, $\{V_{c,w,n}^{(2)} : c \in [\ell, u]\}$ and $\{V_{c,m,n}^{(2)} : c \in [\ell, u]\}$ are V processes with kernels $w_c^{(1)}(\cdot,\cdot)$, $m_c^{(1)}(\cdot,\cdot)$, $w_c^{(2)}(\cdot,\cdot)$, and $m_c^{(2)}(\cdot,\cdot)$ respectively. Specifically, for each c,

$$V_{c,w,n}^{(1)} = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n w_{c,n}^{(1)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j) = \frac{n(n-1)}{n^2} W_{c,n}^{(1)},$$

$$V_{c,m,n}^{(1)} = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n m_{c,n}^{(1)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j) = \frac{n(n-1)}{n^2} M_{c,n}^{(1)},$$

$$V_{c,w,n}^{(2)} = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n w_{c}^{(2)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j) = \frac{n(n-1)}{n^2} W_{c,n}^{(2)},$$

$$V_{c,m,n}^{(2)} = \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n m_{c,n}^{(2)}(\boldsymbol{\mathcal{D}}_i, \boldsymbol{\mathcal{D}}_j) = \frac{n(n-1)}{n^2} M_{c,n}^{(2)}.$$

As $\left\{\frac{1}{\sqrt{n}} W_{c,n}^{(1)} : c \in [\ell, u]\right\} \to 0$ in $\ell^\infty([\ell, u])$ almost surely (Nolan and Pollard, 1987, Theorem 7), and similar results apply to the other three terms, we can conclude that, conditional on $(T_i, U_i, Z_i), i \geq 1$, for almost every sequence $(T_i, U_i, Z_i), i \geq 1$,

$$\left\{\begin{matrix} \sqrt{n}\left(W_{c,n}^{(1)^*} - W_{c,n}^{(1)}\right) \\ \sqrt{n}\left(M_{c,n}^{(1)^*} - M_{c,n}^{(1)}\right) \\ \sqrt{n}\left(W_{c,n}^{(2)^*} - W_{c,n}^{(2)}\right) \\ \sqrt{n}\left(M_{c,n}^{(2)^*} - M_{c,n}^{(2)}\right) \end{matrix} : c \in [\ell, u]\right\} \rightsquigarrow \left\{\begin{matrix} G_{1,c} \\ G_{2,c} \\ G_{3,c} \\ G_{4,c} \end{matrix} : c \in [\ell, u]\right\} \quad \text{in} \quad \ell^\infty(T).$$

Recall that ϕ is the inverse map defined in (S7.2). By Section 2.2.4 and Lemma 12.2 in Kosorok (2008), the map ϕ is Hadamard-differentiable with the same derivative map as discussed in the proof of Theorem 3. Then by the bootstrap version of the functional delta method (Kosorok, 2008, Theorem 12.1), we obtain that

$$\left\{\begin{matrix} \sqrt{n}\left(W_{c,n}^{(1)^*} - W_{c,n}^{(1)}\right) \\ \sqrt{n}\left(M_{c,n}^{(1)^{*-1}} - M_{c,n}^{(1)^{-1}}\right) \\ \sqrt{n}\left(W_{c,n}^{(2)^*} - W_{c,n}^{(2)}\right) \\ \sqrt{n}\left(M_{c,n}^{(2)^{*-1}} - M_{c,n}^{(2)^{-1}}\right) \end{matrix} : c \in [\ell, u]\right\} \rightsquigarrow \left\{\begin{matrix} G_{1,c} \\ -G_{2,c}/(E[M_{c,n}^{(1)}])^2 \\ G_{3,c} \\ -G_{4,c}/(E[M_{c,n}^{(2)}])^2 \end{matrix} : c \in [\ell, u]\right\} \quad \text{in} \quad \ell^\infty(T),$$

conditional on $(T_i, U_i, Z_i), i \geq 1$, for almost every sequence $(T_i, U_i, Z_i), i \geq 1$.

Moreover, by the bootstrap law of large numbers for U processes (Arcones and Giné, 1994) and the continuous mapping theorem (Van der Vaart and Wellner, 1996, Theorem 1.3.6), conditional on $(T_i, U_i, Z_i), i \geq 1$, for almost every sequence $(T_i, U_i, Z_i), i \geq 1$, in $\ell^\infty(T')$,

$$\left\{ \begin{matrix} M_{c,n}^{(1)^{*-1}} \\ M_{c,n}^{(2)^{*-1}} \end{matrix} : c \in [\ell, u] \right\} \rightsquigarrow \left\{ \begin{matrix} E\left[M_{c,n}^{(1)}\right]^{-1} \\ E\left[M_{c,n}^{(2)}\right]^{-1} \end{matrix} : c \in [\ell, u] \right\}.$$

Employing the law of large numbers for U processes (Peña and Giné, 1999, Corollary 5.2.3), we can conclude that almost surely,

$$\left\{ \begin{matrix} W_{c,n}^{(1)} \\ W_{c,n}^{(2)} \end{matrix} : c \in [\ell, u] \right\} \rightsquigarrow \left\{ \begin{matrix} E\left[W_{c,n}^{(1)}\right] \\ E\left[W_{c,n}^{(2)}\right] \end{matrix} : c \in [\ell, u] \right\} \quad \text{in} \quad \ell^\infty(T').$$

Finally, by the Slutsky's theorem (Kosorok, 2008, Theorem 7.15) and continuous mapping theorem (Van der Vaart and Wellner, 1996, Theorem 1.3.6), we have that, conditional on $(T_i, U_i, Z_i), i \geq 1$, for almost every sequence $(T_i, U_i, Z_i), i \geq 1$, the process $\{\sqrt{n}(\hat{\theta}^*_{n,c} - \hat{\theta}_{n,c}) : c \in [\ell, u]\}$ converges to the same Gaussian process $\{G_c : c \in [\ell, u]\}$ in $\ell^\infty([\ell, u])$ defined in (S7.3).

The previous results, together with Theorem 3, imply that, under H_0, the distribution of $\sup_{c\in[\ell,u]}\{\sqrt{n}|\hat{\theta}_{n,c}|\}$ can be approximated by the distribution of $\sup_{c\in[\ell,u]}\{\sqrt{n}|\hat{\theta}^*_{n,c} - \hat{\theta}_{n,c}|\}$ by the continuous mapping theorem. Specifically, we define $J := \sup_{c\in[\ell,u]} |G_c|$, and denote by F_J its cumulative distribution function and by F_J^{-1} its quantile function. Then we have that $\sup_{c\in[\ell,u]} \left\{\sqrt{n}|\hat{\theta}_{n,c}|\right\} \rightsquigarrow J$ and that conditional on $(T_i, U_i, Z_i), i \geq 1$, for almost every sequence $(T_i, U_i, Z_i), i \geq 1$,

$$\sup_{c\in[\ell,u]} \left\{\sqrt{n}|\hat{\theta}^*_{n,c} - \hat{\theta}_{n,c}|\right\} \rightsquigarrow J.$$

Since $\mathbf{G}$ is a tight centered Gaussian process, F_J is absolutely continuous on $(0,\infty)$, and F_J^{-1} is continuous and strictly increasing on $(0,1)$ (Ledoux and Talagrand, 1991; Davydov et al., 1998). Then by Lemma 21.2 (Van der Vaart, 2007) and the definition of weak convergence, conditional on almost all sequences $(T_i, U_i, Z_i), i \geq 1$, for any $\alpha \in (0,1)$, we have as $n \to \infty$,

$$\begin{aligned}\Pr\left(p^*_{adjust} \leq \alpha\right) &= \Pr\left(\sqrt{n} \sup_{c\in[\ell,u]} |\hat{\theta}_{n,c}| \geq (F_n^*)^{-1}(1-\alpha)\right)\\ &\to \Pr\left(J - F_J^{-1}(1-\alpha) \geq 0\right) = \alpha.\end{aligned}$$

The proof is complete. □

S8 Proofs regarding cut-point estimation

S8.1 Consistency and convergence rate

Proof of Theorem 5 - consistency. Note that the alternative hypothesis holds, i.e. $\theta_{c_b} \neq 0$. By the definition of $\hat{c}_n$ and c_b, for any $\epsilon > 0$, we have

$$\Pr\left(|\hat{c}_n - c_b| \geq \epsilon\right) \leq \Pr\left(|\hat{\theta}_{n,c_b}| \leq \sup_{|c-c_b|\geq\epsilon, c\in[\ell,u]} |\hat{\theta}_{n,c}|\right).$$

Then by triangle inequality, we have

$$\begin{aligned}\Pr\left(|\hat{c}_n - c_b| \geq \epsilon\right) &\leq \Pr\left(|\theta_{c_b}| - |\hat{\theta}_{n,c_b} - \theta_{c_b}| \leq \sup_{|c-c_b|\geq\epsilon, c\in[\ell,u]} |\hat{\theta}_{n,c} - \theta_c| + \sup_{|c-c_b|\geq\epsilon, c\in[\ell,u]} |\theta_c|\right)\\ &= \Pr\left(\inf_{|c-c_b|\geq\epsilon, c\in[\ell,u]} |\theta_{c_b}| - |\theta_c| \leq 2\sup_{c\in[\ell,u]} |\hat{\theta}_{n,c} - \theta_c|\right).\end{aligned}$$

Let $\delta = \inf_{|c-c_b|\geq\epsilon, c\in[\ell,u]} |\theta_{c_b}| - |\theta_c|$. By assumption (*C.1*), $\delta > 0$, and thus

$$\Pr\left(|\hat{c}_n - c_b| \geq \epsilon\right) \leq \Pr\left(\sup_{c\in[\ell,u]} |\hat{\theta}_{n,c} - \theta_c| \geq \delta/2\right)$$

$$\leq \Pr\left(\sup_{c\in[\ell,u]}\left|\frac{W_{c,n}^{(1)}}{M_{c,n}^{(1)}} - \frac{E\left[W_{c,n}^{(1)}\right]}{E\left[M_{c,n}^{(1)}\right]}\right| \geq \frac{\delta}{4}\right) + \Pr\left(\sup_{c\in[\ell,u]}\left|\frac{W_{c,n}^{(2)}}{M_{c,n}^{(2)}} - \frac{E\left[W_{c,n}^{(2)}\right]}{E\left[M_{c,n}^{(2)}\right]}\right| \geq \frac{\delta}{4}\right)$$

$$\leq \Pr\left(\sup_{c\in[\ell,u]}\left|\frac{W_{c,n}^{(1)} - E\left[W_{c,n}^{(1)}\right]}{M_{c,n}^{(1)}}\right| \geq \frac{\delta}{8}\right) + \Pr\left(\sup_{c\in[\ell,u]}\left|\frac{1}{M_{c,n}^{(1)}} - \frac{1}{E\left[M_{c,n}^{(1)}\right]}\right| \geq \frac{\delta}{8}\right)$$

$$+ \Pr\left(\sup_{c\in[\ell,u]}\left|\frac{W_{c,n}^{(2)} - E\left[W_{c,n}^{(2)}\right]}{M_{c,n}^{(2)}}\right| \geq \frac{\delta}{8}\right) + \Pr\left(\sup_{c\in[\ell,u]}\left|\frac{1}{M_{c,n}^{(2)}} - \frac{1}{E\left[M_{c,n}^{(2)}\right]}\right| \geq \frac{\delta}{8}\right),$$

where we use the fact that $\sup_{c\in[\ell,u]}\left|E\left[W_{c,n}^{(1)}\right]\right| \leq 1$ and $\sup_{c\in[\ell,u]}\left|E\left[W_{c,n}^{(2)}\right]\right| \leq 1$ (see (S5.6)).

