Electronic Journal of Statistics
Vol. 17 (2023) 548–606
ISSN: 1935-7524
https://doi.org/10.1214/23-EJS2108

# Bootstrap adjusted predictive classification for identification of subgroups with differential treatment effects under generalized linear models*

**Na Li, Yanglei Song and C. Devon Lin**

*Department of Mathematics and Statistics,*
*Queen's University,*
*Kingston, ON, K7L 3N6, Canada*
*e-mail:* na.li@queensu.ca*;* yanglei.song@queensu.ca*;* devon.lin@queensu.ca

**Dongsheng Tu**†

*Canadian Cancer Trials Group,*
*Departments of Public Health Sciences & Mathematics and Statistics,*
*Queen's University,*
*Kingston, ON, K7L 3N6, Canada*
*e-mail:* dtu@ctg.queensu.ca

**Abstract:** Predictive classification considered in this paper concerns the problem of identifying subgroups based on a continuous biomarker through estimation of an unknown cutpoint and assessing whether these subgroups differ in treatment effect relative to some clinical outcome. The problem is considered under a generalized linear model framework for clinical outcomes and formulated as testing the significance of the interaction between the treatment and the subgroup indicator. When the main effect of the subgroup indicator does not exist, the cutpoint is non-identifiable under the null. Existing procedures are not adaptive to the identifiability issue, and do not work well when the main effect is small. In this work, we propose profile score-type and Wald-type test statistics, and further $m$-out-of-$n$ bootstrap techniques to obtain their critical values. The proposed procedures do not rely on the knowledge about the model identifiability, and we establish their asymptotic size validity and study the power under local alternatives in both cases. Further, we show that the standard bootstrap is inconsistent for the non-identifiable case. Simulation results corroborate our theory, and the proposed method is applied to a dataset from a clinical trial on advanced colorectal cancer.

**MSC2020 subject classifications:** Primary 60J12, 62F40; secondary 62H30, 62F05, 62P10.

**Keywords and phrases:** Generalized linear models, $m$-out-of-$n$ bootstrap, nonstandard asymptotics, predictive classification.

Received January 2022.

*This research is supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) and enabled in part by support provided by Compute Canada (www.computecanada.ca).

†Corresponding author.

## 1. Introduction

In clinical trials testing a new treatment against a control, it is a common practice to classify patients into two subgroups based on a continuous biomarker, such as the expression level of a gene or the result from a blood test, so that the identified subgroups have significantly different treatment effects with respect to certain clinical outcome. This problem is known as predictive classification, and the biomarker that induces the subgroups is called a predictive biomarker [5]. Identification and assessment of predictive biomarkers are very active areas of medical research, especially in the current era of personalized medicine [45, 34].

In many studies, the clinical outcome of interest is binary. For example, in cancer clinical trials, one important outcome is whether a patient has responded to or received a clinical benefit from a treatment based on a specific criterion [35]. This motivates us to investigate the problem of predictive classification under generalized linear models (GLMs) for clinical outcomes, which include the binary outcome as a special case. Specifically, for a patient, denote $Y$ as a clinical outcome of interest, $U$ a binary treatment indicator which equals 1 if the patient received an experimental treatment and 0 if a control treatment, and $X$ a continuous biomarker which is used to classify the patient into subgroups based on an *unknown* cutpoint $c_0$. There may be additional covariates $\boldsymbol{W}$ of length $d$ observed. We assume that conditional on $(\boldsymbol{W}, U, X) = (\boldsymbol{w}, u, x)$, the density function of $Y = y$, relative to some $\sigma$-finite measure $\nu$ on $\mathbb{R}$, belongs to the exponential family and is given as follows:

$$\exp(y(\boldsymbol{\eta}_0^T \boldsymbol{z}_{c_0} + \lambda_0 u x_{c_0}) - \phi(\boldsymbol{\eta}_0^T \boldsymbol{z}_{c_0} + \lambda_0 u x_{c_0})), \tag{1.1}$$

where $\phi$ is a given strictly convex function on $\mathbb{R}$, $\boldsymbol{z}_c = (\boldsymbol{w}^T, u, x_c)^T$ with $x_c = I(x \leq c)$ for $c \in \mathbb{R}$, $c_0 \in [\ell, u]$ is the *unknown* cutpoint, and $\boldsymbol{\eta}_0 = (\boldsymbol{\alpha}_0^T, \beta_0, \gamma_0)^T$ and $\lambda_0$ are regression parameters. Then the conditional expectation $\mathrm{E}(Y \mid \boldsymbol{W}, U, X)$ of $Y$ given $(\boldsymbol{W}, U, X)$ satisfies the following GLM:

$$g(\mathrm{E}(Y \mid \boldsymbol{W}, U, X)) = \boldsymbol{\alpha}_0^T \boldsymbol{W} + \beta_0 U + \gamma_0 X_{c_0} + \lambda_0 U X_{c_0}, \tag{1.2}$$

where $g(\cdot) = (\phi')^{-1}(\cdot)$ is a link function. When $g$ is the logistic function, model (1.2) is the popular logistic regression model for the binary responses.

Model (1.2) implies that the treatment effect is respectively $\beta_0 + \lambda_0$ and $\beta_0$ for patients in the subgroup with $X \leq c_0$ and $X > c_0$. Therefore, parameter $\lambda_0$ measures the differential treatment effect between the two subgroups defined by the *unknown* cutpoint $c_0$. Our goal is to test whether the difference is significant, i.e., testing $H_0 : \lambda_0 = 0$.

Given a sample $\{(Y_i, \boldsymbol{W}_i, U_i, X_i), i \in [n] := \{1, \ldots, n\}\}$ of size $n$, if the cutpoint $c_0$ was known to take the value $c$, then one can use the usual score test statistic, denoted as $S_{n,c}$, for testing $H_0 : \lambda_0 = 0$, which, under the null, converges in distribution to the zero mean normal distribution with variance $\sigma^2_{c,\boldsymbol{\eta}_0}$, denoted as $N(0, \sigma^2_{c,\boldsymbol{\eta}_0})$ [31]. Note that the discussions below apply similarly to other tests such as Wald, and we focus on score tests for concreteness. Since $c_0$ is unknown, one may replace it by its estimate, which however is only possible when $c_0$ is *identifiable*. Specifically, if $\gamma_0 = 0$, then under $H_0 : \lambda_0 = 0$, the cutpoint $c_0$ is *non-identifiable*, in the sense that different values of $c_0$ induce the same distribution on the response $Y$. We discuss below procedures for each case.

If it is known *a priori* that the cupoint is identifiable, i.e., $\gamma_0 \neq 0$, we may estimate $c_0$ by the profile maximum likelihood estimator (MLE), $\hat{c}_n$, and act as if $\hat{c}_n$ is its true value. Specifically, denote by $L_n(c, \boldsymbol{\eta}, \lambda)$ the likelihood function under the model (1.1). Let $\hat{\boldsymbol{\eta}}_{n,c}$ be the maximizer of $L_n(c, \boldsymbol{\eta}, 0)$ over $\boldsymbol{\eta} \in \mathbb{R}^{d+2}$ for each fixed $c$, and $\hat{c}_n$ the *smallest* maximizer of $L_n(c, \hat{\boldsymbol{\eta}}_{n,c}, 0)$ over $c \in [\ell, u]$. Then we may use $S_n := S_{n,\hat{c}_n}$ as a profile score-type statistic for testing $H_0 : \lambda_0 = 0$. Since $c_0$ is identifiable, i.e., $\gamma_0 \neq 0$, it can be shown that $\hat{c}_n$ converges to $c_0$ at the rate $n$ under $H_0$ by similar approaches used in, for example, [22, 36, 29, 44, 43, 25, 33], under some closely related models. As a result, $S_n$ converges in distribution to $N(0, \sigma^2_{c_0,\boldsymbol{\eta}_0})$, the same limit as for $S_{n,c_0}$, and the variance $\sigma^2_{c_0,\boldsymbol{\eta}_0}$ can be consistently estimated by replacing $c_0$ and $\boldsymbol{\eta}_0$ by $\hat{c}_n$ and $\hat{\boldsymbol{\eta}}_n$ respectively, where $\hat{\boldsymbol{\eta}}_n := \hat{\boldsymbol{\eta}}_{n,\hat{c}_n}$. See Section 2 for the precise statements.

On the other hand, if it is known *a priori* that the cupoint is non-identifiable, i.e., $\gamma_0 = 0$, the so-called minimum $p$-value method may be used, which is popular in practice [20, 9]. Specifically, for each $c \in [\ell, u]$, if $c_0$ assumes the value $c$, then $S_{n,c}/\sigma_{c,\hat{\boldsymbol{\eta}}_{n,c}}$ converges in distribution to $N(0, 1)$ [31]. The minimum $p$-value method estimates $c_0$ by minimizing the $p$-value or equivalently maximizing the absolute value of the test statistic, and uses the associated $p$-value to test $H_0 : \lambda_0 = 0$, i.e.,

$$p_{n,mp} = 2\{1 - \Phi(M_n)\}, \quad \text{where} \quad M_n = \sup_{c\in[\ell,u]} |S_{n,c}/\sigma_{c,\hat{\boldsymbol{\eta}}_{n,c}}|, \tag{1.3}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Although, without adjustment, $p_{n,mp}$ suffers from the problem of type I error inflation, under *the assumption that* $\gamma_0 = 0$, one may obtain the critical value for $M_n$ from its limiting distribution, as in [17], which considers a general regression model for a continuous outcome. Such approaches are also widely used in a related problem known as prognostic classifications [32, 18, 23, 30, 40], which tests $H_0' : \gamma_0 = 0$ under the assumption $\lambda_0 = 0$ in model (1.2). Finally, we note that the non-identifiable case is non-standard in the sense that $c_0$ is a nuisance parameter under the null [12, 13, 2].

The aforementioned works require the additional assumption regarding the identifiability of the model. However, in practice, there usually is no convincing justification for one case or the other; further, a valid procedure under the identifiability condition may control the size poorly if the model is close to being non-identifiable, i.e., $|\gamma_0|$ being small, and vice versa. Thus it is important to develop a procedure that is *adaptive* to the identifiability issue, and that is valid in *both* cases. In [27], we have shown that for regression models with a continuous outcome, the minimum $p$-value approach as in (1.3), with the score test statistic being replaced by Wald, turns out to work for both cases under both the random and fixed designs. However, the arguments therein rely on the fact that the link function in (1.2) is linear. For general GLMs, $M_n$ in (1.3) would diverge at the rate $\sqrt{n}$, and its failure is also apparent from simulation studies in Section 5.