Since $\mathcal{F}_1$, $\mathcal{F}_2$, $\mathcal{F}_3$ and $\mathcal{F}_4$ defined in (S5.7) are VC-subgraph classes (Van der Vaart and Wellner, 1996, Section 2.6.2), the proof is complete due to condition (*C.0*) and the law of large numbers for U processes (Peña and Giné, 1999, Corollary 5.2.3). □

Proof of Theorem 5 - convergence rate. We assume $\theta_{c_b} > 0$, noting that the arguments are similar for the case $\theta_{c_b} < 0$. The goal is to show that for any $\epsilon > 0$, there exists $M > 0$ such that

$$\limsup_{n\to\infty} \Pr(n^{1/3}|\hat{c}_n - c_b| > 2^M) < \epsilon.$$

We use a modified "peeling device" (Van der Vaart and Wellner, 1996, Theorem 3.2.5). Specifically, for each n, the parameter space minus the point c_b can be partitioned into "peels" $S_{n,j} = \{c \in [\ell, u] : 2^{j-1} < n^{1/3}|c - c_b| \leq 2^j\}$ for $j \geq 1$. Due to condition $(C.0)$, and since $\theta_{c_b} > 0$ and the function $c \mapsto \theta_c$ is continuous on $[\ell, u]$, there exists $\iota > 0$ such that

$$2\lambda(1-\lambda) \inf_{c\in[\ell,u]} \min\left\{(1-F(c))^2, F^2(c)\right\} \geq \iota, \quad \text{and} \quad \inf_{|c-c_b|\leq\iota,\, c\in[\ell,u]} \theta_c > 0. \tag{S8.1}$$

For this positive ι, we define an event A_n as follows:

$$A_n := \left\{ \inf_{c\in[\ell,u]} M^{(1)}_{c,n} \geq \iota/2,\ \inf_{c\in[\ell,u]} M^{(2)}_{c,n} \geq \iota/2,\ \inf_{|c-c_b|\leq\iota, c\in[\ell,u]} \hat{\theta}_{n,c} > 0,\ |\hat{c}_n - c_b| \leq \zeta/2 \right\}, \tag{S8.2}$$

where $\zeta > 0$ appears in condition $(C.2)$. By the law of large numbers for U processes (Peña and Giné, 1999, Corollary 5.2.3) and the continuous mapping theorem (Van der Vaart and Wellner, 1996, Theorem 1.3.6), we have

$$\sup_{c\in[\ell,u]} \left|M^{(1)}_{c,n} - 2\lambda(1-\lambda)(1-F(c))^2\right| \xrightarrow{P} 0, \quad \sup_{c\in[\ell,u]} \left|M^{(2)}_{c,n} - 2\lambda(1-\lambda)F^2(c)\right| \xrightarrow{P} 0,$$
$$\sup_{c\in[\ell,u]} \left|\hat{\theta}_{n,c} - \theta_c\right| \xrightarrow{P} 0.$$

As a result, together with the consistency of $\hat{c}_n$, we have $\lim_{n\to\infty} \Pr(A_n^{\complement}) = 0$.

Then due to condition $(C.2)$, we have

$$\Pr\left(n^{1/3}|\hat{c}_n - c_b| > 2^M\right) \leq \Pr\left(n^{1/3}|\hat{c}_n - c_b| > 2^M, A_n\right) + \Pr\left(A_n^{\complement}\right)$$
$$\leq \sum_{\substack{j>M \\ 2^{j-1} \leq n^{1/3}\zeta/2}} \Pr\left(\hat{c}_n \in S_{n,j}, A_n\right) + o(1),$$

where $o(1)$ denotes a sequence of deterministic numbers that converges to zero as $n \to \infty$. Since by definition $\hat{\theta}_{n,\hat{c}_n} - \hat{\theta}_{n,c_b} \geq 0$, then on the event $\{\hat{c}_n \in S_{n,j}\}$, it implies that $\sup_{c \in S_{n,j}} \left(\hat{\theta}_{n,c} - \hat{\theta}_{n,c_b}\right) \geq 0$. Thus, we obtain that

$$\Pr\left(n^{1/3}|\hat{c}_n - c_b| > 2^M\right) \leq \sum_{\substack{j>M \\ 2^{j-1} \leq n^{1/3}\zeta/2}} \Pr\left(\sup_{c \in S_{n,j}} \left(\hat{\theta}_{n,c} - \hat{\theta}_{n,c_b}\right) \geq 0, A_n\right) + o(1).$$

Further, by the definition of $S_{n,j}$ and condition $(C.2)$, if $2^{j-1} \leq n^{1/3}\zeta/2$, then

$$|c - c_b| \leq \frac{2^j}{n^{1/3}} \leq \zeta, \quad |c - c_b| > \frac{2^{j-1}}{n^{1/3}} \quad \Rightarrow \quad \theta_c - \theta_{c_b} \leq -\kappa \frac{2^{2j-2}}{n^{2/3}},$$

where the constant κ appears in condition $(C.2)$ and recall that we assume $\theta_{c_b} > 0$. Thus, by the Markov inequality, we have

$$\Pr\left(n^{1/3}|\hat{c}_n - c_b| > 2^M\right)$$
$$\leq \sum_{\substack{j>M \\ 2^{j-1} \leq n^{1/3}\zeta/2}} \Pr\left(\sup_{c \in S_{n,j}} \left|\hat{\theta}_{n,c} - \theta_c - (\hat{\theta}_{n,c_b} - \theta_{c_b})\right| \geq \kappa \frac{2^{2j-2}}{n^{2/3}}, A_n\right) + o\left(1\right)$$
$$\leq \kappa^{-1} \sum_{\substack{j>M \\ 2^{j-1} \leq n^{1/3}\zeta/2}} E\left[\sup_{c \in S_{n,j}} \left(\left|\hat{\theta}_{n,c} - \theta_c - (\hat{\theta}_{n,c_b} - \theta_{c_b})\right|\right) \mathbb{1}_{A_n}\right] \frac{n^{2/3}}{2^{2j-2}} + o(1),$$

Finally, by Lemma 2, for some constant $C > 0$,

$$\Pr\left(n^{1/3}|\hat{c}_n - c_b| > 2^M\right) \leq C\kappa^{-1}\sum_{j>M}\frac{\sqrt{2^j/n^{1/3}}n^{2/3}}{\sqrt{n}2^{2j-2}} + o\left(1\right) = C\kappa^{-1}\sum_{j>M}2^{-\frac{3}{2}j+2} + o\left(1\right).$$

We note that $\sum_{j>M}2^{-\frac{3}{2}j}$ converges to zero as $M \to \infty$. The proof is complete. □

Recall the constant $\iota > 0$ in (S8.1), and the constant $\zeta > 0$ in condition (*C.2*). Recall the event A_n defined in (S8.2).

Lemma 2. *Assume the conditions in Theorem 5 hold. There exists $K > 0$, that only depends on F and ι, such that for every $n \geq 1$ and $0 < \delta \leq \zeta$, we have*

$$E\left[\sup_{|c-c_b|\leq\delta,\ c\in[\ell,u]}\left(\left|\hat{\theta}_{n,c} - \theta_c - (\hat{\theta}_{n,c_b} - \theta_{c_b})\right|\right)\mathbb{1}_{A_n}\right] \leq K\frac{\sqrt{\delta}}{\sqrt{n}}.$$

Proof. In this proof, we use K to denote a constant, that only depends on F, ι, which may vary from line to line. Further, c is always assumed to be in $[\ell, u]$. By the triangle inequality,

$$\begin{aligned}&\sqrt{n}E\left[\sup_{|c-c_b|\leq\delta}\left(\left|\hat{\theta}_{n,c} - \theta_c - (\hat{\theta}_{n,c_b} - \theta_{c_b})\right|\right)\mathbb{1}_{A_n}\right]\\ &\leq \sqrt{n}\sum_{i=1}^{2}E\left[\sup_{|c-c_b|\leq\delta}\left(\left|\frac{W_{c,n}^{(i)}}{M_{c,n}^{(i)}} - \frac{E\left[W_{c,n}^{(i)}\right]}{E\left[M_{c,n}^{(i)}\right]} - \left(\frac{W_{c_b,n}^{(i)}}{M_{c_b,n}^{(i)}} - \frac{E\left[W_{c_b,n}^{(i)}\right]}{E\left[M_{c_b,n}^{(i)}\right]}\right)\right|\right)\mathbb{1}_{A_n}\right].\end{aligned} \tag{S8.3}$$

We deal with the case $i = 1$, noting that the arguments for the case $i = 2$ are similar. For the term above with $i = 1$, it can be upper bounded by the sum of four terms I_δ, II_δ, III_δ and IV_δ, where

$$I_\delta = \sqrt{n} E\left[\sup_{|c-c_b|\le\delta}\left(\left|\frac{1}{M_{c,n}^{(1)}}\left(W_{c,n}^{(1)} - E\left[W_{c,n}^{(1)}\right] - W_{c_b,n}^{(1)} + E\left[W_{c_b,n}^{(1)}\right]\right)\right|\right)\mathbb{1}_{A_n}\right],$$

$$II_\delta = \sqrt{n} E\left[\sup_{|c-c_b|\le\delta}\left(\left|\frac{E\left[W_{c,n}^{(1)}\right]}{M_{c,n}^{(1)} E\left[M_{c,n}^{(1)}\right]}\left(M_{c,n}^{(1)} - E\left[M_{c,n}^{(1)}\right] - M_{c_b,n}^{(1)} + E\left[M_{c_b,n}^{(1)}\right]\right)\right|\right)\mathbb{1}_{A_n}\right],$$

$$III_\delta = \sqrt{n} E\left[\sup_{|c-c_b|\le\delta}\left(\left|\left(W_{c_b,n}^{(1)} - E\left[W_{c_b,n}^{(1)}\right]\right)\left(\frac{1}{M_{c,n}^{(1)}} - \frac{1}{M_{c_b,n}^{(1)}}\right)\right|\right)\mathbb{1}_{A_n}\right],$$

$$IV_\delta = \sqrt{n} E\left[\sup_{|c-c_b|\le\delta}\left(\left|\left(M_{c_b,n}^{(1)} - E\left[M_{c_b,n}^{(1)}\right]\right)\left(\frac{E\left[W_{c,n}^{(1)}\right]}{M_{c,n}^{(1)} E\left[M_{c,n}^{(1)}\right]} - \frac{E\left[W_{c_b,n}^{(1)}\right]}{M_{c_b,n}^{(1)} E\left[M_{c_b,n}^{(1)}\right]}\right)\right|\right)\mathbb{1}_{A_n}\right].$$

Then it suffices to show that $I_\delta, II_\delta, III_\delta, IV_\delta \le K\sqrt{\delta}$ for any $0 \le \delta \le \zeta$ and $n \ge 1$. Fix some $0 \le \delta \le \zeta$ and $n \ge 1$.

Upper bounding I_δ. Define a class of functions $\mathcal{F}_\delta^1 := \{w_{c,n}^{(1)} - w_{c_b,n}^{(1)} : |c - c_b| \le \delta\}$, where recall that $w_{c,n}^{(1)}(\cdot,\cdot)$ is a function of $\mathcal{S}^2$ defined in (S5.4). Since on the event A_n, $\sup_{|c-c_n|\le\delta} 1/M_{c,n}^{(1)} \le 2/\iota$, we have

$$\begin{aligned} I_\delta &\le K\sqrt{n} E\left[\sup_{|c-c_b|\le\delta}\left|W_{c,n}^{(1)} - E\left[W_{c,n}^{(1)}\right] - W_{c_b,n}^{(1)} + E\left[W_{c_b,n}^{(1)}\right]\right|\right], \\ &= K\sqrt{n} E\left[\|U_n^{(2)}(f) - P^2 f\|_{\mathcal{F}_\delta^1}\right], \end{aligned}$$

where we recall the notation for U-process in (S5.1) and the definition of $\|\cdot\|_{\mathcal{F}_\delta^1}$ in Appendix S5. By Hoeffding decomposition (Van der Vaart, 2007), the centered U process can be decomposed into two parts: for $f \in \mathcal{F}_\delta^1$,

$$U_n^{(2)}(f) - P^2 f = 2U_n^{(1)}(\pi_1 f) + U_n^{(2)}(\pi_2 f), \tag{S8.4}$$

where $\pi_1 f(\boldsymbol{\mathcal{D}}_1) = E[f(\boldsymbol{\mathcal{D}}_1, \boldsymbol{\mathcal{D}}_2)|\boldsymbol{\mathcal{D}}_1] - P^2 f$, $\pi_2 f(\boldsymbol{\mathcal{D}}_1, \boldsymbol{\mathcal{D}}_2) = f(\boldsymbol{\mathcal{D}}_1, \boldsymbol{\mathcal{D}}_2) - \pi_1 f(\boldsymbol{\mathcal{D}}_1) - \pi_1 f(\boldsymbol{\mathcal{D}}_2) + P^2 f$, and both $U_n^{(1)}(\pi_1 f)$ and $U_n^{(2)}(\pi_2 f)$ denote the U processes with kernels $\pi_1 f$ and $\pi_2 f$ respectively. Thus,

$$\begin{aligned} I_c &\leq K\sqrt{n}\left(E\left[\left\|U_n^{(1)}(\pi_1 f)\right\|_{\mathcal{F}_\delta^1}\right] + E\left[\left\|U_n^{(2)}(\pi_2 f)\right\|_{\mathcal{F}_\delta^1}\right]\right) \\ &\leq K J_1(1, \mathcal{F}_\delta^1, F_\delta^1)\left\|PF_\delta^1\right\|_{P,2} + \frac{K}{\sqrt{n}} J_2(1, \mathcal{F}_\delta^1, F_\delta^1)\left\|F_\delta^1\right\|_{P^2,2}, \end{aligned}$$

where the last inequality holds by maximal inequality for U processes (Chen and Kato, 2019, Corollary 5.6), $J_1(1, \mathcal{F}_\delta^1, F_\delta^1), J_2(1, \mathcal{F}_\delta^1, F_\delta^1)$ are uniform entropy, defined as

$$J_k(1, \mathcal{F}_\delta^1, F_\delta^1) = \int_0^1 \sup_Q \left[1 + \log N(P^{2-k}\mathcal{F}_\delta^1, \|\cdot\|_{Q,2}, \tau\left\|P^{2-k}F_\delta^1\right\|_{Q,2})\right]^{k/2} d\tau, \quad k = 1, 2,$$

where recall that $N(T, d, \epsilon)$ denotes the ϵ-covering number for pseudometric space (T, d), $\sup_Q$ is taken over all finitely discrete distributions on $\mathcal{S}^k$, and $F_\delta^1(\boldsymbol{d}_1, \boldsymbol{d}_2) := I(c_b - \delta \leq z_1 \leq c_b + \delta) + I(c_b - \delta \leq z_2 \leq c_b + \delta)$ for $\boldsymbol{d}_1 = (t_1, u_1, z_1), \boldsymbol{d}_2 = (t_2, u_2, z_2) \in \mathcal{S}^2$ is the envelope function for $\mathcal{F}_\delta^1$.

Since $\mathcal{F}_\delta^1$ is a VC-type class by Lemma A.6, Corollary A.1 in Chernozhukov et al. (2014), thus $J_k(1, \mathcal{F}_\delta^1, F_\delta^1) \leq K$ for $k = 1, 2$. Further, by Jensen's Inequality and the mean value theorem, we have

$$\left\|PF_\delta^1\right\|_{P,2} \leq \left\|F_\delta^1\right\|_{P^2,2} \leq \sqrt{K\left(F(c_b + \delta) - F(c_b - \delta)\right)} \leq K\sqrt{\delta},$$

where the last inequality is because F is continuous differentiable on $[c_b - \zeta, c_b + \zeta]$ by condition $(C.2)$. Thus, I_δ is upper bounded by $K\sqrt{\delta}$.

Upper bounding II_δ. Define a class of functions $\mathcal{F}_\delta^2 = \{m_{c,n}^{(1)} - m_{c_b,n}^{(1)} : |c - c_b| \leq \delta\}$ with envelope function $F_\delta^2(\boldsymbol{d}_1, \boldsymbol{d}_2) = I(c_b - \delta \leq z_1 \leq c_b + \delta) + I(c_b - \delta \leq z_2 \leq c_b + \delta)$ for $\boldsymbol{d}_1 = (t_1, u_1, z_1), \boldsymbol{d}_2 = (t_2, u_2, z_2) \in \mathcal{S}^2$; note that $F_2^\delta = F_1^\delta$. As $\mathcal{F}_\delta^2$ is also a VC-type class (Chernozhukov et al., 2014, Lemma A.6, Corollary A.1), by similar arguments as those used for I_δ, we have that

$$II_\delta \leq KJ_1(1, \mathcal{F}_\delta^2, F_\delta^2) \left\| PF_\delta^2 \right\|_{P,2} + \frac{K}{\sqrt{n}} J_2(1, \mathcal{F}_\delta^2, F_\delta^2) \left\| F_\delta^2 \right\|_{P^2,2} \leq K\sqrt{\delta}.$$

Upper bounding III_δ. By Cauchy-Schwarz inequality, it can be upper bounded by the product of $III_{\delta,1}$ and $III_{\delta,2}$, which are defined as follows:

$$III_{\delta,1} := \sqrt{n}\sqrt{E\left[\left(W_{c_b,n}^{(1)} - E\left[W_{c_b,n}^{(1)}\right]\right)^2\right]}, \quad III_{\delta,2} := \sqrt{E\left[\sup_{|c-c_b|\leq\delta}\left|\frac{1}{M_{c,n}^{(1)}} - \frac{1}{M_{c_b,n}^{(1)}}\right|^2 \mathbb{1}_{A_n}\right]}.$$

Note that $\sqrt{n}(\left(W_{c_b,n}^{(1)} - E\left[W_{c_b,n}^{(1)}\right]\right))$ is a normalized U statistic with a bounded kernel $w_{c_b}^{(1)}(\cdot,\cdot)$ in (S5.4); thus, by (Van der Vaart, 2007, Theorem 12.3), $III_{\delta,1} \leq K$ for $n \geq 1$. For $III_{\delta,2}$, it can be further decomposed and upper bounded by $K\sqrt{III_{\delta,2}^{(1)}} + K\sqrt{III_{\delta,2}^{(2)}}$, where

$$III_{\delta,2}^{(1)} := E\left[\sup_{|c-c_b|\leq\delta}\left|\frac{M_{c,n}^{(1)} - E\left[M_{c,n}^{(1)}\right] - \left(M_{c_b,n}^{(1)} - E\left[M_{c_b,n}^{(1)}\right)\right]}{M_{c,n}^{(1)} M_{c_b,n}^{(1)}}\right|^2 \mathbb{1}_{A_n}\right],$$

$$III_{\delta,2}^{(2)} := E\left[\sup_{|c-c_b|\leq\delta}\left|\frac{E\left[M_{c,n}^{(1)}\right] - E\left[M_{c_b,n}^{(1)}\right]}{M_{c,n}^{(1)} M_{c_b,n}^{(1)}}\right|^2 \mathbb{1}_{A_n}\right].$$

By similar arguments as those used for I_δ, $III_{\delta,2}^{(1)} \leq K\delta$. Further, by the definition of A_n and the mean value theorem, due to condition $(C.2)$, $III_{\delta,2}^{(2)} \leq K\delta$. As a result, $III_\delta \leq K\sqrt{\delta}$.

By similar arguments as before, we can show $IV_\delta \leq K\sqrt{\delta}$. The proof is complete.

□

S8.2 Discussions on the rate of convergence

In Theorem 5, under Assumption $(C.2)$, the estimator converges at rate $n^{-1/3}$ rather than the usual parametric rate $n^{-1/2}$. Below we explain the origin of this nonstandard cube-root rate. Moreover, when θ_c is not smooth at c_b, the convergence rate further accelerates to n^{-1}.

Specifically, assume without loss of generality that $\theta_{c_b} > 0$. Under Assumption $(C.2)$, we have the following expansion of the population-level criterion $c \mapsto \theta_c$: for all sufficiently small $\delta \in \mathbb{R}$,

$$\theta_{c_b+\delta} - \theta_{c_b} \leq -\kappa\delta^2. \tag{S8.5}$$

This quadratic drift follows directly from the local condition $|\theta_c| - |\theta_{c_b}| \leq -\kappa|c - c_b|^2$.

Furthermore, Lemma 2 above gives the order of the stochastic fluctuation:

$$\hat{\theta}_{n,c_b+\delta} - \theta_{c_b+\delta} - (\hat{\theta}_{n,c_b} - \theta_{c_b}) = O_P\left(\frac{\sqrt{\delta}}{\sqrt{n}}\right). \tag{S8.6}$$

The factor $\sqrt{\delta}$ arises because the standard deviation of the indicator $\mathbb{1}\{c_b < Z < c_b + \delta\}$ equals

$$\sqrt{F(c_b + \delta) - F(c_b)} = O(\sqrt{\delta}),$$

so the empirical fluctuation over an interval of width δ is of order $\sqrt{\delta}/\sqrt{n}$.

The convergence rate is then obtained by balancing the deterministic drift and the stochastic fluctuation:

$$\delta^2 = \frac{\sqrt{\delta}}{\sqrt{n}} \qquad \Rightarrow \qquad \delta = n^{-1/3}.$$

For a *regular* parameter, the quadratic expansion in (S8.5) still holds. However, smoothness of the estimator, in contrast to the indicator structure above, implies that the fluctuation term in (S8.6) becomes $O(\delta)$ instead of $O(\sqrt{\delta})$. Balancing the deterministic drift and the stochastic fluctuation yields

$$\delta^2 = \frac{\delta}{\sqrt{n}} \qquad \Rightarrow \qquad \delta = n^{-1/2},$$

which is the usual parametric rate.

As discussed in Remark 8, our main results focus on the smooth setup. When $c \mapsto \theta_c$ is not smooth, the rate of convergence increases to n^{-1}, as established in the following theorem.

(*C2'*) Assume that the function $c \mapsto \theta_c$ is continuous on $[\ell, u]$. Further, there exist $\zeta > 0$ and $\kappa > 0$ such that if $|c - c_b| \leq \zeta$ and $c \in [\ell, u]$, $|\theta_c| - |\theta_{c_b}| \leq -\kappa|c - c_b|$, and that the function $c \mapsto F(c)$ is continuously differentiable on $[c_b - \zeta, c_b + \zeta]$.

Remark 2. Compared to condition (*C.2*), we require in (*C2'*) the condition that $|\theta_c| - |\theta_{c_b}| \leq -\kappa|c - c_b|$ for c sufficiently close to c_b. This occurs, for example, when $\theta_c = \theta_{c_b} - |c - c_b| + o(|c - c_b|)$.

Theorem 1. *Suppose that conditions (C.0), (C.1) and (C2') hold. Then* $n|\hat{c}_n - c_b|$ *is bounded in probability as* $n \to \infty$.

Proof. By the first part of Theorem 5, which only requires conditions (*C.0*) and (*C.1*), we have the consistency: $\hat{c}_n \xrightarrow{P} c_b$ as $n \to \infty$.