In this work, we propose to use the profile score-type or Wald-type statistic for testing $H_0 : \lambda_0 = 0$, and obtain its critical value by an $m$-out-of-$n$ bootstrap. For concreteness, here we focus on the score-type, i.e., $S_n$. Specifically, we first sample *with replacement* $m_n \leq n$ pairs of *covariates* $(\boldsymbol{W}_i^*, U_i^*, X_i^*),\ i \in [m_n]$, and then for each $i \in [m_n]$, generate $Y_i^*$ from the density (1.1) with $\lambda_0 = 0$ and $\boldsymbol{\eta}_0, c_0$ replaced by $\hat{\boldsymbol{\eta}}_n, \hat{c}_n$, respectively. Finally, we obtain the critical value for $S_n$ from the bootstrap distribution of the score-type statistic, $S_n^*$, based on the bootstrap samples. The proposed procedure does not rely on the knowledge of identifiability, and we establish its asymptotic size validity in both cases as long as the bootstrap sample size is of a smaller order compared to the original sample size, i.e., $m_n/n \to 0$.

Further, in the *identifiable* case, we show that the condition $m_n/n \to 0$ can be dropped, and that the standard bootstrap, i.e., $m_n = n$, provides asymptotically correct critical values for $S_n$. This is interesting because by [36, 44, 43] the standard bootstrap is not asymptotically consistent for constructing confidence intervals for the MLE $\hat{c}_n$ of the cutpoint $c_0$, but we show that it is in fact so for the MLE $\hat{\boldsymbol{\eta}}_n$ of the regular parameter $\boldsymbol{\eta}_0$, and also for $S_n$. In the *non-identifiable* case, we prove that the standard bootstrap is inconsistent, in the sense that the bootstrap distribution does not converge weakly to the limiting distribution of the test statistic, in probability.

Finally, we study the rejection probabilities, i.e. power, of the proposed procedure under local alternatives, $H_{1,n} : \lambda_0 = B_0/\sqrt{n}$, for some fixed constant $B_0 \neq 0$. In the identifiable case, the asymptotic power is the same as if the unknown cutpoint $c_0$ was known. In the non-identifiable case, the form of power is more complicated, but it tends to one as $|B_0|$ approaches $\infty$.

In terms of the literature, the $m$-out-of-$n$ bootstrap, which usually generates bootstrap samples of size $m = o(n)$, is proposed by [7] as a modification to the standard bootstrap techniques using $m = n$, which are reliable for regular models [6, 15, 38], but may fail in cases such as non-smooth estimation problems, estimators with a cube-root convergence rate, or models with unknown nuisance parameters; see [24, 36, 10, 37, 1] and the references therein. The asymptotic validity of $m$-out-of-$n$ bootstrap techniques have been established under several non-standard models. For example, [36, 44] show the inconsistency of the standard bootstrap and the consistency of the $m$-out-of-$n$ bootstrap in constructing confidence intervals for the cutpoint parameter under regression models with continuous responses; in addition, [43] establishes similar results for the Cox model. Further, [24] proves the consistency of the $m$-out-of-$n$ bootstrap for estimating the limiting distribution of non-nuisance parameters under the M-estimation framework, provided that the estimators of the nuisance parameters enjoy a faster convergence rate. To the best of our knowledge, for the GLM in (1.1), due to the identifiability issue, the validity of the proposed procedures does not follow directly from previous works.

The reminder of the paper is organized as follows. In Section 2 we introduce the score-type and Wald-type test statistics, and derive the limiting distributions of $S_n$ in the identifiable and non-identifiable cases, respectively. In Section 3, we propose an $m$-out-of-$n$ bootstrap procedure for obtaining the critical value for $S_n$, establish the asymptotic size validity and study its power under local alternatives; further, we show the inconsistency of standard bootstrap for the non-identifiable case. In Section 4 we propose an $m$-out-of-$n$ bootstrap for the Wald-type test statistic. We present in Section 5 simulation studies to evaluate the finite-sample performance of the proposed methods, and in Section 6 an application to an advanced colorectal cancer dataset. We conclude in Section 7 and present proofs in Appendix.

## 2. Profile estimates and test statistics

Our primary objective is to test $H_0 : \lambda_0 = 0$ under model (1.2) based on $\boldsymbol{\mathcal{D}}_i = (Y_i, \boldsymbol{W}_i, U_i, X_i),\ i \in [n]$, which are independently and identically distributed observations with the same distribution as $(Y, \boldsymbol{W}, U, X)$, where random variables are defined on some probability space $(\Omega, \mathcal{G}, \mathrm{pr})$. Under the assumption that the response $Y$ is from the exponential family defined by (1.1), the log-likelihood function of parameters $c \in [\ell, u]$, $\boldsymbol{\eta} = (\boldsymbol{\alpha}^T, \beta, \gamma)^T \in \mathbb{R}^{d+2}$, and $\lambda \in \mathbb{R}$ can be written as

$$\begin{aligned} L_n(c, \boldsymbol{\eta}, \lambda) &:= n^{-1} \sum_{i=1}^{n} \varphi_{c,\boldsymbol{\eta},\lambda}(\boldsymbol{\mathcal{D}}_i), \qquad \text{where} \\ \varphi_{c,\boldsymbol{\eta},\lambda}(y, \boldsymbol{w}, u, x) &:= y\left(\boldsymbol{\eta}^T \boldsymbol{z}_c + \lambda u x_c\right) - \phi\left(\boldsymbol{\eta}^T \boldsymbol{z}_c + \lambda u x_c\right), \end{aligned} \tag{2.1}$$

and $\boldsymbol{z}_c = (\boldsymbol{w}^T, u, x_c)^T$ with $x_c = I(x \le c)$.

Suppose for now the value of the cutpoint $c_0$ is known to be $c$. Then, based on the likelihood function (2.1), the score test statistic for $H_0 : \lambda_0 = 0$ is:

$$S_{n,c} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i X_{i,c}(Y_i - \phi'(\hat{\boldsymbol{\eta}}_{n,c}^T \boldsymbol{Z}_{i,c})), \tag{2.2}$$

where $\hat{\boldsymbol{\eta}}_{n,c} := \operatorname{argmax}_{\boldsymbol{\eta} \in \mathbb{R}^{d+2}} L_n(c, \boldsymbol{\eta}, 0)$ is the MLE of $\boldsymbol{\eta}_0$ for a given $c$ under the null $H_0 : \lambda_0 = 0$.

Since the cutpoint $c_0$ is in fact unknown, in view of the discussion in Section 1, we replace it by the profile estimate $\hat{c}_n$ under the null, and use its associated score test statistic, $S_n$, for testing $H_0 : \lambda_0 = 0$, i.e.,

$$\begin{aligned} \hat{c}_n &:= \operatorname*{sargmax}_{c \in [\ell, u]} L_n(c, \hat{\boldsymbol{\eta}}_{n,c}, 0), \qquad \hat{\boldsymbol{\eta}}_n := \hat{\boldsymbol{\eta}}_{n,\hat{c}_n}, \\ S_n &:= S_{n,\hat{c}_n} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i X_{i,\hat{c}_n}(Y_i - \phi'(\hat{\boldsymbol{\eta}}_n^T \boldsymbol{Z}_{i,\hat{c}_n})), \end{aligned} \tag{2.3}$$

where "sargmax" denotes the maximizer corresponding to the smallest $c$; see a precise definition in [36, Page 4]. Note that $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n)$ are the joint MLE for $(c_0, \boldsymbol{\eta}_0)$ under the null $H_0 : \lambda_0 = 0$.

Another approach for testing $H_0 : \lambda_0 = 0$ is to use the Wald-type test statistic. Similarly, if the value of the cutpoint $c_0$ was known to be $c$, we can estimate $(\boldsymbol{\eta}_0, \lambda_0)$ by its MLE $(\tilde{\boldsymbol{\eta}}_{n,c}, \tilde{\lambda}_{n,c}) := \operatorname{argmax}_{(\boldsymbol{\eta},\lambda)\in\mathbb{R}^{d+3}} L_n(c, \boldsymbol{\eta}, \lambda)$, and use $W_{n,c} := \sqrt{n}\tilde{\lambda}_{n,c}$ for testing $H_0 : \lambda_0 = 0$. Now since $c_0$ is unknown, we replace it by a profile estimator $\tilde{c}_n$, and use its associated Wald test statistic, $W_n$, for testing $H_0 : \lambda_0 = 0$, i.e.,

$$\tilde{c}_n := \operatorname*{sargmax}_{c\in[\ell,u]} L_n(c, \tilde{\boldsymbol{\eta}}_{n,c}, \tilde{\lambda}_{n,c}), \quad \tilde{\lambda}_n := \tilde{\lambda}_{\tilde{c}_n}, \qquad W_n := W_{n,\tilde{c}_n}. \tag{2.4}$$

Because of the similarity in the theoretical development for the score-type and Wald-type test statistics, we only present the details below for the score-type test statistic.

**Remark 2.1.** *In practice, to find the "sargmax" over $c \in [\ell, u]$ in (2.3) and (2.4), it suffices to consider those $c \in \{X_i : i \in [n]\} \cap [\ell, u]$.*

### *2.1. Limiting distributions of the score-type test statistic*

Define $F_i(x) := \mathrm{pr}(X \le x|U = i)$ for $i \in \{0, 1\}$ and $x \in \mathbb{R}$. For a square matrix $\boldsymbol{A}$, denote by $\lambda_{\min}(\boldsymbol{A})$ its smallest eigenvalue. We make the following assumption regarding the *covariates* in model (1.2).

(C.I) $0 < \mathrm{E}[U] < 1$; For $i \in \{0, 1\}$, $F_i$ is continuous on $[\ell, u]$, continuously differentiable with positive derivative in a neighbourhood of $c_0$, and $0 < F_i(\ell) < F_i(u) < 1$; $\|\boldsymbol{W}\| \le C_w$ for some positive constant $C_w$, where $\|\cdot\|$ is the Euclidean norm; $\lambda_{\min}(\mathrm{E}[\boldsymbol{W}\boldsymbol{W}^T|U = i, X > u]) > 0$ and $\lambda_{\min}(\mathrm{E}[\boldsymbol{W}\boldsymbol{W}^T|U = i, X \le \ell]) > 0$.

**Remark 2.2.** *We assume that the additional covariates $\boldsymbol{W}$ are bounded, which simplifies significantly our presentation and proof. Other assumptions are mainly to exclude degenerate situations; in particular, the last two conditions assume that $\boldsymbol{W}$ are (conditionally on $U$ and $X$) not collinear.*

First, we consider the identifiable case. Let $\mathbb{Z}_{\boldsymbol{\eta}} := N(0, (\boldsymbol{V}^{(1)}_{c_0,\boldsymbol{\eta}_0})^{-1})$ be a normal random vector of length $d + 2$, independent of another normal random variable $\mathbb{Z}_S := N(0, \sigma^2_{c_0,\boldsymbol{\eta}_0})$, where for any $(c, \boldsymbol{\eta}) \in [\ell, u] \times \mathbb{R}^{d+2}$,

$$\begin{aligned} \boldsymbol{V}^{(1)}_{c,\boldsymbol{\eta}} &:= \mathrm{E}[\phi''(\boldsymbol{\eta}^T\boldsymbol{Z}_c)\boldsymbol{Z}_c\boldsymbol{Z}_c^T], \quad \sigma^2_{c,\boldsymbol{\eta}} := V^{(3)}_{c,\boldsymbol{\eta}} - \boldsymbol{V}^{(2)}_{c,\boldsymbol{\eta}}(\boldsymbol{V}^{(1)}_{c,\boldsymbol{\eta}})^{-1}(\boldsymbol{V}^{(2)}_{c,\boldsymbol{\eta}})^T, \\ \text{with} \quad \boldsymbol{V}^{(2)}_{c,\boldsymbol{\eta}} &:= \mathrm{E}[\phi''(\boldsymbol{\eta}^T\boldsymbol{Z}_c)UX_c\boldsymbol{Z}_c^T], \quad V^{(3)}_{c,\boldsymbol{\eta}} := \mathrm{E}[\phi''(\boldsymbol{\eta}^T\boldsymbol{Z}_c)UX_c]. \end{aligned} \tag{2.5}$$

Define $\Theta_+ := \boldsymbol{\alpha}_0^T\boldsymbol{W} + \beta_0 U + \gamma_0$, and $\Theta_- := \boldsymbol{\alpha}_0^T\boldsymbol{W} + \beta_0 U$. Let $Y_+$ and $Y_-$ be two random variables such that their conditional $\nu$-density, given $(\boldsymbol{W}, U, X)$, are respectively $\exp(y_+\Theta_+ - \phi(\Theta_+))$ and $\exp(y_-\Theta_- - \phi(\Theta_-))$. Further, consider a sequence of independent and identically distributed pairs of random variables, $\{(\xi_{n,+}, \xi_{n,-}) : n \in \mathbb{N}\}$, such that $\xi_{1,+}$ and $\xi_{1,-}$ are independent, and

$$\xi_{1,+} \overset{\mathrm{d}}{=} \gamma_0 Y_+ - (\phi(\Theta_+) - \phi(\Theta_-)) \text{ given } X = c_0,$$

$$\xi_{1,-} \quad \overset{\mathrm{d}}{=} \quad -\gamma_0 Y_- + (\phi(\Theta_+) - \phi(\Theta_-)) \text{ given } X = c_0,$$

where $\overset{\mathrm{d}}{=}$ means that two sides have the same distribution. Let $N_+(\cdot)$ and $N_-(\cdot)$ be two Poisson processes with intensity $F'_X(c_0) > 0$, where $F_X(\cdot)$ is the cumulative distribution function for $X$. In addition, we assume $N_+(\cdot)$, $N_-(\cdot)$, $\{(\xi_{n,+}, \xi_{n,-}) : n \in \mathbb{N}\}$ and $\mathbb{Z}_{\boldsymbol{\eta}}, \mathbb{Z}_S$ are all independent. Finally, define

$$\mathbb{Z}_c := \operatorname*{sargmax}_{t\in\mathbb{R}} \mathbb{D}(t), \quad \text{where} \quad \mathbb{D}(t) := \begin{cases} \sum_{i=1}^{N_+(t)} \xi_{i,+} & \text{if } t \geq 0, \\ \sum_{i=1}^{N_-(-t)} \xi_{i,-} & \text{if } t < 0. \end{cases}$$

Note that we use the convention $\sum_{i=1}^{0} \xi_{i,+} = \sum_{i=1}^{0} \xi_{i,-} = 0$.

We use the notation $\rightsquigarrow$ for the weak convergence of probability measures.

**Theorem 2.1.** *Consider the identifiable case under the null, i.e.* $\gamma_0 \neq 0$ *and* $\lambda_0 = 0$. *Assume* (C.I) *holds. Then* $n(\hat{c}_n - c_0)$ *is bounded in probability, and* $(\sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0), S_n) \rightsquigarrow (\mathbb{Z}_{\boldsymbol{\eta}}, \mathbb{Z}_S)$.

*In addition, if the conditional distribution of* $(\boldsymbol{W}, U)$ *given* $X = c$ *is continuous in a neighbourhood of* $c_0$ *with respect to the weak convergence*[1], *then*

$$(n(\hat{c}_n - c_0),\ \sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0),\ S_n) \rightsquigarrow (\mathbb{Z}_c,\ \mathbb{Z}_{\boldsymbol{\eta}},\ \mathbb{Z}_S).$$

*Proof.* See Sections C.1 and C.3 of the Appendix. □

**Remark 2.3.** *It is well known that for change-point models, the weak limits of the MLE for the regular parameters (*$\mathbb{Z}_{\boldsymbol{\eta}}$ *above) and for the change-point (*$\mathbb{Z}_c$ *above) are independent, and the latter is the smallest maximizer of a two-sided, compound Poisson process (*$\mathbb{D}(\cdot)$ *above). See [22, 36, 44] for related results in the regression models, and [43] in the Cox models.*

Now consider the non-identifiable case, i.e. $\gamma_0 = 0$. For each $c \in \mathbb{R}$, define $\tilde{\boldsymbol{Z}}_c = (\boldsymbol{W}^T, U, X_c, UX_c)^T$, and denote by $\ell^\infty(A)$ the space of bounded functions on an arbitrary index set $A$. Let $\{((\boldsymbol{\Delta}_c^{(1)})^T, \Delta_c^{(2)})^T : c \in [\ell, u]\}$ be a zero mean Gaussian process, that is tight in $(\ell^\infty([\ell, u]))^{d+3}$, whose covariance function is given as follows: for any $c_1, c_2 \in [\ell, u]$,

$$\mathrm{COV}(((\boldsymbol{\Delta}_{c_1}^{(1)})^T, \Delta_{c_1}^{(2)})^T, ((\boldsymbol{\Delta}_{c_2}^{(1)})^T, \Delta_{c_2}^{(2)})^T) = \mathrm{E}[\phi''(\boldsymbol{\alpha}_0^T \boldsymbol{W} + \beta_0 U) \tilde{\boldsymbol{Z}}_{c_1} \tilde{\boldsymbol{Z}}_{c_2}^T]. \quad (2.6)$$

Note that for each $c$, $\boldsymbol{\Delta}_c^{(1)}$ is of length $d + 2$ and $\boldsymbol{\Delta}_c^{(2)}$ of length 1, and that the existence of such a Gaussian process is established in Theorem 2.2.

Finally, define the following quantities associated with the Gaussian process:

$$\begin{aligned} &\mathbb{C} = \operatorname*{argmax}_{c\in[\ell,u]} \left(\boldsymbol{\Delta}_c^{(1)}\right)^T (\boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(1)})^{-1} \boldsymbol{\Delta}_c^{(1)}, \quad \mathbb{H} = (\boldsymbol{V}_{\mathbb{C},\boldsymbol{\eta}_0}^{(1)})^{-1} \boldsymbol{\Delta}_{\mathbb{C}}^{(1)}, \\ &\mathbb{S} = \Delta_{\mathbb{C}}^{(2)} - \boldsymbol{V}_{\mathbb{C},\boldsymbol{\eta}_0}^{(2)} \mathbb{H}, \end{aligned} \quad (2.7)$$

where $\boldsymbol{V}_{c,\boldsymbol{\eta}}^{(1)}$ and $\boldsymbol{V}_{c,\boldsymbol{\eta}}^{(2)}$ are defined in (2.5).

[1]That is, for any sequence $c_n \to c$, we have $\mathrm{E}[f(\boldsymbol{W}, U)|X = c_n] \to \mathrm{E}[f(\boldsymbol{W}, U)|X = c]$ for any continuous, bounded function $f$.

**Theorem 2.2.** *Consider the non-identifiable case under the null, i.e.* $\gamma_0 = \lambda_0 = 0$*, and assume* (C.I) *holds. Then there exists a zero mean Gaussian process, that is tight in* $(\ell^\infty([\ell, u]))^{d+3}$ *and that has the covariance function given by* (2.6)*. Further,*

$$(\hat{c}_n,\ \sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0),\ S_n) \rightsquigarrow (\mathbb{C},\ \mathbb{H},\ \mathbb{S}).$$

*Proof.* See Section C.2 of the Appendix. □

In this work, our primary focus is on the limiting behavior of the score-type test statistic $S_n$, which is distinct in the two cases. Specifically, in the identifiable case, $\hat{c}_n$ is a consistent estimator of $c_0$ with the rate $n$, and, as a result, the limiting distribution of $S_n$ is the same as that for $S_{n,c_0}$ [31], despite the fact that $c_0$ is unknown.

In the non-identifiable case, however, $\hat{c}_n$ converges to a non-degenerate limit without scaling, and the limiting distribution of $S_n$ is non-Gaussian. For illustration, in Appendix C.4, we show that if $\boldsymbol{W} = 1$, $X$ has the uniform distribution over $(0, 1)$, and $U$ is independent of $X$, then $\mathbb{C}$ is the maximizer of a Brownian bridge, while $\mathbb{S}$ is the value of another independent Brownian bridge evaluated at $\mathbb{C}$, up to a multiplicative constant.

In practice, it is usually unknown whether the observations are from an identifiable model or not. Therefore, it is important to develop a procedure for obtaining the critical values for the test statistic $S_n$, that does not rely on the knowledge of identifiability, but nonetheless is valid in both cases. In the next section, we propose an $m$-out-of-$n$ bootstrap method for this purpose, and establish its validity.

## 3. Bootstrap method for the score-type profile tests

We propose the following $m$-out-of-$n$ bootstrap procedure for the score-type profile test statistic $S_n$ in (2.3), where $m_n \le n$ below are user-specified integers.