Here, we aim to show that for any $\epsilon > 0$, there exists $M > 0$ such that

$$\limsup_{n \to \infty} \Pr(n|\hat{c}_n - c_b| > 2^M) < \epsilon.$$

Without loss of generality, consider the case $\theta_{c_b} > 0$. Using a peeling argument (Van der Vaart and Wellner, 1996), we partition the parameter space into shells $S_{n,j} = \{c : 2^{j-1} < n|c - c_b| \leq 2^j\}$. Due to assumption (*C2'*),

$$|c - c_b| > \frac{2^{j-1}}{n} \quad \Rightarrow \quad \theta_c - \theta_{c_b} \leq -\kappa \frac{2^{j-1}}{n}.$$

By the Markov inequality, we have

$$\begin{aligned}
&\Pr\left(n|\hat{c}_n - c_b| > 2^M\right) \\
&\leq \sum_{\substack{j>M \\ 2^{j-1}\leq n\zeta/2}} \Pr\left(\sup_{c\in S_{n,j}} \left|\hat{\theta}_{n,c} - \theta_c - (\hat{\theta}_{n,c_b} - \theta_{c_b})\right| \geq \kappa\frac{2^{j-1}}{n}, A_n\right) + o\left(1\right) \\
&\leq \kappa^{-1} \sum_{\substack{j>M \\ 2^{j-1}\leq n\zeta/2}} E\left[\sup_{c\in S_{n,j}} \left(\left|\hat{\theta}_{n,c} - \theta_c - (\hat{\theta}_{n,c_b} - \theta_{c_b})\right|\right) \mathbb{1}_{A_n}\right] \frac{n}{2^{j-1}} + o(1),
\end{aligned}$$

Then, by Lemma 2 that

$$E\left[\sup_{|c-c_b|\leq\delta,\, c\in[\ell,u]} \left(\left|\hat{\theta}_{n,c} - \theta_c - (\hat{\theta}_{n,c_b} - \theta_{c_b})\right|\right) \mathbb{1}_{A_n}\right] \leq K\frac{\sqrt{\delta}}{\sqrt{n}},$$

for some constant $C > 0$, we have

$$\Pr\left(n|\hat{c}_n - c_b| > 2^M\right) \leq C\kappa^{-1} \sum_{j>M} \frac{\sqrt{\frac{2^j}{n}n}}{\sqrt{n}2^{j-1}} + o\left(1\right) = C\kappa^{-1} \sum_{j>M} 2^{-\frac{1}{2}j+1} + o\left(1\right).$$

Since $\sum_{j>M} 2^{-j/2}$ converges to zero as $M \to \infty$, we conclude that $\hat{c}_n - c_b = O_p(n^{-1})$. The proof is complete. □

Under condition (*C2'*), for all sufficiently small $\delta \in \mathbb{R}$,

$$\theta_{c_b+\delta} - \theta_{c_b} \leq -\kappa\delta.$$

Furthermore, (S8.6) continues to hold:

$$\hat{\theta}_{n,c_b+\delta} - \theta_{c_b+\delta} - (\hat{\theta}_{n,c_b} - \theta_{c_b}) = O_P\left(\frac{\sqrt{\delta}}{\sqrt{n}}\right).$$

The convergence rate is thus:

$$\delta = \frac{\sqrt{\delta}}{\sqrt{n}} \qquad \Rightarrow \qquad \delta = n^{-1}.$$

Finally, note that our paper focuses on the smooth case described in Theorem 6, since in most practical applications the predictive effects vary gradually around the optimal cutpoint.

S8.3 Limiting distribution of the profile estimator

Proof of Theorem 6. As in the proof of Theorem 5, we assume $\theta_{c_b} > 0$, noting that arguments are similar for the case $\theta_{c_b} < 0$. Recall the constant ι in (S8.1); in particular, $\inf_{|c-c_b|\leq\iota,\, c\in[\ell,u]} \theta_c > 0$.

Let $\ell_n := (\ell - c_b)n^{1/3}$ and $u_n := (u - c_b)n^{1/3}$. By condition (*C.3*), $\lim_{n\to\infty} \ell_n = \lim_{n\to\infty} u_n = \infty$. Define two stochastic processes indexed by $\mathbb{R}$ as follows:

$$\tilde{M}_n(h) := \begin{cases} n^{2/3}\left(\left|\hat{\theta}_{n,c_b+h/n^{1/3}}\right| - \left|\hat{\theta}_{n,c_b}\right|\right), & \text{if } h \in [\ell_n, u_n] \\ -\infty, & \text{otherwise} \end{cases}.$$

$$M_n(h) := n^{2/3}\left(\hat{\theta}_{n,c_b+h/n^{1/3}} - \hat{\theta}_{n,c_b}\right), \quad \text{for } h \in \mathbb{R}.$$

Then by definition,

$$n^{1/3}(\hat{c}_n - c_b) = \operatorname*{arg\,max}_{h\in\mathbb{R}} \tilde{M}_n(h). \tag{S8.7}$$

Further, fix any $K > 0$. For large enough n, we have $K/n^{1/3} \leq$

$\{\iota, |u - c_b|, |c_b - \ell|\}$. Thus for $|h| \leq K$ and large enough n,

$$\begin{aligned}\tilde{M}_n(h) &= \tilde{M}_n(h)\mathbb{1}\left\{\inf_{|c-c_b|\leq\iota} \hat{\theta}_{n,c} > 0\right\} + \tilde{M}_n(h)\mathbb{1}\left\{\inf_{|c-c_b|\leq\iota} \hat{\theta}_{n,c} \leq 0\right\} \\ &= M_n(h) + o_P(1),\end{aligned} \tag{S8.8}$$

where $o_P(1)$ is uniform in $|h| \leq K$, and the last equality holds since as shown in the proof of Theorem 5, $\sup_{c\in[\ell,u]} |\hat{\theta}_{n,c} - \theta_c| = o_P(1)$ and by assumption $\inf_{|c-c_b|\leq\iota,\, c\in[\ell,u]} \theta_c > 0$.

Since $n^{1/3}(\hat{c}_n - c_b) = O_P(1)$ as shown in Theorem 5, in view of (S8.7) and (S8.8), by Theorem 3.2.2 in Van der Vaart and Wellner (1996), if there exists a tight, zero-mean Gaussian process $\{\widetilde{G}(h) : h \in \mathbb{R}\}$, with continuous sample paths and a unique (random) maximizer, such that $\{M_n(h) : |h| \leq K\} \rightsquigarrow \{\widetilde{G}(h) : |h| \leq K\}$ in $\ell^\infty([-K, K])$ for any $K > 0$, then we have

$$n^{1/3}(\hat{c}_n - c_b) \ \rightsquigarrow\ \arg\max_{h\in\mathbb{R}} \widetilde{G}(h). \tag{S8.9}$$

Next, we show the existence of $\{\widetilde{G}(h) : h \in \mathbb{R}\}$ and identify its distribution. Fix $K > 0$.

Note the following decomposition: for $|h| \leq K$,

$$M_n(h) = n^{2/3}\left\{\hat{\theta}_{n,c_b+h/n^{1/3}} - \hat{\theta}_{n,c_b}\right\} = L_n(h) + n^{2/3}\left(\theta_{c_b+h/n^{1/3}} - \theta_{c_b}\right),$$

where $L_n(h)$ is defined as

$$L_n(h) = n^{2/3}\left\{\frac{W^{(1)}_{c_b+h/n^{1/3},n}}{M^{(1)}_{c_b+h/n^{1/3},n}} - \frac{W^{(1)}_{c_b,n}}{M^{(1)}_{c_b,n}} - \frac{E\left[W^{(1)}_{c_b+h/n^{1/3},n}\right]}{E\left[M^{(1)}_{c_b+h/n^{1/3},n}\right]} + \frac{E\left[W^{(1)}_{c_b,n}\right]}{E\left[M^{(1)}_{c_b,n}\right]}\right\}$$
$$- n^{2/3}\left\{\frac{W^{(2)}_{c_b+h/n^{1/3},n}}{M^{(2)}_{c_b+h/n^{1/3},n}} - \frac{W^{(2)}_{c_b,n}}{M^{(2)}_{c_b,n}} - \frac{E\left[W^{(2)}_{c_b+h/n^{1/3},n}\right]}{E\left[M^{(2)}_{c_b+h/n^{1/3},n}\right]} + \frac{E\left[W^{(2)}_{c_b,n}\right]}{E\left[M^{(2)}_{c_b,n}\right]}\right\}.$$

By assumption $(C.3)$, the first and second derivatives, θ'_{c_b} and θ''_{c_b}, of the function $c \mapsto \theta_c$ at c_b exist. Since we assume $\theta_{c_b} > 0$ and c_b is a maximizer as defined in (1.4), due to assumption $(C.3)$, $\theta'_{c_b} = 0$ and $\theta''_{c_b} < 0$. Then by Taylor expansion and condition $(C.3)$,

$$M_n(h) = L_n(h) - \frac{1}{2}h^2\left|\theta''_{c_b}\right| + o_P(1), \tag{S8.10}$$

where $o_P(1)$ is uniform over $h \in [-K, K]$.

We further decompose $L_n(h)$ as $L_n^{(1)}(h) - L_n^{(2)}(h)$ for $|h| \le K$, where for $i \in \{1, 2\}$,

$$L_n^{(i)}(h) = I_n^{(i)}(h) + II_n^{(i)}(h) + III_n^{(i)}(h) + IV_n^{(i)}(h), \tag{S8.11}$$

and

$$I_n^{(i)}(h) = \frac{1}{M^{(i)}_{c_b+h/n^{1/3},n}} \sqrt{n} \left[n^{1/6} \left(W^{(i)}_{c_b+h/n^{1/3},n} - E\left[W^{(i)}_{c_b+h/n^{1/3},n} \right] - W^{(i)}_{c_b,n} + E\left[W^{(i)}_{c_b,n} \right] \right) \right]$$

$$II_n^{(i)}(h) = -\frac{E\left[W^{(i)}_{c_b+h/n^{1/3},n} \right]}{M^{(i)}_{c_b+h/n^{1/3},n} E\left[M^{(i)}_{c_b+h/n^{1/3},n} \right]} \sqrt{n} \left[n^{1/6} \left(M^{(i)}_{c_b+h/n^{1/3},n} - E\left[M^{(i)}_{c_b+h/n^{1/3},n} \right] - M^{(i)}_{c_b,n} + E\left[M^{(i)}_{c_b,n} \right] \right) \right]$$

$$III_n^{(i)}(h) = n^{2/3} \left[W^{(i)}_{c_b,n} - E\left[W^{(i)}_{c_b,n} \right] - \frac{E\left[W^{(i)}_{c_b,n} \right]}{E\left[M^{(i)}_{c_b,n} \right]} \left(M^{(i)}_{c_b,n} - E\left[M^{(i)}_{c_b,n} \right] \right) \right] \left(\frac{1}{M^{(i)}_{c_b+h/n^{1/3},n}} - \frac{1}{M^{(i)}_{c_b,n}} \right)$$

$$IV_n^{(i)}(h) = -\frac{\sqrt{n} \left(M^{(i)}_{c_b,n} - E\left[M^{(i)}_{c_b,n} \right] \right)}{M^{(i)}_{c_b+h/n^{1/3},n}} n^{1/6} \left(\frac{E\left[W^{(i)}_{c_b+h/n^{1/3},n} \right]}{E\left[M^{(i)}_{c_b+h/n^{1/3},n} \right]} - \frac{E\left[W^{(i)}_{c_b,n} \right]}{E\left[M^{(i)}_{c_b,n} \right]} \right).$$