STEP 1. Based on data $\boldsymbol{\mathcal{D}}_i,\ i \in [n]$, compute the MLE $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n)$ under the null $H_0 : \lambda_0 = 0$ and the score-type test statistic $S_n$ as in (2.3).

STEP 2. Randomly sample *with replacement* from $(\boldsymbol{W}_i, U_i, X_i),\ \ i \in [n]$ to obtain a bootstrap sample of size $m_n$ denoted as $(\boldsymbol{W}_i^*, U_i^*, X_i^*),\ i \in [m_n]$. For each $i \in [m_n]$, given $(\boldsymbol{W}_i^*, U_i^*, X_i^*) = (\boldsymbol{w}, u, x)$, generate $Y_i^*$ is from the distribution with the following $\nu$-density:

$$y \ \mapsto\ \exp(y(\hat{\boldsymbol{\eta}}_n^T \boldsymbol{z}_{\hat{c}_n}) - \phi(\hat{\boldsymbol{\eta}}_n^T \boldsymbol{z}_{\hat{c}_n})).$$

The bootstrap sample is $\boldsymbol{\mathcal{D}}_i^* = (Y_i^*, \boldsymbol{W}_i^*, U_i^*, X_i^*),\ i \in [m_n]$.

STEP 3. Based on the bootstrapped data $\boldsymbol{\mathcal{D}}_i^*,\ i \in [m_n]$, compute the MLE $(\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*)$ and the score-type test statistic $S_n^*$ as in (2.3) with $n$ replaced by $m_n$

and $\boldsymbol{\mathcal{D}}_i$ replaced by $\boldsymbol{\mathcal{D}}_i^*$. That is,

$$(\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*) := \underset{(c,\boldsymbol{\eta})\in[\ell,u]\times\mathbb{R}^{d+2}}{\operatorname{sargmax}} {m_n}^{-1} \sum_{i=1}^{m_n} \varphi_{c,\boldsymbol{\eta}}(\boldsymbol{\mathcal{D}}_i^*), \quad \text{where} \quad \varphi_{c,\boldsymbol{\eta}} := \varphi_{c,\boldsymbol{\eta},0},$$
$$S_n^* := \frac{1}{\sqrt{m_n}} \sum_{i=1}^{m_n} U_i^* X_{i,\hat{c}_n^*}^* (Y_i^* - \phi'((\hat{\boldsymbol{\eta}}_n^*)^T \boldsymbol{Z}_{i,\hat{c}_n^*}^*)). \tag{3.1}$$

STEP 4. Denote by $\mathrm{pr}_{|\boldsymbol{\mathcal{D}}}$ the conditional probability given the data $\boldsymbol{\mathcal{D}}_i, i \in [n]$. Then the $p$-value is defined as

$$p_n^S = 1 - \mathrm{pr}_{|\boldsymbol{\mathcal{D}}}(|S_n^*| \leq |S_n|). \tag{3.2}$$

If $p_n^S$ is smaller than a user-given significance level, we reject the null hypothesis $H_0 : \lambda_0 = 0$. Otherwise we cannot reject the null.

**Remark 3.1.** *In practice the p-value can be approximated by:*

$$\hat{p}_n^S = B^{-1} \sum_{b=1}^{B} I(|S_n^{b*}| > |S_n|),$$

*where $S_n^{b*}$ is the value of the score-type test statistic $S_n^*$ defined in (3.1) based on the b-th bootstrap sample and $B$ is the number of bootstrap repetitions.*

### *3.1. Asymptotic consistency of the m-out-of-n bootstrap*

In this section, we establish the asymptotic theory for the bootstrap procedure for the score-type test statistic. We denote by $o_{\mathrm{pr}}(1)$ a sequence of random variables that goes to zero in probability as $n \to \infty$. Recall that $m_n \leq n$ is the size of the bootstrap sample, and that $(\mathbb{Z}_{\boldsymbol{\eta}}, \mathbb{Z}_S)$ and $(\mathbb{C}, \mathbb{H}, \mathbb{S})$ are the weak limits appearing Theorems 2.1 and 2.2.

**Theorem 3.1.** *Consider the null, i.e., $\lambda_0 = 0$. Assume that* (C.I) *holds and that $m_n \to \infty$ as $n \to \infty$.*

*(i). If the model is identifiable, i.e., $\gamma_0 \neq 0$, then*

$$\sup_{\boldsymbol{t}\in\mathbb{R}^{d+3}} \left| \mathrm{pr}_{|\boldsymbol{\mathcal{D}}}((\sqrt{m_n}(\hat{\boldsymbol{\eta}}_n^* - \hat{\boldsymbol{\eta}}_n), S_n^*) \leq \boldsymbol{t}) - \mathrm{pr}((\mathbb{Z}_{\boldsymbol{\eta}}, \mathbb{Z}_S) \leq \boldsymbol{t}) \right| = o_{\mathrm{pr}}(1).$$

*(ii). If the model is non-identifiable, i.e., $\gamma_0 = 0$, and if, additionally, $m_n/n \to 0$ as $n \to \infty$, then*

$$\sup_{\boldsymbol{t}\in\mathbb{R}^{d+4}} \left| \mathrm{pr}_{|\boldsymbol{\mathcal{D}}}((\hat{c}_n^*, \sqrt{m_n}(\hat{\boldsymbol{\eta}}_n^* - \hat{\boldsymbol{\eta}}_n), S_n^*) \leq \boldsymbol{t}) - \mathrm{pr}((\mathbb{C}, \mathbb{H}, \mathbb{S}) \leq \boldsymbol{t}) \right| = o_{\mathrm{pr}}(1).$$

*Proof.* The proof for (i) and (ii) can be found, respectively, in Sections C.1 and C.2 of the Appendix. □

The above theorem, together with Theorems 2.1 and 2.2, establishes the asymptotic size validity of the proposed $m$-out-of-$n$ bootstrap procedure for both the identifiable and non-identifiable cases, as long as the bootstrap sample size is of a smaller order than the original sample size, i.e., $m_n/n \to 0$, despite the fact that the procedure does not use the knowledge about the identifiability.

In the *identifiable* case (see part (i) above), however, the condition $m_n/n \to 0$ is not required, and $m_n = n$ is allowed for obtaining the critical value for the test statistic $S_n$, which corresponds to the standard bootstrap. It is noteworthy that by [36, 44, 43] the standard bootstrap is not asymptotically valid for constructing confidence intervals for $\hat{c}_n$, but, by Theorem 3.1 above, is in fact so for the MLE $\hat{\boldsymbol{\eta}}_n$ and $S_n$. That is, for the same model, bootstrap methods may work for some statistics, but fail for others.

In the *non-identifiable* case (see part (ii) above), we show in the next subsection that the standard bootstrap, corresponding to $m_n = n$, is inconsistent, in the sense that the bootstrap distribution of $S_n^*$ in (3.1) does not converge weakly to the limiting distribution (i.e. $\mathbb{S}$) of the test statistic $S_n$, in probability. From the simulation studies in Section 5, with $m_n = n^\kappa$ (rounded to an integer), we observe that as $\kappa$ varies from 0.9 to 1, the proposed procedure controls the size well and is not sensitive to the choice of $\kappa$; thus $\kappa = 0.95$ seems to be a reasonable choice in practice.

**Remark 3.2.** *In the identifiable case, if $m_n/n$ does vanish, then again by [36, 44, 43], the $m$-out-of-$n$ bootstrap procedure is also consistent for $\hat{c}_n$, which are now standard results and omitted, since our primary focus is on testing $H_0$ : $\lambda_0 = 0$ using $S_n$.*

### *3.2. Inconsistency of the standard bootstrap for the non-identifiable case*

In this subsection, we establish the inconsistency of the standard bootstrap, i.e., $m_n = n$, for the non-identifiable case, i.e., $\gamma_0 = 0$. To make this statement precise, we recall the notations in Subsection 2.1 and introduce additional ones.

Denote by $\hat{\mathcal{R}}_n$ the empirical distribution of the covariates $\{(\boldsymbol{W}_i, U_i, X_i) : i \in [n]\}$. For each integer $k \geq 1$, denote by $\mathcal{M}(\mathbb{R}^k)$ the space of Borel probability measures on $\mathbb{R}^k$, and we equip it with the Prokhorov metric $d_{\mathsf{Prok}}(\cdot,\cdot)$ [8, Section 6.5], which characterizes the weak convergence and under which $\mathcal{M}(\mathbb{R}^k)$ is a complete and separable metric space.

The $m$-out-of-$n$ bootstrap procedure with $m_n = n$ requires three inputs: the empirical distribution $\hat{\mathcal{R}}_n$ from which bootstrap covariates $\{(\boldsymbol{W}_i^*, U_i^*, X_i^*) : i \in [n]\}$ are drawn, and the estimators $\hat{c}_n$ and $\hat{\boldsymbol{\eta}}_n$ which are used to generate the bootstrap responses $\{Y_i^* : i \in [n]\}$. Denote by $\mathcal{L}_n(c, \boldsymbol{\eta}, \mathcal{R})$ the distribution of the bootstrap test statistic $S_n^*$ when $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n, \hat{\mathcal{R}}_n)$ takes the value $(c, \boldsymbol{\eta}, \mathcal{R})$. That is, $\mathcal{L}_n$ is a measurable mapping from $[\ell, u] \times \mathbb{R}^{d+2} \times \mathcal{M}(\mathbb{R}^{d+2})$ to $\mathcal{M}(\mathbb{R})$, and $\mathcal{L}_n(\hat{c}_n, \hat{\boldsymbol{\eta}}_n, \hat{\mathcal{R}}_n)$ is the bootstrap distribution of $S_n^*$ given the data, which is a random element in $\mathcal{M}(\mathbb{R})$.

Finally, recall that $\mathbb{S}$ in (2.7) is the limiting distribution of the test statistic $S_n$, which a deterministic element in $\mathcal{M}(\mathbb{R})$. The next theorem shows that the Prokhorov distance between the bootstrap distribution and the target does not converge to zero in probability.

**Theorem 3.2.** *Consider the null, i.e.,* $\lambda_0 = 0$*, and the non-identifiable case, i.e.,* $\gamma_0 = 0$*. Further, consider the standard bootstrap with* $m_n = n$*. Assume that* (C.I) *holds. There exists some* $\epsilon > 0$ *such that*

$$\liminf_{n\to\infty} \operatorname{pr}\left(d_{\mathsf{Prok}}(\mathcal{L}_n(\hat{c}_n, \hat{\boldsymbol{\eta}}_n, \hat{\mathcal{R}}_n), \mathbb{S}) \geq \epsilon\right) > 0,$$

*where* $\mathbb{S}$ *in the second argument of* $d_{\mathsf{Prok}}(\cdot,\cdot)$ *refers to its distribution.*

*Proof.* See Appendix C.5. □

We briefly discuss the proof strategy for the non-identifiable case, i.e., $\gamma_0 = 0$ and the standard bootstrap, i.e., $m_n = n$. From Theorem 2.2, $(\hat{c}_n, \sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)) \rightsquigarrow (\mathbb{C}, \mathbb{H})$, and the empirical distribution $\hat{\mathcal{R}}_n$ converges weakly (i.e., in terms of $d_{\mathsf{Prok}}$) to the population distribution $\mathcal{R}_\infty$ of the covariates $(\boldsymbol{W}, U, X)$ almost surely [14, Theorem 11.4.1]. Due to Skorohod's representation theorem [8, Theorem 6.7], there exist a sequence of random variables $\{(c_n^\dagger, \boldsymbol{\eta}_n^\dagger, \mathcal{R}_n^\dagger) : n \geq 1\}$ and $(\mathbb{C}^\dagger, \mathbb{H}^\dagger)$ such that $(c_n^\dagger, \boldsymbol{\eta}_n^\dagger, \mathcal{R}_n^\dagger)$ has the same distribution as $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n, \hat{\mathcal{R}}_n)$ for each $n \geq 1$, $(\mathbb{C}^\dagger, \mathbb{H}^\dagger)$ as $(\mathbb{C}, \mathbb{H})$, and for each $\omega \in \Omega$,

$$\lim_{n\to\infty} (c_n^\dagger(\omega), \sqrt{n}(\boldsymbol{\eta}_n^\dagger(\omega) - \boldsymbol{\eta}_0), \mathcal{R}_n^\dagger(\omega)) = (\mathbb{C}^\dagger(\omega), \mathbb{H}^\dagger(\omega), \mathcal{R}_\infty).$$

Denote by $\gamma_n^\dagger(\omega)$ the last component of $\boldsymbol{\eta}_n^\dagger(\omega)$, and by $\mathbb{H}_\gamma^\dagger(\omega)$ the last component of $\mathbb{H}^\dagger(\omega)$. Since $\gamma_0 = 0$, we have $\sqrt{n}\gamma_n^\dagger(\omega) \to \mathbb{H}_\gamma^\dagger(\omega)$. In Appendix C.5 (Theorem C.1), we show that

$$d_{\mathsf{Prok}}\left(\mathcal{L}_n(c_n^\dagger, \boldsymbol{\eta}_n^\dagger, \mathcal{R}_n^\dagger),\ \mathcal{L}_\infty(\mathbb{C}^\dagger, \mathbb{H}_\gamma^\dagger)\right) = o_{\mathrm{pr}}(1),$$

where $\mathcal{L}_\infty$ is some measurable map from $[\ell, u] \times \mathbb{R}$ to $\mathcal{M}(\mathbb{R})$. Since $\mathcal{L}_\infty(\mathbb{C}^\dagger, \mathbb{H}_\gamma^\dagger)$ is a random measure, i.e., depending on $\omega \in \Omega$, and the law of $\mathbb{S}$ is fixed, we have that for some $\epsilon > 0$,

$$\liminf_{n\to\infty} \operatorname{pr}\left(d_{\mathsf{Prok}}(\mathcal{L}_n(c_n^\dagger, \boldsymbol{\eta}_n^\dagger, \mathcal{R}_n^\dagger), \mathbb{S}) \geq \epsilon\right) > 0,$$

which is equivalent to the conclusion in the above theorem, as $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n, \hat{\mathcal{R}}_n)$ and $(c_n^\dagger, \boldsymbol{\eta}_n^\dagger, \mathcal{R}_n^\dagger)$ have the same distribution by construction.

**Remark 3.3.** *If* $m_n/n \to 0$*, we have* $\sqrt{m_n}\hat{\gamma}_n = \sqrt{n}\hat{\gamma}_n \times \sqrt{m_n/n} = o_{\mathrm{pr}}(1)$*. Further, it turns out that* $\mathcal{L}_\infty(c, 0)$ *is equal to the law of* $\mathbb{S}$ *for any* $c \in [\ell, u]$*. This explains the consistency in Theorem 3.1(ii) when the bootstrap sample size* $m_n$ *is of a smaller order compared to the original sample size* $n$*.*

**Remark 3.4.** *Theorem 3.1 shows that the* $m$*-out-of-*$n$ *bootstrap with* $m_n/n \to 0$ *is consistent in the sense that for each* fixed $\boldsymbol{\eta}_0 \in \mathbb{R}^{d+2}$ *and* $c_0 \in [\ell, u]$*,*

$\lim_{n\to\infty} \mathrm{pr}(p_n^S \le \alpha) = \alpha$. *The procedure is robust against model identifiability issues, that is, the consistency holds whether* $\gamma_0 = 0$ *or not.*

*However, we note that the procedure does not control the size uniformly over the parameter space* $\mathbb{R}^{d+2} \times [\ell, u]$, *in the sense [4, 3] that under* $H_0 : \lambda_0 = 0$,

$$\lim_{n\to\infty} \sup_{(\boldsymbol{\eta},c)\in\mathbb{R}^{d+2}\times[\ell,u]} \mathrm{pr}_{\boldsymbol{\eta},c}(p_n^S \le \alpha) \le \alpha,$$

*where* $\mathrm{pr}_{\boldsymbol{\eta},c}$ *denotes the probability when the value of* $(\boldsymbol{\eta}_0, c_0)$ *is* $(\boldsymbol{\eta}, c)$. *This is a stronger requirement, which allows the parameters* $(\boldsymbol{\eta}_0, c_0)$ *to change with the sample size* $n$, *as opposed to being fixed relative to* $n$. *Indeed, if* $\gamma_0 = B_1/\sqrt{n}$ *for some constant* $B_1 \neq 0$ *and other parameters are fixed, which corresponds to the weakly identifiable case in [3], the limiting distribution of the test statistic* $S_n$ *would be* $\mathcal{L}_\infty(c_0, B_1)$ *by Theorem A.5 in the Appendix. Further, if* $m_n/n \to 0$, *by similar arguments as for Theorem 3.1(ii) and 3.4, the bootstrap distribution of* $S_n^*$ *converges weakly to* $\mathbb{S}$ *in probability, whose law differs from* $\mathcal{L}_\infty(c_0, B_1)$ *if* $B_1 \neq 0$.

### *3.3. Power analysis under local alternatives*

In this subsection, we study the rejection probabilities under the following local alternatives:

$$H_{1,n} \;:\; \lambda_n = B_0/\sqrt{n}, \text{ for some constant } B_0 \neq 0. \tag{3.3}$$

That is, conditional on $(\boldsymbol{W}, U, X) = (\boldsymbol{w}, u, x)$, the $\nu$-density of $Y = y$ is

$$\exp\left(y\left(\boldsymbol{\eta}_0^T \boldsymbol{z}_{c_0} + \lambda_n u x_{c_0}\right) - \phi(\boldsymbol{\eta}_0^T \boldsymbol{z}_{c_0} + \lambda_n u x_{c_0})\right).$$

For simplicity, we assume that the constant $B_0$ and other parameters $(\boldsymbol{\eta}_0, c_0)$, as well as the distribution of $(\boldsymbol{W}, U, X)$, do not depend on the sample size $n$.

We start with the identifiable case, and recall that $\mathbb{Z}_S$ has the zero mean normal distribution with variance $\sigma^2_{c_0,\boldsymbol{\eta}_0}$ given in (2.5).

**Theorem 3.3.** *Assume that* $\gamma_0 \neq 0$, *and* (C.I) *holds. Consider the local alternatives* $H_{1,n}$ *in* (3.3). *As* $n \to \infty$, $S_n$ *converges in distribution to* $\mathbb{Z}_S + B_0\sigma^2_{c_0,\boldsymbol{\eta}_0}$.

*Further, consider the bootstrap procedure with* $m_n \to \infty$ *as* $n \to \infty$. *For each* $\alpha \in (0, 1)$,

$$\lim_{n\to\infty} \mathrm{pr}(p_n^S \le \alpha) = \Phi(\Phi^{-1}(\alpha/2) + B_0\sigma_{c_0,\boldsymbol{\eta}_0}) + \Phi(\Phi^{-1}(\alpha/2) - B_0\sigma_{c_0,\boldsymbol{\eta}_0}),$$

*where* $\Phi(\cdot)$ *is the cumulative distribution function of* $N(0, 1)$.

*Proof.* See Appendix D.1. □

For the identifiable case, the power under the local alternatives in (3.3) is the same as if the unknown cutpoint $c_0$ was known [41]. This is because if $\gamma_0 \neq 0$, the cutpoint $c_0$ can still be estimated at a super parametric rate under $H_{1,n}$.

Next, we consider the non-identifiable case. Recall the definitions of $\boldsymbol{V}_{c,\boldsymbol{\eta}}^{(1)}$ and $\boldsymbol{V}_{c,\boldsymbol{\eta}}^{(2)}$ in (2.5), and the zero mean Gaussian process $\{((\boldsymbol{\Delta}_c^{(1)})^T, \Delta_c^{(2)})^T : c \in [\ell, u]\}$ in Subsection 2.1. Define

$$\overline{\mathbb{C}} := \operatorname*{sargmax}_{c\in[\ell,u]} \frac{1}{2}\left(\boldsymbol{\Delta}_c^{(1)} + B_0\boldsymbol{V}_c^{(4)}\right)^T (\boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(1)})^{-1}(\boldsymbol{\Delta}_c^{(1)} + B_0\boldsymbol{V}_c^{(4)}),$$
$$\overline{\mathbb{S}} := \Delta_{\overline{\mathbb{C}}}^{(2)} - \boldsymbol{V}_{\overline{\mathbb{C}},\boldsymbol{\eta}_0}^{(2)}(\boldsymbol{V}_{\overline{\mathbb{C}},\boldsymbol{\eta}_0}^{(1)})^{-1}\boldsymbol{\Delta}_{\overline{\mathbb{C}}}^{(1)} + B_0\left(V_{\overline{\mathbb{C}}}^{(5)} - \boldsymbol{V}_{\overline{\mathbb{C}},\boldsymbol{\eta}_0}^{(2)}(\boldsymbol{V}_{\overline{\mathbb{C}},\boldsymbol{\eta}_0}^{(1)})^{-1}\boldsymbol{V}_{\overline{\mathbb{C}}}^{(4)}\right),$$

where for each $c \in [\ell, u]$,

$$\boldsymbol{V}_c^{(4)} = \mathrm{E}[\phi''(\boldsymbol{\eta}_0^T\boldsymbol{Z}_{c_0})UX_{c_0}\boldsymbol{Z}_c], \qquad V_c^{(5)} = \mathrm{E}[\phi''(\boldsymbol{\eta}_0^T\boldsymbol{Z}_{c_0})UX_{c_0}X_c]. \tag{3.4}$$

Note that both $\overline{\mathbb{C}}$ and $\overline{\mathbb{S}}$ depend on $B_0$, which is omitted for simplicity. In addition to the bias term, i.e., the last term in the definition of $\overline{\mathbb{S}}$ above, the first two terms are different from $\mathbb{S}$ in (2.7), since they are evaluated at $\overline{\mathbb{C}}$, instead of $\mathbb{C}$ in (2.7).