By Lemma 8, we have that

$$L_n(h) = I_n^{(1)}(h) + II_n^{(1)}(h) - I_n^{(2)}(h) - II_n^{(2)}(h) + o_P(1), \tag{S8.12}$$

where $o_P(1)$ is uniform over $h \in [-K, K]$. For $n \geq 1$ and $h \in \mathbb{R}$, define

$$\mathbb{Z}_n(h) := \begin{Bmatrix} n^{2/3} \left(W^{(1)}_{c_b+h/n^{1/3},n} - E\left[W^{(1)}_{c_b+h/n^{1/3},n} \right] - W^{(1)}_{c_b,n} + E\left[W^{(1)}_{c_b,n} \right] \right) \\ n^{2/3} \left(M^{(1)}_{c_b+h/n^{1/3},n} - E\left[M^{(1)}_{c_b+h/n^{1/3},n} \right] - M^{(1)}_{c_b,n} + E\left[M^{(1)}_{c_b,n} \right] \right) \\ n^{2/3} \left(W^{(2)}_{c_b+h/n^{1/3},n} - E\left[W^{(2)}_{c_b+h/n^{1/3},n} \right] - W^{(2)}_{c_b,n} + E\left[W^{(2)}_{c_b,n} \right] \right) \\ n^{2/3} \left(M^{(2)}_{c_b+h/n^{1/3},n} - E\left[M^{(2)}_{c_b+h/n^{1/3},n} \right] - M^{(2)}_{c_b,n} + E\left[M^{(2)}_{c_b,n} \right] \right) \end{Bmatrix}. \tag{S8.13}$$

By Lemma 3, there exists a tight, zero-mean Gaussian process $\{\widetilde{G}_1(h), \widetilde{G}_2(h), \widetilde{G}_3(h), \widetilde{G}_4(h) : |h| \leq K\}$ with covariance function $\{\gamma_{s,t}\mathcal{V} : s, t \in [-K, K]\}$ (see Equations

(S8.18) and (S8.25)) such that, in $\ell^\infty([-K,K]\times[4])$,

$$\{\mathbb{Z}_n(h) : |h| \le K\} \rightsquigarrow \left\{ \left(\widetilde{G}_1(h), \widetilde{G}_2(h), \widetilde{G}_3(h), \widetilde{G}_4(h)\right)^\top : |h| \le K \right\}.$$

Also, by the law of large numbers for U processes (Peña and Giné, 1999, Corollary 5.2.3) and due to condition (*C.2*), in $\ell^\infty([-K,K]\times[6])$,

$$\left\{ \begin{matrix} E\left[M^{(1)}_{c_b+h/n^{1/3},n}\right] \\ M^{(1)}_{c_b+h/n^{1/3},n} \\ E\left[W^{(1)}_{c_b+h/n^{1/3},n}\right] \\ E\left[M^{(2)}_{c_b+h/n^{1/3},n}\right] \\ M^{(2)}_{c_b+h/n^{1/3},n} \\ E\left[W^{(2)}_{c_b+h/n^{1/3},n}\right] \end{matrix} : |h| \le K \right\} \rightsquigarrow \left\{ \begin{matrix} E\left[M^{(1)}_{c_b,n}\right] \\ E\left[M^{(1)}_{c_b,n}\right] \\ E\left[W^{(1)}_{c_b,n}\right] \\ E\left[M^{(2)}_{c_b,n}\right] \\ E\left[M^{(2)}_{c_b,n}\right] \\ E\left[W^{(2)}_{c_b,n}\right] \end{matrix} \right\}$$

Then by Slutsky's theorem (Kosorok, 2008, Theorem 7.15) and continuous mapping Theorem (Van der Vaart and Wellner, 1996, Theorem 1.3.6), due to (S8.12) and (S8.10), we obtain that

$$\{M_n(h) : |h| \le K\} \rightsquigarrow \left\{ \widetilde{G}(h) - \frac{1}{2}h^2 \left|\theta''_{c_b}\right| : |h| \le K \right\} \text{ in } \ell^\infty([-K,K]),$$

where $\widetilde{G}(h) := \boldsymbol{\beta}\left(\widetilde{G}_1(h), \widetilde{G}_2(h), \widetilde{G}_3(h), \widetilde{G}_4(h)\right)^\top$, and

$$\boldsymbol{\beta} := \left(\frac{1}{E\left[M^{(1)}_{c_b,n}\right]}, -\frac{E\left[W^{(1)}_{c_b,n}\right]}{E\left[M^{(1)}_{c_b,n}\right]^2}, -\frac{1}{E\left[M^{(2)}_{c_b,n}\right]}, \frac{E\left[W^{(2)}_{c_b,n}\right]}{E\left[M^{(2)}_{c_b,n}\right]^2} \right). \tag{S8.14}$$

Recall the definitions of the scalar $\gamma_{s,t}$ and matrix $\mathcal{V}$ in Equations (S8.18) and (S8.25) from Lemma 3. Then the covariance function of $\{\widetilde{G}_h :$

$|h| \leq K\}$ is given as follows: for $s, t \in [-K, K]$,

$$\mathrm{Cov}(\widetilde{G}_s, \widetilde{G}_t) = \boldsymbol{\beta}\gamma_{s,t}\mathcal{V}\boldsymbol{\beta}^\top = 4\mathbb{1}_{\mathrm{sgn}(s)=\mathrm{sgn}(t)} \min(|s|, |t|) f(c_b)\lambda(1-\lambda)(\boldsymbol{\beta}\mathcal{V}\boldsymbol{\beta}^\top).$$

Thus, if we denote by $\{\mathbb{Z}(h) : |h| \leq K\}$ a standard two-sided Brownian motion with $\mathbb{Z}(0) = 0$ (Van der Vaart and Wellner, 1996), then $\{\widetilde{G}(h) : |h| \leq K\} \overset{\mathrm{d}}{=} \{\nu\mathbb{Z}(h) : |h| \leq K\}$, where $\overset{\mathrm{d}}{=}$ means equal in distribution and

$$\nu := \sqrt{4f(c_b)\lambda(1-\lambda)\boldsymbol{\beta}\mathcal{V}\boldsymbol{\beta}^\top}. \tag{S8.15}$$

Note that $\boldsymbol{\beta}$ and $\mathcal{V}$ are defined in (S8.14) and (S8.25) respectively and $f(c_b)$ is the first derivative of the function $c \mapsto F(c)$ at c_b. Thus in view of (S8.9), we have

$$n^{1/3}(\hat{c}_n - c_b) \ \rightsquigarrow \ \arg\max_{h\in\mathbb{R}} \left(\nu\mathbb{Z}(h) - \frac{1}{2}h^2 \left|\theta''_{c_b}\right|\right).$$

By Problem 5 in Chapter 3.2 of Van der Vaart and Wellner (1996),

$$\arg\max_{h\in\mathbb{R}} \left(\nu\mathbb{Z}(h) - \frac{1}{2}h^2 \left|\theta''_{c_b}\right|\right) \overset{\mathrm{d}}{=} \left(2\nu/\left|\theta''_{c_b}\right|\right)^{2/3} \mathbb{C},$$

where $\mathbb{C} = \arg\max_{h\in\mathbb{R}} \{\mathbb{Z}(h) - h^2\}$ follows the Chernoff's distribution (Groeneboom and Wellner, 2001). The proof is complete. □

For $n \geq 1$ and $h \in \mathbb{R}$, define the following functions on $\mathcal{S}^2$:

$$\widetilde{f}^{(1)}_{n,h} := n^{1/6}\left(w^{(1)}_{c_b+h/n^{1/3}} - w^{(1)}_{c_b}\right), \quad \widetilde{f}^{(2)}_{n,h} := n^{1/6}\left(m^{(1)}_{c_b+h/n^{1/3}} - m^{(1)}_{c_b}\right),$$
$$\widetilde{f}^{(3)}_{n,h} := n^{1/6}\left(w^{(2)}_{c_b+h/n^{1/3}} - w^{(2)}_{c_b}\right), \quad \widetilde{f}^{(4)}_{n,h} := n^{1/6}\left(m^{(2)}_{c_b+h/n^{1/3}} - m^{(2)}_{c_b}\right), \tag{S8.16}$$

where $w_c^{(i)}(\cdot,\cdot)$ and $m_c^{(i)}(\cdot,\cdot)$ for $i=1,2$ are defined in (S5.4). Further, for $n \geq 1$ and $h \in \mathbb{R}$, define the following functions on $\mathcal{S}$:

$$\bar{f}_{n,h}^{(1)} := P\widetilde{f}_{n,h}^{(1)} = n^{1/6}\left(\bar{w}_{c_b+h/n^{1/3}}^{(1)} - \bar{w}_{c_b}^{(1)}\right), \quad \bar{f}_{n,h}^{(2)} := P\widetilde{f}_{n,h}^{(2)} = n^{1/6}\left(\bar{m}_{c_b+h/n^{1/3}}^{(1)} - \bar{m}_{c_b}^{(1)}\right),$$
$$\bar{f}_{n,h}^{(3)} := P\widetilde{f}_{n,h}^{(3)} = n^{1/6}\left(\bar{w}_{c_b+h/n^{1/3}}^{(2)} - \bar{w}_{c_b}^{(2)}\right), \quad \bar{f}_{n,h}^{(4)} := P\widetilde{f}_{n,h}^{(4)} = n^{1/6}\left(\bar{m}_{c_b+h/n^{1/3}}^{(2)} - \bar{m}_{c_b}^{(2)}\right), \tag{S8.17}$$

where $\bar{w}_c^{(i)}(\cdot)$ and $\bar{m}_c^{(i)}(\cdot)$ for $i=1,2$ are defined in (S5.5).

Denote by $f(c_b)$ the first derivative of the function $c \mapsto F(c)$ at c_b, and recall the definition of $\mathbb{Z}_n(h)$ in (S8.13)

Lemma 3. *Assume the conditions of Theorem 6 hold, and fix some $K>0$. There exists a tight, zero-mean Gaussian process $\{\widetilde{G}_1(h), \widetilde{G}_2(h), \widetilde{G}_3(h), \widetilde{G}_4(h) : |h| \leq K\}$ such that, in $\ell^\infty([-K,K]\times[4])$,*

$$\{\mathbb{Z}_n(h) : |h| \leq K\} \rightsquigarrow \left\{\left(\widetilde{G}_1(h), \widetilde{G}_2(h), \widetilde{G}_3(h), \widetilde{G}_4(h)\right)^\top : |h| \leq K\right\},$$

and that the covariance function is given as follows: for $s,t \in [-K,K]$,

$$Cov\left(\left(\widetilde{G}_1(s), \widetilde{G}_2(s), \widetilde{G}_3(s), \widetilde{G}_4(s)\right), \left(\widetilde{G}_1(t), \widetilde{G}_2(t), \widetilde{G}_3(t), \widetilde{G}_4(t)\right)\right) = \gamma_{s,t}\mathcal{V},$$

where $\mathcal{V}$ is a 4×4 matrix defined in (S8.25) *and $\gamma_{s,t}$ is a scalar given by*

$$\gamma_{s,t} := 4\mathbb{1}_{\mathrm{sgn}(s)=\mathrm{sgn}(t)}\min(|s|,|t|)f(c_b)\lambda(1-\lambda). \tag{S8.18}$$

Proof. Recall the definition of $\mathcal{S}$ in (S5.2). For $n \geq 1$, define two functions as follows: for $\boldsymbol{d}_1 = (t_1, u_1, z_1), \boldsymbol{d}_2 = (t_2, u_2, z_2) \in \mathcal{S}^2$

$$F_n(\boldsymbol{d}_1, \boldsymbol{d}_2) := n^{1/6} \left(\mathbb{1}\{|z_1 - c_b| \leq K/n^{1/3}\} + \mathbb{1}\{|z_2 - c_b| \leq K/n^{1/3}\} \right),$$

$$\bar{F}_n(\boldsymbol{d}_1) := n^{1/6} \left(\mathbb{1}\{|z_1 - c_b| \leq K/n^{1/3}\} + F(c_b + K/n^{1/3}) - F(c_b - K/n^{1/3}) \right). \tag{S8.19}$$

Note that $\bar{F}_n = PF_n$. By the definitions in (S8.16) and (S8.17), for any $\boldsymbol{d}_1, \boldsymbol{d}_2 \in \mathcal{S}^2$ and $|h| \leq K$,

$$\max_{1 \leq j \leq 4} \left| \widetilde{f}^{(j)}_{n,h}(\boldsymbol{d}_1, \boldsymbol{d}_2) \right| \leq F_n(\boldsymbol{d}_1, \boldsymbol{d}_2), \quad \max_{1 \leq j \leq 4} \left| \bar{f}^{(j)}_{n,h}(\boldsymbol{d}_1) \right| \leq \bar{F}_n(\boldsymbol{d}_1). \tag{S8.20}$$

Due to condition (*C.2*),

$$\|F_n\|_{P^2,1} = O(n^{-1/3}), \quad \|F_n\|_{P^2,2} = O(1), \quad \|\bar{F}_n\|_{P,1} = O(n^{-1/3}), \quad \|\bar{F}_n\|_{P,2} = O(1). \tag{S8.21}$$

Further, note that

$$\mathbb{Z}_n(h) = \sqrt{n} \left(U_n^{(2)} \left(\widetilde{f}^{(1)}_{n,h} - P^2 \widetilde{f}^{(1)}_{n,h} \right), \ \ldots, U_n^{(2)} \left(\widetilde{f}^{(4)}_{n,h} - P^2 \widetilde{f}^{(4)}_{n,h} \right) \right)^\top,$$

where we recall the notation for U-processes in (S5.1).