**Theorem 3.4.** *Assume that $\gamma_0 = 0$, and* (C.I) *holds. Consider the local alternatives $H_{1,n}$ in* (3.3). *As $n \to \infty$, $S_n$ converges in distribution to $\overline{\mathbb{S}}$.*

*Further, consider the bootstrap procedure with $m_n/n \to 0$ as $n \to \infty$. Fix some level $\alpha \in (0,1)$ and denote by $q_{\alpha,\mathbb{S}}$ the upper $\alpha$-th quantile of $|\mathbb{S}|$. Then*

$$\lim_{n\to\infty} \mathrm{pr}(p_n^S \le \alpha) = \mathrm{pr}(|\overline{\mathbb{S}}| \ge q_{\alpha,\mathbb{S}}).$$

*Proof.* See Appendix D.2. □

In the non-identifiable case, although the power function does not have a simple form, under mild conditions, the rejection probability for the proposed procedure approaches one under the local alternatives in (3.3), when the magnitude of $B_0$ diverges, that is, $\lim_{|B_0|\to\infty} \mathrm{pr}(|\overline{\mathbb{S}}| \ge q_{\alpha,\mathbb{S}}) = 1$. To see this, due to the definition of $\overline{\mathbb{C}}$, as $|B_0| \to \infty$, $\overline{\mathbb{C}}$ converges in probability to

$$c^\ddagger := \operatorname*{argmax}_{c\in[\ell,u]} (\boldsymbol{V}_c^{(4)})^T(\boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(1)})^{-1}\boldsymbol{V}_c^{(4)},$$

if the maximizer is unique. Further, if $V_{c^\ddagger}^{(5)} - \boldsymbol{V}_{c^\ddagger,\boldsymbol{\eta}_0}^{(2)}(\boldsymbol{V}_{c^\ddagger,\boldsymbol{\eta}_0}^{(1)})^{-1}\boldsymbol{V}_{c^\ddagger}^{(4)}$ is *non-zero*, then $|\overline{\mathbb{S}}|$ approaches $\infty$ in probability as $|B_0| \to \infty$.

## 4. Wald-type profile tests based on the $m$-out-of-$n$ bootstrap

In this section, we propose the following $m$-out-of-$n$ bootstrap procedure for the Wald-type test statistic, $W_n$, in (2.4). As before, $m_n \le n$ below is a user-specified integer.

Step 1. Compute the MLE $(\tilde{c}_n, \tilde{\boldsymbol{\eta}}_n, \tilde{\lambda}_n)$ for $(c_0, \boldsymbol{\eta}_0, \lambda_0)$ and the test statistic $W_n$ as in (2.4), based on data $\boldsymbol{\mathcal{D}}_i,\ i \in [n]$.

STEP 2. Randomly sample *with replacement* from $(\boldsymbol{W}_i, U_i, X_i)$, $i \in [n]$ to obtain a bootstrap sample of size $m_n$ denoted as $(\boldsymbol{W}_i^*, U_i^*, X_i^*)$, $i \in [m_n]$. For each $i \in [m_n]$, given $(\boldsymbol{W}_i^*, U_i^*, X_i^*) = (\boldsymbol{w}, u, x)$, generate $Y_i^*$ from the distribution with the following $\nu$-density:

$$y \;\mapsto\; \exp(y(\tilde{\boldsymbol{\eta}}_n^T z_{\tilde{c}_n}) - \phi(\tilde{\boldsymbol{\eta}}_n^T z_{\tilde{c}_n})).$$

The bootstrap sample is $\tilde{\boldsymbol{\mathcal{D}}}_i^* = (Y_i^*, \boldsymbol{W}_i^*, U_i^*, X_i^*)$, $i \in [m_n]$.

STEP 3. Based on the bootstrapped data $\tilde{\boldsymbol{\mathcal{D}}}_i^*$, $i \in [m_n]$, compute the MLE $(\tilde{c}_n^*, \tilde{\boldsymbol{\eta}}_n^*, \tilde{\lambda}_n^*)$ and the test statistic $W_n^*$ using (2.4) with $n$ replaced by $m_n$ and $\boldsymbol{\mathcal{D}}_i$ replaced by $\tilde{\boldsymbol{\mathcal{D}}}_i^*$. That is,

$$(\tilde{c}_n^*, \tilde{\boldsymbol{\eta}}_n^*, \tilde{\lambda}_n^*) := \underset{(c,\boldsymbol{\eta},\lambda)\in[\ell,u]\times\mathbb{R}^{d+3}}{\operatorname{sargmax}} {m_n}^{-1} \sum_{i=1}^{m_n} \varphi_{c,\boldsymbol{\eta},\lambda}(\tilde{\boldsymbol{\mathcal{D}}}_i^*), \qquad W_n^* := \sqrt{m_n}\tilde{\lambda}_n^*.$$

STEP 4. Recall that $\mathrm{pr}_{|\boldsymbol{\mathcal{D}}}$ denotes the conditional probability given the data $\boldsymbol{\mathcal{D}}_i$, $i \in [n]$. Then the $p$-value is defined as

$$p_n^W = 1 - \mathrm{pr}_{|\boldsymbol{\mathcal{D}}}(|W_n^*| \le |W_n|).$$

If $p_n^W$ is smaller than a user-given significance level, we reject the null hypothesis $H_0 : \lambda_0 = 0$. Otherwise we cannot reject the null.

As mentioned above, we omit the precise statements regarding the properties of Wald-type profile tests, due to its similarity to score-type profile tests, and also because in Section 5 we notice that its performance is not as good in terms of empirical sizes and powers, when the sample size is moderate ($100 \sim 500$).

To understand this issue, we conduct the following simulation study: $\boldsymbol{W} = 1$, $X$ is uniformly distributed over $(0, 1)$, $U$ is independent of $X$ with $\mathrm{E}[U] = 0.5$, the response follows the logistic regression model, and $c_0 = 0.5$, $[\ell, u] = [0.15, 0.85]$, $\boldsymbol{\eta}_0 = (1, -1.5, 0)^T$, $\lambda_0 = 0$. In Figure 1, we present histograms and density lines, with $7.5 \times 10^5$ repetitions, of the sampling distributions of $\sqrt{n}\tilde{\lambda}_n$ for $n = 100$, 300, 1000, $10^6$. From Figure 1 we observe that $\sqrt{n}\tilde{\lambda}_n$ approaches its limiting distribution (in the last panel) rather slowly, not until $n$ is 1000. The theoretical investigation of the slow convergence rate is left for future work.

## 5. Simulation study

In this section we conduct simulation studies to evaluate the finite-sample performance of the proposed methods. In each repetition of the simulation, for each $i \in [n]$ where the sample size $n \in \{200, 300, 500\}$, we consider the following independent covariates: an intercept $\boldsymbol{W}_i = 1$, a continuous biomarker $X_i$ from the uniform distribution on $(0, 1)$, and a treatment indicator $U_i$ from the Bernoulli distribution with a success probability 0.5, and further a binary outcome $Y_i$, whose conditional expectation, given $(\boldsymbol{W}_i, X_i, U_i)$, is specified by model (1.2) with the logistic link function, and $c_0 = 0.5$. For regression parameters, we consider $\lambda_0 = 0$ under the $H_0$ and $\lambda_0 = 2$ under $H_1$, and four different choices for

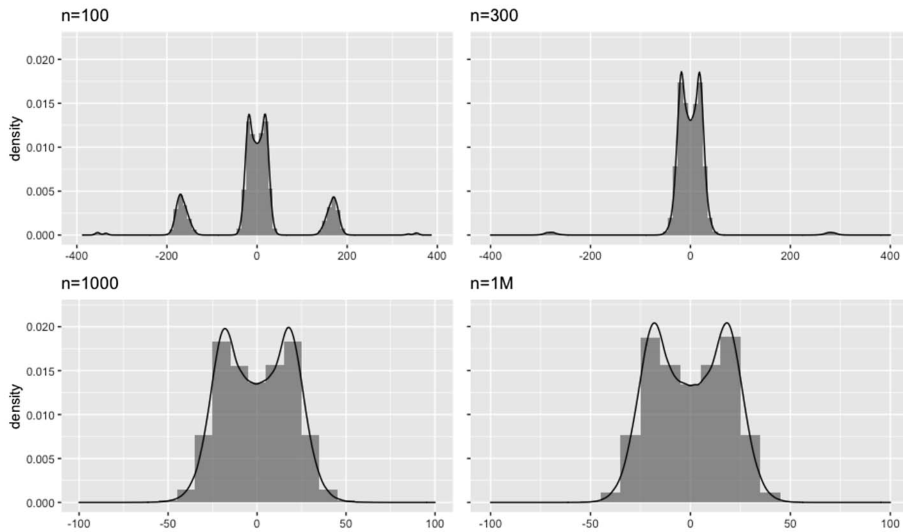


FIG 1. *The sampling distributions of $W_n$ under the null for different $n$. "1M" denotes $10^6$.*

$\boldsymbol{\eta}_0 \in \{\boldsymbol{\eta}^{(k)} : k \in [4]\}$, which are specified in the captions of Tables 1 and 2. Specifically, under the null, $\boldsymbol{\eta}^{(3)}, \boldsymbol{\eta}^{(4)}$ correspond to the identifiable case, and $\boldsymbol{\eta}^{(2)}$ to the non-identifiable case, while $\boldsymbol{\eta}^{(1)}$ belongs to the identifiable case with a small main effect (i.e., $|\gamma_0|$ is small). The empirical sizes and powers of the tests below, defined as the proportion of rejections under $H_0$ and $H_1$ respectively, are calculated with $R = 2000$ repetitions at the level 5%.

The bootstrap score-type and Wald-type tests, proposed respectively in Sections 3 and 4, are referred as "B-Score" and "B-Wald" in this section. For each sample size $n$, we let $[\ell, u] = [15\%, 85\%]$, consider the bootstrap sample size $m_n = n^\kappa$ (rounded to an integer) with $\kappa \in \{0.9, 0.9375, 0.95, 1\}$, and use $B = 2000$ bootstrap repetitions. We compare the proposed methods with several tests mentioned in Section 1, which are asymptotically valid under either the identifiable case or non-identifiable case, but not both.

First, in Theorem 2.1, we show that if the cutpoint $c_0$ is identifiable (i.e., $\gamma_0 \neq 0$), the profile score test statistic $S_n$ in (2.3) convergences in distribution to $N(0, \sigma^2_{c_0,\boldsymbol{\eta}_0})$, where $\sigma^2_{c,\boldsymbol{\eta}}$ is defined in (2.5). By a similar argument as for Theorem 2.1, one can show that $\sigma^2_{\hat{c}_n,\hat{\boldsymbol{\eta}}_n}$ is a consistent estimator for $\sigma^2_{c_0,\boldsymbol{\eta}_0}$ if $\gamma_0 \neq 0$, where $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n)$ are the MLE in (2.3). Thus we may reject the null if $|S_n/\sigma_{\hat{c}_n,\hat{\boldsymbol{\eta}}_n}| \geq \Phi^{-1}(0.975)$, where $\Phi(\cdot)$ is the cumulative distribution function of $N(0, 1)$. This procedure obtains the critical value by the asymptotic approximation under the *identifiable* assumption, and thus is referred as "A-Score".

Second, we consider the minimum $p$-value methods, with and without adjustment. For the unadjusted version, which is called as "MP", we reject the null if $\sup_{c\in[\ell,u]} |S_{n,c}/\hat{\sigma}_{c,\hat{\boldsymbol{\eta}}_{n,c}}| \geq \Phi^{-1}(0.975)$, where $S_{n,c}$ and $\hat{\boldsymbol{\eta}}_{n,c}$ are defined in (2.2), and $\sigma^2_{c,\boldsymbol{\eta}}$ is defined in (2.5). For the adjustment under the *non-identifiable* assumption, which is referred to as "MP(adj)", we define the following test statistic $S_{adj}$ [40] and $p$-value $p_{n,adj}$ [8],

$$p_{n,adj} = 2\sum_{k=1}^{\infty}(-1)^{k+1}\exp(-2k^2S_{adj}^2), \quad \text{where } S_{adj} = \sup_{c\in[\ell,u]}\left|\frac{\sum_{i=1}^{n} X_{i,c}U_i\tilde{\xi}_i}{\tilde{\sigma}_{mp}\sqrt{\sum_{i=1}^{n} U_i}}\right|,$$

with $\tilde{\xi}_i := Y_i - \phi'(\tilde{\boldsymbol{\alpha}}_n + \tilde{\beta}_n U_i)$, $\tilde{\sigma}_{mp}^2 := n^{-1}\sum_{i=1}^{n}\tilde{\xi}_i^2 - (n^{-1}\sum_{i=1}^{n}\tilde{\xi}_i)^2$, and $\tilde{\boldsymbol{\alpha}}_n$ and $\tilde{\beta}_n$ are the MLE of $\boldsymbol{\alpha}_0$ and $\beta_0$ under the assumption $\gamma_0 = \lambda_0 = 0$.

The empirical sizes and powers for the above tests are summarized in Tables 1 and 2. From Table 1, the empirical sizes of the B-Score method are close to the nominal 5% level under both the identifiable and non-identifiable cases, and are not sensitive to the choice of bootstrap sample size $m_n$. The empirical sizes from the MP method can be seven times of the nominated level in all cases. Although the A-Score test works reasonably well under the identifiable case ($\boldsymbol{\eta}^{(3)}$ and $\boldsymbol{\eta}^{(4)}$), it loses the control of Type I error in the non-identifiable case ($\boldsymbol{\eta}^{(2)}$) and the small-main-effect case ($\boldsymbol{\eta}^{(1)}$). In contrast, the MP(adj) method works well in the non-identifiable case ($\boldsymbol{\eta}^{(2)}$) but behaves poorly in the identifiable case ($\boldsymbol{\eta}^{(3)}$ and $\boldsymbol{\eta}^{(4)}$), and also when $|\gamma_0|$ is small ($\boldsymbol{\eta}^{(1)}$). The results on empirical sizes indicate that A-Score and MP(adj) methods are quite sensitive to their corresponding identifiability assumption.

The empirical sizes of the B-Wald are very conservative, and from Table 2, its empirical powers are poor compared to the B-Score method especially when the sample size is not large (say, less than 500), which makes B-Wald less desirable. The A-Score, MP and MP(adj), as expected, have a slightly better power than the proposed methods, but they fail to control the size in many situations.

## 6. Application to a colorectal cancer dataset

We consider data from a CO.17 trial conducted by the Canadian Cancer Trials Group [19], which randomized 572 patients with advanced colorectal cancers to receive Cetuximab plus best supportive care (BSC) or BSC alone. One of the important clinical outcomes in this trial is the response to the treatment as assessed by Response Evaluation Criteria in Solid Tumours (RECIST), which is categorized as complete response (CR) if all target lesions of this patient disappear, partial response (PR) if there is at least 30% decrease in the sum of longest diameters of the target lesions, progressive disease (PD) if there is at least 20% increase in the sum of longest diameters of the target lesions, or stable disease (SD) if there is not sufficient decrease or increase in the sum of longest diameter to qualify for PR or PD [39]. A patient is said to have benefited clinically when the response was either CR, PR or SD, and it is of interest to identify subgroups of patients who would have different treatment effects with respect to this clinical outcome based on baseline values of some biomarkers.

In this analysis, we consider the following three potential, continuous biomarkers (i.e. $X$): mRNA expression of the gene epiregulin (EREG), the levels of lactate dehydrogenase (LDH) and alkaline phosphatase (ALKPH) in the blood.

TABLE 1
*The empirical sizes (in percentage) for testing $H_0 : \lambda_0 = 0$ at the level 5% for the logistic model. Here, $\boldsymbol{\eta}^{(1)} = (-1.4, 1.2, 0.2)^T$, $\boldsymbol{\eta}^{(2)} = (1, -1.5, 0)^T$,$\boldsymbol{\eta}^{(3)} = (-1.4, 1.2, 1)^T$, $\boldsymbol{\eta}^{(4)} = (-1.4, 1.2, 2)^T$, $\lambda_0 = 2$ under the alternative.*

| | B-Score | | | | B-Wald | | | | A-Score | MP | MP(adj) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $m_n$ | $n^{0.9}$ | $n^{0.9375}$ | $n^{0.95}$ | $n$ | $n^{0.9}$ | $n^{0.9375}$ | $n^{0.95}$ | $n$ | | | |
| $n = 200$ | | | | | | | | | | | |
| $\boldsymbol{\eta}^{(1)}$ | 5.4 | 5.7 | 5.8 | 5.9 | 4.5 | 3.8 | 4.4 | 4.4 | 26.0 | 34.0 | 8.7 |
| $\boldsymbol{\eta}^{(2)}$ | 6.2 | 6.3 | 6.1 | 6.4 | 2.2 | 2.4 | 2.8 | 2.0 | 28.4 | 37.7 | 5.0 |
| $\boldsymbol{\eta}^{(3)}$ | 7.0 | 7.5 | 6.5 | 6.9 | 1.1 | 0.7 | 1.5 | 0.9 | 11.9 | 35.5 | 5.4 |
| $\boldsymbol{\eta}^{(4)}$ | 6.4 | 6.8 | 6.6 | 6.0 | 0.6 | 2.0 | 2.4 | 4.0 | 5.7 | 35.4 | 9.4 |
| $n = 300$ | | | | | | | | | | | |
| $\boldsymbol{\eta}^{(1)}$ | 6.3 | 6.6 | 5.5 | 5.4 | 1.3 | 0.8 | 1.0 | 3.6 | 26.5 | 36.0 | 9.9 |
| $\boldsymbol{\eta}^{(2)}$ | 5.3 | 5.2 | 4.2 | 5.0 | 0.2 | 0.4 | 1.3 | 2.9 | 27.8 | 37.4 | 4.9 |
| $\boldsymbol{\eta}^{(3)}$ | 7.1 | 5.5 | 7.5 | 5.4 | 0.2 | 0.3 | 0.6 | 1.8 | 8.6 | 37.7 | 7.5 |
| $\boldsymbol{\eta}^{(4)}$ | 6.6 | 5.6 | 5.8 | 5.6 | 0.0 | 1.2 | 2.6 | 4.0 | 5.1 | 40.3 | 9.8 |
| $n = 500$ | | | | | | | | | | | |
| $\boldsymbol{\eta}^{(1)}$ | 5.2 | 5.0 | 6.0 | 5.0 | 1.6 | 2.3 | 3.3 | 6.0 | 22.6 | 36.8 | 13.2 |
| $\boldsymbol{\eta}^{(2)}$ | 5.6 | 5.7 | 4.8 | 5.8 | 1.6 | 3.6 | 4.0 | 8.2 | 28.2 | 39.9 | 4.7 |
| $\boldsymbol{\eta}^{(3)}$ | 5.6 | 6.3 | 6.1 | 4.9 | 1.2 | 1.6 | 1.6 | 2.8 | 7.3 | 41.1 | 9.3 |
| $\boldsymbol{\eta}^{(4)}$ | 6.1 | 5.6 | 6.7 | 6.0 | 2.6 | 3.2 | 4.1 | 4.4 | 4.5 | 46.4 | 9.9 |