By similar arguments as in Lemma 2, in particular, using (Chen and Kato, 2019, Corollary 5.6), we have that the difference between $\{\mathbb{Z}_n(h) : |h| \leq K\}$ and its Hajek projection process is uniformly controlled, that is,

$$\sup_{|h| \leq K} \left\| \mathbb{Z}_n(h) - 2\sqrt{n} \left(U_n^{(1)} \left(\bar{f}^{(1)}_{n,h} - P \bar{f}^{(1)}_{n,h} \right), \ \ldots, U_n^{(1)} \left(\bar{f}^{(4)}_{n,h} - P \bar{f}^{(4)}_{n,h} \right) \right)^\top \right\| = o_P(1). \tag{S8.22}$$

We now apply Theorem 2.11.22 in Van der Vaart and Wellner (1996) to the empirical process with classes of functions changing with n, that is,

$$\left\{2\sqrt{n}\left(U_n^{(1)}\left(\bar{f}_{n,h}^{(1)} - P\bar{f}_{n,h}^{(1)}\right),\ \ldots, U_n^{(1)}\left(\bar{f}_{n,h}^{(4)} - P\bar{f}_{n,h}^{(4)}\right)\right) : |h| \leq K\right\},$$

for which $\bar{F}_n$ in (S8.19) is an envelope function.

First, as discussed in (S8.21), $P\bar{F}_n^2 = O(1)$. Second, for any $\eta > 0$, we have $\bar{F}_n(\cdot) \leq 2n^{1/6} < \eta\sqrt{n}$ surely for sufficiently large n, which implies that

$$P\left[\bar{F}_n^2 \mathbb{1}\{\bar{F}_n^2 > \eta\sqrt{n}\}\right] \to 0, \text{ as } n \to \infty.$$

Third, let δ_n be an arbitrary sequence of positive numbers which decreases to zero, that is, $\delta_n \downarrow 0$ as $n \to \infty$. Fix any pair $s, t \in [-K, K]$ such that $s < t < s + \delta_n$. Due to the definition of $\bar{f}_{n,h}^{(1)}$ in (S8.17) and $\bar{w}_c^{(1)}(\cdot)$ in (S5.5), we have that for $\boldsymbol{d} = (\tau, u, z) \in \mathcal{S}$,

$$\begin{aligned}\bar{f}_{n,s}^{(1)}(\boldsymbol{d}) - \bar{f}_{n,t}^{(1)}(\boldsymbol{d}) =& \xi_{n,s}(\boldsymbol{d}) n^{1/6} \mathbb{1}\{z \in (c_b + s/n^{1/3}, c_b + t/n^{1/3}]\} \\ &+ n^{1/6}(\xi_{n,s}(\boldsymbol{d}) - \xi_{n,t}(\boldsymbol{d}))\mathbb{1}\{z > c_b + t/n^{1/3}\}.\end{aligned}$$

where, due to condition $(C.3)$, for $h \in \mathbb{R}$,

$$\begin{aligned}\xi_{n,h}(\boldsymbol{d}) :=& \big((1 - G^{(+)}(\tau|c_b + h/n^{1/3}, 1))(1-u)\lambda + G^{(+)}(\tau|c_b + h/n^{1/3}, 0)u(1-\lambda)\big) \\ &\times (1 - F(c_b + h/n^{1/3})),\end{aligned} \tag{S8.23}$$

and recall that $G^{(+)}(t|c, k) = \Pr(T \leq t | Z > c, U = k), k = 0, 1$ is defined in

(2.5). Since for any $h \in \mathbb{R}$, $0 \leq \xi_{n,h}(\cdot) \leq 1$, we have

$$P(\bar{f}^{(1)}_{n,s}(\boldsymbol{\mathcal{D}}_i) - \bar{f}^{(1)}_{n,t}(\boldsymbol{\mathcal{D}}_i))^2 \leq 2n^{1/3}(F(c_b + t/n^{1/3}) - F(c_b + s/n^{1/3})) + 2n^{1/3} P\left(\xi_{n,s} - \xi_{n,t}\right)^2.$$

Thus, due to condition (*C.2*), Taylor's Theorem and Lemma 6, as $n \to \infty$,

$$\sup_{s,t \in [-K,K], |s-t| \leq \delta_n} P(\bar{f}^{(j)}_{n,s} - \bar{f}^{(j)}_{n,t})^2 \to 0,$$

for $j = 1$. Similar arguments show that the above also holds for $j = 2, 3, 4$. Thus we verify the conditions listed in Equation (2.11.21) of Van der Vaart and Wellner (1996).

Furthermore, define for $j = 1, \ldots, 4$,

$$\widetilde{\mathcal{F}}^{(j)}_n := \{\widetilde{f}^{(j)}_{n,h} : |h| \leq K\}, \quad \bar{\mathcal{F}}^{(j)}_n := \{\bar{f}^{(j)}_{n,h} : |h| \leq K\}.$$

By (Van der Vaart and Wellner, 1996, Theorem 2.6.7), for some absolute constants A and v, $\widetilde{\mathcal{F}}^{(j)}_n$, together with the envelope function $\widetilde{F}_n$, is VC-type with characteristic (A, v) for $n \geq 1$ and $j = 1, \ldots, 4$; see (Chen and Kato, 2019, Definition 2.1). Then by (Chen and Kato, 2019, Lemma 5.4), $\bar{\mathcal{F}}^{(j)}_n$, together with the envelope function $\bar{F}_n$, is VC-type with characteristic $(4\sqrt{A}, 2v)$ for $n \geq 1$ and $j = 1, \ldots, 4$. Thus, the uniform entropy condition holds: for $j = 1, \ldots, 4$,

$$\int_0^{\delta_n} \sup_Q \sqrt{\log N(\bar{\mathcal{F}}^{(j)}_n, \|\cdot\|_{Q,2}, \epsilon \|\bar{F}_n\|_{Q,2})} d\epsilon \to 0$$

for every $\delta_n \downarrow 0$, where the supremum is taken over all discretely supported probability measure Q on $\mathcal{S}$.

Finally, for $n \geq 1$ and $s, t \in [-K, K]$, define

$$\Sigma_n(s,t) := 4\text{cov}\left(\left(\bar{f}_{n,s}^{(1)}, \bar{f}_{n,s}^{(2)}, \bar{f}_{n,s}^{(3)}, \bar{f}_{n,s}^{(4)}\right)^\top, \left(\bar{f}_{n,t}^{(1)}, \bar{f}_{n,t}^{(2)}, \bar{f}_{n,t}^{(3)}, \bar{f}_{n,t}^{(4)}\right)^\top \right). \quad \text{(S8.24)}$$

In Lemma 4, we show that for any $s, t \in [-K, K]$, $\lim_{n\to\infty} \Sigma_n(s,t) = \gamma_{s,t}\mathcal{V}$, where $\gamma_{s,t}$ is a scalar defined in (S8.18) and $\mathcal{V}$ is a 4×4 *symmetric* matrix such that

$$\begin{aligned}
\mathcal{V}_{1,1} &= (1 - F(c_b))^2 \left\{ \lambda \int \left(1 - G^{(+)}(x|c_b, 1)\right)^2 \mathcal{L}_{c_b}^{(1)}(dx) + (1-\lambda) \int \left(G^{(+)}(y|c_b, 0)\right)^2 \mathcal{L}_{c_b}^{(2)}(dy) \right\}, \\
\mathcal{V}_{1,2} &= (1 - F(c_b))^2 \left\{ \lambda \int \left(1 - G^{(+)}(x|c_b, 1)\right) \mathcal{L}_{c_b}^{(1)}(dx) + (1-\lambda) \int G^{(+)}(y|c_b, 0) \mathcal{L}_{c_b}^{(2)}(dy) \right\}, \\
\mathcal{V}_{1,3} &= -(1 - F(c_b))F(c_b) \Big\{ \lambda \int \left(1 - G^{(+)}(x|c_b, 1)\right) \left(1 - G^{(-)}(x|c_b, 1)\right) \mathcal{L}_{c_b}^{(1)}(dx) \\
&\qquad + (1-\lambda) \int G^{(+)}(y|c_b, 0) G^{(-)}(y|c_b, 0) \mathcal{L}_{c_b}^{(2)}(dy) \Big\}, \\
\mathcal{V}_{1,4} &= -(1 - F(c_b))F(c_b) \left\{ \lambda \int \left(1 - G^{(+)}(x|c_b, 1)\right) \mathcal{L}_{c_b}^{(1)}(dx) + (1-\lambda) \int G^{(+)}(y|c_b, 0) \mathcal{L}_{c_b}^{(2)}(dy) \right\}, \\
\mathcal{V}_{2,2} &= (1 - F(c_b))^2, \\
\mathcal{V}_{2,3} &= -(1 - F(c_b))F(c_b) \left\{ \lambda \int \left(1 - G^{(-)}(x|c_b, 1)\right) \mathcal{L}_{c_b}^{(1)}(dx) + (1-\lambda) \int G^{(-)}(y|c_b, 0) \mathcal{L}_{c_b}^{(2)}(dy) \right\}, \\
\mathcal{V}_{2,4} &= -(1 - F(c_b))F(c_b), \\
\mathcal{V}_{3,3} &= F^2(c_b) \left\{ \lambda \int \left(1 - G^{(-)}(x|c_b, 1)\right)^2 \mathcal{L}_{c_b}^{(1)}(dx) + (1-\lambda) \int \left(G^{(-)}(y|c_b, 0)\right)^2 \mathcal{L}_{c_b}^{(2)}(dy) \right\}, \\
\mathcal{V}_{3,4} &= F^2(c_b) \left\{ \lambda \int \left(1 - G^{(-)}(x|c_b, 1)\right) \mathcal{L}_{c_b}^{(1)}(dx) + (1-\lambda) \int G^{(-)}(y|c_b, 0) \mathcal{L}_{c_b}^{(2)}(dy) \right\}, \\
\mathcal{V}_{4,4} &= F^2(c_b),
\end{aligned} \quad \text{(S8.25)}$$

Then the proof is complete by (Van der Vaart and Wellner, 1996, Theorem 2.11.22). □

Recall the definition of $\Sigma_n(s,t)$ in (S8.24), the scalar $\gamma_{s,t}$ in (S8.18), and the 4×4 *symmetric* matrix $\mathcal{V}$ in (S8.25).

Lemma 4. *Suppose the conditions in Theorem 6 hold. Fix $s,t \in \mathbb{R}$ and assume $s \leq t$. Then we have*

$$\lim_{n\to\infty} \Sigma_n(s,t) = \gamma_{s,t}\mathcal{V},$$

Proof. We only prove the convergence result for the top left coordinate, noting that the arguments for the other coordinates are similar. Further, we only consider the case $0 \leq s \leq t$, and note that the arguments for the case $s \leq 0 \leq t$ and $s \leq t \leq 0$ are similar.

Specifically, the top left coordinate is

$$\Sigma_{n,1,1}(s,t) := 4\mathrm{Cov}\left(\bar{f}^{(1)}_{n,s},\ \bar{f}^{(1)}_{n,t}\right).$$

Recall the definition of $\xi_{n,h}(\cdot)$ in (S8.23) and define for $h \in \{s,t\}$ and $\boldsymbol{d} = (\tau, u, z) \in \mathcal{S}$,

$$\Delta_{n,h}(\boldsymbol{d}) := \bar{f}^{(1)}_{n,h}(\boldsymbol{d}) - \xi_{n,0}(\boldsymbol{d})n^{1/6}\left(\mathbb{1}\{z > c_b + h/n^{1/3}\} - \mathbb{1}\{z > c_b\}\right). \tag{S8.26}$$

By Lemma 7, we have for $h \in \{s,t\}$, $\|\Delta_{n,h}\|_{P,2} = o(1)$. Due to condition

$(C.2)$, for $h \in \{s,t\}$, we have

$$\left\|n^{1/6}\left(\mathbb{1}\{z > c_b + h/n^{1/3}\} - \mathbb{1}\{z > c_b\}\right)\right\|_{P,1} = O(n^{-1/6}),$$

$$\left\|n^{1/6}\left(\mathbb{1}\{z > c_b + h/n^{1/3}\} - \mathbb{1}\{z > c_b\}\right)\right\|_{P,2} = O(1).$$

Thus, by Cauchy-Schwarz inequality,

$$\begin{aligned}\Sigma_{n,1,1}(s,t) = 4E\Big[&\xi_{n,0}(\mathcal{D})n^{1/6}\left(\mathbb{1}\{Z > c_b + s/n^{1/3}\} - \mathbb{1}\{Z > c_b\}\right)\\ &\xi_{n,0}(\mathcal{D})n^{1/6}\left(\mathbb{1}\{Z > c_b + t/n^{1/3}\} - \mathbb{1}\{Z > c_b\}\right)\Big] + o(1),\end{aligned}$$

where $\mathcal{D} = (T, U, Z)$ is the generic random vector as discussed in Section 2.

Since $0 \le s \le t$, we have

$$\Sigma_{n,1,1}(s,t) = 4n^{1/3}\int_{c_b}^{c_b + \min\{s,t\}/n^{1/3}} f(z)E\left[\xi^2_{n,0}(\mathcal{D})|Z = z\right]dz + o(1),$$

where $f(z) := F'(z)$ exists and is continuous in a small neighbourbood of c_b due to condition $(C.2)$. Note that $\xi^2_{n,0}(\cdot)$ does not depend on n. In Lemma 5, we show that the function $z \mapsto E\left[\xi^2_{n,0}(\mathcal{D})|Z = z\right]$ is continuous at c_b. Thus, we have

$$\Sigma_{n,1,1}(s,t) = 4n^{1/3}\min\{s,t\}f(c_b)E\left[\xi^2_{n,0}(\mathcal{D})|Z = c_b\right] + o(1).$$

Finally, we note that

$$\begin{aligned}E\left[\xi^2_{n,0}(\mathcal{D})|Z = c_b\right] =&(1 - F(c_b))^2\lambda(1-\lambda)\\ &\times\left(\lambda\int(1 - G^{(+)}(x|c_b,1))^2\mathcal{L}^{(1)}_{c_b}(dx) + (1-\lambda)\int(G^{(+)}(y|c_b,1))^2\mathcal{L}^{(2)}_{c_b}(dy)\right),\end{aligned}$$

which implies that for $0 \le s \le t$, $\lim_{n\to\infty} \Sigma_{n,1,1}(s,t) = \gamma_{s,t}\mathcal{V}_{1,1}$. The proof is complete. □

S8.4 Discussions on Theorem 3

The test statistic $\hat{S}_n/\sqrt{n}$ defined in (4.9) is the supremum of the estimators $\hat{\theta}_{n,c}$ for $c \in [\ell, u]$, which is defined in (2.6). The numerators and denominators in (2.6), which appear in the definition of $\hat{\theta}_{n,c}$, $c \in [\ell, u]$, are U-processes. As shown above, these U-processes are Donsker and can be analyzed using classical empirical processes techniques Peña and Giné (1999); Van der Vaart and Wellner (1996). Further, the non-parametric bootstrap is valid for inference as shown in Theorem 4.

Supreme-type statistics have been widely used and studied in related scenarios and broader areas of statistics. For example, in Cai et al. (2010), supreme-type statistics are used to construct a simultaneous confidence band for average treatment differences across a range of scores defined by baseline covariates. Fuentes et al. (2018) utilizes a supreme-type statistic to construct simultaneous confidence intervals for the means of k selected populations, assuming independence and normality with a common variance. In Li et al. (2023b,a), supremum-type statistics are used to test treatment-biomarker interactions with an unknown cutpoint, under linear and generalized-linear frameworks. Supremum-type statistics have also been used extensively in non-parametric statistics; see, e.g., Bickel and Rosenblatt (1973); Einmahl and Mason (2005).

Finally, we should also highlight the recent developments in the distribution approximation and bootstrap of suprema of stochastic processes; see, e.g., Chernozhukov et al. (2013, 2014); Chen and Kato (2019). These works provide powerful tools for dealing with non-Donsker empirical and U-processes. In contrast, we deal with Donsker U-processes and also need to apply the functional delta method, which seems not covered in the references mentioned.

S8.5 Supporting lemmas

Recall the definition of $\xi_{n,h}(\cdot)$ in (S8.23), which is a function on $\mathcal{S}$ in (S5.2). Note that $\xi_{n,0}$ does not depend on n.

Lemma 5. *Suppose the conditions in Theorem 6 hold. Then the function* $z \mapsto E\left[\xi_{n,0}^2(\mathcal{D})|Z=z\right]$ *is continuous at* c_b.

Proof. Note that by definition,

$$E\left[\xi_{n,0}^2(\mathcal{D})|Z=z\right] = (1-F(c_b))^2\,\lambda(1-\lambda)\Big\{\lambda\int\left(1-G^{(+)}(x|c_b,1)\right)^2\mathcal{L}_z^{(1)}(dx) + (1-\lambda)\int\left(G^{(+)}(y|c_b,0)\right)^2\mathcal{L}_z^{(2)}(dy)\Big\}.$$

By condition (*C.3*), as $z \to c_b$, $\mathcal{L}_z^{(1)} \rightsquigarrow \mathcal{L}_{c_b}^{(1)}$, $\mathcal{L}_z^{(2)} \rightsquigarrow \mathcal{L}_{c_b}^{(2)}$. Again by condition (*C.3*), $G^{(+)}(\cdot|c_b, 1)$ and $G^{(-)}(\cdot|c_b, 0)$ are both continuous and bounded functions. As a result, by the definition of weak convergence, as $z \to c_b$,

$$E\left[\xi_{n,0}^2(\mathcal{D})|Z = z\right] \to E\left[\xi_{n,0}^2(\mathcal{D})|Z = c_b\right],$$

which completes the proof. □

Lemma 6. *Suppose the conditions in Theorem 6 hold and fix any $K > 0$. Then*

$$n^{1/3} \sup_{s,t\in[-K,K]} P\left(\xi_{n,s} - \xi_{n,t}\right)^2 = O(n^{-1/3}).$$

Proof. By definition and the triangle inequality, for $\boldsymbol{d} = (\tau, u, z) \in \mathcal{S}$,

$$|\xi_{n,s} - \xi_{n,t}| \leq \sum_{u=0}^{1} \left|G^{(+)}(\tau|c_b + s/n^{1/3}, u)) - G^{(+)}(\tau|c_b + t/n^{1/3}, u))\right| + \left|F(c_b + s/n^{1/3}) - F(c_b + t/n^{1/3})\right|.$$

Due to condition (*C.2*), and by Taylor's theorem,

$$n^{1/3} \sup_{s,t\in[-K,K]} \left(F(c_b + s/n^{1/3}) - F(c_b + t/n^{1/3})\right)^2 = O(n^{-1/3}).$$

Now we fix some $u \in \{0, 1\}$ and assume without loss of generality $-K \leq s < t \leq \min\{s + \delta_n, K\}$. By the definition of $G^{(+)}(\cdot)$ in (2.5), for $|h| \leq K$,

$$G^{(+)}(\tau|c_b + h/n^{1/3}, u) = \frac{\Pr\left(T \leq \tau, Z > c_b + h/n^{1/3}|U = u\right)}{1 - F(c_b + h/n^{1/3})},$$

which implies that

$$\left|G^{(+)}(\tau|c_b + s/n^{1/3}, u) - G^{(+)}(\tau|c_b + t/n^{1/3}, u)\right| \leq \frac{\Pr\left(c_b + s/n^{1/3} < Z \leq c_b + t/n^{1/3}\right)}{1 - F(c_b + s/n^{1/3})} + \frac{F(c_b + t/n^{1/3}) - F(c_b + s/n^{1/3})}{(1 - F(c_b + s/n^{1/3}))(1 - F(c_b + t/n^{1/3}))}.$$

Then the proof is complete again due to condition ($C.2$) and Taylor's theorem. □

Recall the definition of $\Delta_{n,h}(\boldsymbol{d})$ in (S8.26).

Lemma 7. *Suppose the conditions in Theorem 6 hold and fix $h \geq 0$. Then*

$$\|\Delta_{n,h}\|_{P,2} = o(1).$$

Proof. Recall the definition of $\xi_{n,h}$ in (S8.23) and the calculation above this equation. Since $\bar{f}^{(1)}_{n,0}(\cdot) = 0$, we have for $\boldsymbol{d} = (\tau, u, z) \in \mathcal{S}$,

$$\Delta_{n,h}(\boldsymbol{d}) = n^{1/6}(\xi_{n,h}(\boldsymbol{d}) - \xi_{n,0}(\boldsymbol{d}))\mathbb{1}\{z > c_b + h/n^{1/3}\}.$$

As a result, $\|\Delta_{n,h}\|_{P,2} \leq n^{1/3} P(\xi_{n,h} - \xi_{n,0})^2 = o(1)$ due to Lemma 6. □

Recall the definition of $III^{(i)}_n(h)$ and $IV^{(i)}_n(h)$ following equation (S8.11).

Lemma 8. *Suppose the conditions in Theorem 6 hold. For any $K > 0$ and $i \in \{1, 2\}$,*

$$\sup_{|h| \leq K} |III^{(i)}_n(h)| = o_P(1), \quad \sup_{|h| \leq K} |IV^{(i)}_n(h)| = o_P(1).$$

Proof. Fix some $i \in \{0,1\}$. Note that $III_n^{(i)}(h)$ can be decomposed as follows:

$$III_n^{(i)}(h) = -A_{c_b,n}^{(i)} \frac{\widetilde{III_n}^{(i)}(h)}{M_{c_b+h/n^{1/3},n}^{(i)} M_{c_b,n}^{(i)}} - \sqrt{n} A_{c_b,n}^{(i)} \cdot n^{1/6} \frac{E\left[M_{c_b+h/n^{1/3},n}^{(i)}\right] - E\left[M_{c_b,n}^{(i)}\right]}{M_{c_b+h/n^{1/3},n}^{(i)} M_{c_b,n}^{(i)}},$$

where

$$\widetilde{III_n}^{(i)}(h) := n^{2/3}\left(M_{c_b+h/n^{1/3},n}^{(i)} - E\left[M_{c_b+h/n^{1/3},n}^{(i)}\right] - M_{c_b,n}^{(i)} + E\left[M_{c_b,n}^{(i)}\right]\right)$$

$$A_{c_b,n}^{(i)} := W_{c_b,n}^{(i)} - E\left[W_{c_b,n}^{(i)}\right] - \frac{E\left[W_{c_b,n}^{(i)}\right]}{E\left[M_{c_b,n}^{(i)}\right]}\left(M_{c_b,n}^{(i)} - E\left[M_{c_b,n}^{(i)}\right]\right).$$

By the central limit theorem for U-statistics (Van der Vaart, 2007, Theorem 12.3),

$$A_{c_b,n}^{(i)} = o_P(1), \quad \sqrt{n} A_{c_b,n}^{(i)} = O_P(1).$$

By Lemma 3 and Taylor's theorem, due to (*C.2*),

$$\sup_{|h|\le K}\left|\widetilde{III_n}^{(i)}(h)\right| = O_P(1), \quad n^{1/6}\left(E\left[M_{c_b+h/n^{1/3},n}^{(i)}\right] - E\left[M_{c_b,n}^{(i)}\right]\right) = o(1),$$

where $o(1)$ is uniform in $|h| \le K$. Thus, by the law of large numbers for U processes (Peña and Giné, 1999, Corollary 5.2.3) and due to condition (*C.0*), we obtain that $\sup_{|h|\le K} |III_n^{(i)}(h)| = o_P(1)$.

For $IV_n^{(i)}(h)$, by the central limit theorem for U statistics (Van der Vaart, 2007, Theorem 12.3), we have

$$\sqrt{n}\left(M_{c_b,n}^{(i)} - E\left[M_{c_b,n}^{(i)}\right]\right) = O_P(1).$$

By Taylor's theorem and due to (*C.2*),

$$n^{1/6}\left(\frac{E\left[W^{(i)}_{c_b+h/n^{1/3},n}\right]}{E\left[M^{(i)}_{c_b+h/n^{1/3},n}\right]}-\frac{E\left[W^{(i)}_{c_b,n}\right]}{E\left[M^{(i)}_{c_b,n}\right]}\right)=o(1)$$

where $o(1)$ is uniform in $|h|\leq K$. Thus, again by the law of large numbers for U processes (Peña and Giné, 1999, Corollary 5.2.3) and due to condition (*C.0*), we obtain that $\sup_{|h|\leq K}|IV^{(i)}_n(h)|=o_P(1)$. □

S8.6 A Special Case: Finitely Discrete Biomarkers

This section considers a special but common case in subgroup analysis where the biomarker Z is finitely discrete, taking values in a set V with m values: $v_1<v_2<\ldots<v_m$. Assume that $\Pr(Z=v_i)>0$ for $i\in[m]$, and that $\ell<v_1<v_m<u$ for simplicity.

Note that Theorem 3, which concerns testing the hypothesis in (1.3), does not require F to be continuous, and thus it continues to hold when Z is finitely discrete. Next, we focus on the cutpoint estimation problem and define the optimal cutpoint set as

$$B=\{v_i\in V:|\theta_{v_i}|=\kappa^*\},\ \text{where}\ \kappa^*=\max_{v_j\in V}|\theta_{v_j}|.$$

Recall the definition of $\hat{c}_n$ in (4.10). Due to remark 4, $\hat{c}_n$ takes values in V. The next theorem establishes the convergence rate under this framework.

Theorem 2. *Suppose that condition (C.0) holds. Then, there exists a positive constant C, depending on m, λ and F, such that*

$$\Pr(\hat{c}_n \notin B) \leq Ce^{-nC}.$$

Remark 3. Thus, when the biomarker Z is finitely discrete, the probability that $\hat{c}_n$ does not belong to the optimal set decays exponentially with n.

Proof of Theorem 2. Recall that $V = \{v_1, ..., v_m\}$ represents the set of values that the biomarker Z can take. Let $\delta = \kappa^* - \max_{j \notin B} |\theta_{v_j}|$. Assume without loss of generality that $\delta > 0$; otherwise, $B = V$ and $\Pr(\hat{c}_n \notin B) = 0$ directly. Due to assumption (*C.0*), we fix some $\epsilon > 0$ such that

$$\epsilon \leq \min_{c \in V} \{(1-\lambda)(1-F(c)), \lambda(1-F(c)), (1-\lambda)F(c), \lambda F(c)\}.$$

Define

$$\begin{aligned} B_n := \Bigg\{ &\max_{c \in V} \left| \frac{1}{n} \sum_{i=1}^{n} (1-U_i) Z_i^{c+} - (1-\lambda)(1-F(c)) \right| \leq \epsilon/2, \\ &\max_{c \in V} \left| \frac{1}{n} \sum_{i=1}^{n} U_i Z_i^{c+} - \lambda(1-F(c)) \right| \leq \epsilon/2, \\ &\max_{c \in V} \left| \frac{1}{n} \sum_{i=1}^{n} (1-U_i) Z_i^{c-} - (1-\lambda)F(c) \right| \leq \epsilon/2, \\ &\max_{c \in V} \left| \frac{1}{n} \sum_{i=1}^{n} U_i Z_i^{c-} - \lambda F(c) \right| \leq \epsilon/2 \Bigg\}. \end{aligned}$$

Let $\epsilon_1 = \min\{\delta, \epsilon\}$, then for any $j = 1, ..., m$,

$$\begin{aligned}&\Pr(|\hat{\theta}_{v_j} - \theta_{v_j}| > \delta/2) \leq \Pr(|\hat{\theta}_{v_j} - \theta_{v_j}| > \epsilon_1/2)\\ &\leq \Pr(B_n^{\complement}) + \sum_{i=1}^{2} \left\{ \Pr\left(\left| W_{v_j,n}^{(i)} - E[W_{v_j,n}^{(i)}] \right| > \frac{\epsilon_1^5}{256} \right) + \Pr\left(\left| M_{v_j,n}^{(i)} - E[M_{v_j,n}^{(i)}] \right| > \frac{\epsilon_1^5}{256} \right) \right\}.\end{aligned}$$

As $U_i, T_{i,j}, Z_i^{v_j+}$ are bounded by 1, by McDiarmid's inequality (McDiarmid, 1989), there exists a constant $C > 0$ such that, for $j = 1, ..., m$, the above quality can be further bounded as follows:

$$\Pr(|\hat{\theta}_{v_j} - \theta_{v_j}| > \delta/2) \leq Ce^{-n\epsilon_1^{10}/C}.$$

Finally, since

$$\cap_{j=1}^{m} \{|\hat{\theta}_{v_j} - \theta_{v_j}| \leq \delta/2\} \subseteq \{\hat{c}_n \in B\},$$

and by union bound, we have

$$\Pr(\hat{c}_n \notin B) \leq mCe^{-n\epsilon_1^{10}/C}.$$

The proof is complete, as the right-hand side decays exponentially with n. □

Bibliography

Arcones, M. A. and E. Gine (1993). Limit theorems for u-processes. *The Annals of Probability 21.*

Arcones, M. A. and E. Giné (1994). U-processes indexed by vapnik-Červonenkis classes of functions with applications to asymptotics and bootstrap of u-statistics with estimated parameters. *Stochastic Processes and their Applications 52*, 17–38.

Bickel, P. J. and M. Rosenblatt (1973). On some global measures of the deviations of density function estimates. *The Annals of Statistics*, 1071–1095.

Cai, T., L. Tian, P. H. Wong, and L. J. Wei (2010, 09). Analysis of randomized comparative clinical trial data for personalized treatment selections. *Biostatistics 12*, 270–282.

Chen, X. and K. Kato (2019). Jackknife multiplier bootstrap: finite sample approximations to the u-process supremum with applications. *Probability Theory and Related Fields 176*, 1097–1163.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018, 01). Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal 21*, C1–C68.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014). Gaussian approximation of suprema of empirical processes. *The Annals of Statistics 42*.

Chernozhukov, V., S. Lee, and A. M. Rosen (2013). Intersection bounds: Estimation and inference. *Econometrica 81*, 667–737.

Davydov, Y. A., M. A. Lifshits, and N. V. Smorodina (1998). *Local Properties of Distributions of Stochastic Functionals*. American Mathematical Soc.

De Schryver, M. and J. De Neve (2019, 08). A tutorial on probabilistic index models: Regression models for the effect size p(y1 ¡ y2). *Psychological Methods 24*, 403–418.

Einmahl, U. and D. M. Mason (2005). Uniform in bandwidth consistency of kernel-type function estimators.

Fan, A., R. Song, and W. Lu (2017). Change-plane analysis for subgroup detection and sample size calculation. *Journal of the American Statistical Association 112*(518), 769–778.

Fokianos, K., J. Qin, B. Kedem, and D. A. Short (2001). A Semiparametric Approach to the One-Way Layout. *Technometrics 43*(1), 56–65.

Fokianos, K. and J. F. Troendle (2007). Inference for the relative treatment effect with the density ratio model. *Statistical Modelling 7*, 155–173.

Fuentes, C., G. Casella, and M. T. Wells (2018, 01). Confidence intervals for the means of the selected populations. *Electronic Journal of Statistics 12*.

Giné, E. and R. Nickl (2021). *Mathematical foundations of infinite-dimensional statistical models.* Cambridge university press.

Groeneboom, P. and J. A. Wellner (2001). Computing chernoff's distribution. *Journal of Computational and Graphical Statistics 10*, 388–400.

Jiang, S., B. E. Chen, and D. Tu (2016). Inference on treatment-covariate interaction based on a nonparametric measure of treatment effects and censored survival data. *Statistics in Medicine 35*, 2715–2725.

Jiang, S. and D. Tu (2012). Inference on the probability as a measurement of treatment effect under a density ratio model and random censoring. *Computational Statistics and Data Analysis 56*, 1069–1078.

Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference.* Springer New York.

Ledoux, M. and M. Talagrand (1991). *Probability in Banach Spaces: Isoperimetry and Processes.* Springer.

Li, J., Y. Li, B. Jin, and M. R. Kosorok (2021). Multithreshold change plane model: Estimation theory and applications in subgroup identification. *Statistics in medicine 40*(15), 3440–3459.

Li, N., Y. Song, C. D. Lin, and D. Tu (2023a). Bootstrap adjusted predictive classification for identification of subgroups with differential treatment effects under generalized linear models. *Electronic Journal of Statistics 17*(1), 548 – 606.

Li, N., Y. Song, D. Lin, and D. Tu (2023b). Bootstrap adjustment to minimum p-value method for predictive classification. *Statistica Sinica 33*, 2065–2086.

McDiarmid, C. (1989). *On the method of bounded differences*, pp. 148–188. London Mathematical Society Lecture Note Series. Cambridge University Press.

Mi, X., P. Tighe, F. Zou, and B. Zou (2021). A deep learning semiparametric regression for adjusting complex confounding structures. *The Annals of Applied Statistics 15*, 1086–1100.

Nolan, D. and D. Pollard (1987). U-processes: Rates of convergence. *The Annals of Statistics 15*, 780–799.

Peña, V. H. and E. Giné (1999). *Decoupling: From Dependence to Independence*. Springer.

Qin, J. and B. Zhang (1997). A goodness-of-fit test for logistic regression models based on case-control data. *Biometrika 84*(3), 609–618.

Thas, O., J. D. Neve, L. Clement, and J.-P. Ottoy (2012, 07). Probabilistic index models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 74*, 623–671.

Van der Vaart, A. W. (2007). *Asymptotic statistics*. Cambridge University Press.

Van der Vaart, A. W. and J. A. Wellner (1996). *Weak convergence and empirical processes: With Applications to Statistics*. Springer New York.