TABLE 2
*The empirical powers (in percentage) for testing $H_0 : \lambda_0 = 0$ at the level 5% for the logistic model. Here, $\boldsymbol{\eta}^{(1)} = (-1.4, 1.2, 0.2)^T$, $\boldsymbol{\eta}^{(2)} = (1, -1.5, 0)^T$,$\boldsymbol{\eta}^{(3)} = (-1.4, 1.2, 1)^T$, $\boldsymbol{\eta}^{(4)} = (-1.4, 1.2, 2)^T$, $\lambda_0 = 2$ under the alternative.*

| | B-Score | | | | B-Wald | | | | A-Score | MP | MP(adj) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $m_n$ | $n^{0.9}$ | $n^{0.9375}$ | $n^{0.95}$ | $n$ | $n^{0.9}$ | $n^{0.9375}$ | $n^{0.95}$ | $n$ | | | |
| $n = 200$ | | | | | | | | | | | |
| $\boldsymbol{\eta}^{(1)}$ | 73.9 | 74.0 | 73.0 | 73.2 | 2.2 | 2.3 | 3.5 | 13.3 | 81.0 | 92.5 | 97.3 |
| $\boldsymbol{\eta}^{(2)}$ | 79.2 | 78.2 | 76.5 | 75.6 | 0.9 | 1.8 | 2.4 | 16.2 | 86.7 | 96.0 | 96.7 |
| $\boldsymbol{\eta}^{(3)}$ | 72.5 | 70.9 | 71.9 | 70.6 | 9.3 | 24.0 | 30.3 | 53.8 | 73.1 | 88.1 | 94.8 |
| $\boldsymbol{\eta}^{(4)}$ | 53.2 | 53.0 | 53.1 | 54.2 | 45.2 | 57.2 | 61.4 | 66.4 | 48.3 | 56.8 | 98.4 |
| $n = 300$ | | | | | | | | | | | |
| $\boldsymbol{\eta}^{(1)}$ | 89.1 | 88.5 | 89.1 | 88.7 | 3.3 | 5.7 | 35.2 | 59.5 | 93.0 | 98.2 | 98.9 |
| $\boldsymbol{\eta}^{(2)}$ | 90.2 | 91.5 | 92.1 | 92.0 | 5.0 | 14.8 | 48.9 | 67.1 | 96.3 | 99.3 | 98.9 |
| $\boldsymbol{\eta}^{(3)}$ | 88.9 | 87.2 | 87.5 | 86.9 | 32.5 | 61.3 | 80.2 | 89.0 | 87.7 | 97.1 | 98.6 |
| $\boldsymbol{\eta}^{(4)}$ | 73.1 | 72.1 | 71.9 | 72.2 | 50.5 | 73.9 | 80.0 | 81.1 | 69.1 | 83.7 | 99.5 |
| $n = 500$ | | | | | | | | | | | |
| $\boldsymbol{\eta}^{(1)}$ | 98.3 | 99.0 | 98.7 | 98.1 | 83.5 | 90.7 | 92.2 | 95.2 | 99.0 | 99.8 | 99.3 |
| $\boldsymbol{\eta}^{(2)}$ | 98.9 | 99.3 | 98.9 | 98.7 | 92.9 | 94.2 | 95.2 | 96.8 | 99.7 | 99.9 | 99.2 |
| $\boldsymbol{\eta}^{(3)}$ | 97.7 | 98.2 | 98.1 | 98.2 | 97.9 | 98.7 | 98.4 | 98.9 | 98.4 | 99.7 | 99.5 |
| $\boldsymbol{\eta}^{(4)}$ | 92.1 | 91.4 | 91.7 | 92.5 | 94.5 | 94.6 | 94.0 | 95.9 | 89.4 | 98.3 | 99.0 |

The clinical outcome $Y$ is binary, with value 1 if a patient had a clinical benefit and 0 otherwise. The treatment indicator $U$ is 1 if a patient received Cetuximab plus BSC and 0 if BSC alone. We consider the age of patients and the intercept as the additional covariate $\boldsymbol{W}$. The sampling distributions of the three biomarkers and the age covariate are showniIn Figure 2. For each biomarker

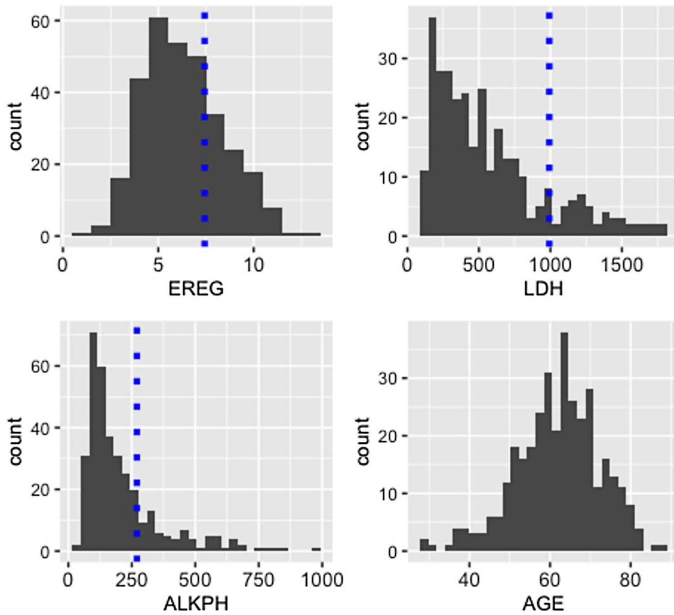


FIG 2. *The sampling distribution of EREG, LDH, ALKPH, and AGE. The vertical dash line for each of the three biomarkers indicates the cutpoint estimated by* (2.4).

TABLE 3
*The p-values for testing $H_0 : \lambda_0 = 0$ from the B-Score, MP and A-Score methods. The* 95% *confidence interval for the p-value from B-Score method shown in parentheses under the p-value.*

| Biomarker | B-Score | | MP | A-Score |
|---|---|---|---|---|
| $m_n$ | $n^{0.9}$ | $n^{0.95}$ | | |
| EREG | $5\times10^{-3}$ | $7\times10^{-3}$ | $1.4\times10^{-3}$ | $1.3\times10^{-3}$ |
| | (0.001, 0.008) | (0.000, 0.008) | | |
| LDH | 0.02 | 0.03 | $2.5\times10^{-3}$ | 0.014 |
| | (0.011, 0.028) | (0.012, 0.030) | | |
| ALKPH | 0.07 | 0.09 | $4.4\times10^{-3}$ | 0.027 |
| | (0.069, 0.105) | (0.066, 0.100) | | |

(i.e. $X$), we consider the regression model (1.2) with the logistic link function for identification of subgroups with differential treatment effects.

First, separately for each biomaker, we applied the proposed B-Score method with $m_n = n^{0.9}$, $n^{0.95}$, MP method, and A-Score method defined in Section 5 to test whether there exist subgroups defined by an unknown cutpoint which have significantly different treatment effects, i.e., testing $H_0 : \lambda_0 = 0$ in model (1.2) with an unspecified $c_0$. The $p$-values from these methods are presented in Table 3. Due to randomness of the resampling methods [26], we computed 95%

Table 4
*An explanatory subgroup analysis that shows the proportion of patients with clinical benefits by treatment and subgroups for biomarkers EREG, LDH and ALKPH.*

| | | Cetuximab+BSC | | BSC | | | |
|---|---|---|---|---|---|---|---|
| Biomarker | Subgroup | Total # | # (%) with Clinical Benefit | Total # | # (%) with Clinical Benefit | Treat Effect | $p$-value |
| EREG | $\leq 7.43$ | 116 | 65 (56%) | 99 | 17 (17%) | 39% | $5.7\times10^{-9}$ |
| | $> 7.43$ | 46 | 8 (17%) | 39 | 9 (23%) | -6% | 0.70 |
| LDH | $\leq 992$ | 150 | 71 (47%) | 109 | 25 (23%) | 24% | $1.0\times10^{-4}$ |
| | $> 992$ | 40 | 20 (50%) | 25 | 0 (0%) | 50% | $7.1\times10^{-5}$ |
| ALKPH | $\leq 270$ | 152 | 77 (51%) | 114 | 25 (22%) | 29% | $3.5\times10^{-6}$ |
| | $> 270$ | 50 | 20 (40%) | 21 | 0 (0%) | 40% | $1.7\times10^{-3}$ |

confidence intervals for the bootstrap $p$-values from the B-score method with $1,000$ repeated bootstrap tests (each with $B = 2,000$ bootstrap samples), which are also shown in Table 3.

From Table 3, one can see that, for both EREG and LDH, all methods suggest rejecting the null at the 5% nominal level; the $p$-values from MP and A-Score methods are, however, more significant than that from the B-Score methods. For ALKPH, the B-Score method fails to reject $H_0$ at the 5% level, in contrast to the other two methods. These results are consistent with the simulation results in Section 5 which demonstrated that MP and A-Score methods are more liberal than the B-Score method. Note that the confidence intervals for the $p$-values from the B-Score method lead to the same conclusions.

To make sense of the above results from testing $H_0 : \lambda_0 = 0$, we show in Table 4 the results from the *explanatory* subgroup analysis based on each biomarker. Specifically, we first obtained the MLE estimate $\tilde{c}_n$ of the cutpoint for each of the potential biomarkers through (2.4), i.e., without assuming $\lambda_0 = 0$, which is 7.43 for EREG, 992 for LDH, and 270 for ALKPH; these cutpoints are shown by the vertical dash lines in Figure 2 relative to the sampling distributions of the biomarkers. Next, for each biomarker, based on the estimated cutpoint, patients are divided into two subgroups; further, for each subgroup, we counted the number of patients (shown in the Total # column) in each of the two treatment groups ("Cetuximab + BSC" and "BSC"), and also calculated the number and proportion (in %) of patients with clinical benefit. For example, there were 116 patients in the subgroup with EREG $\leq 7.43$ who received the treatment with cetuximab+BSC, and 65 (56%) of them had clinical benefit [2]. Finally, for each subgroup, we computed the difference in the proportions (in %) of patients with clinical benefits between patients treated with cetuximab+BSC and BSC alone (shown in the Treat Effect column) with a positive difference indicating that the treatment with Cetuximab plus BSC is "more beneficial" than that with BSC alone, and the $p$-value from the chi-square test for the difference. From Table 4, we observe that the absolute difference in the treatment effect between the subgroups was the smallest for ALKPH among the three biomarkers (11% in comparison with 45% and 26% respectively for EREG and LDH). Since, from the B-Score tests (see Table 3), the difference in treatment effect was significant at

[2]Due to missing data in the biomarkers, the number of patients corresponding to different biomarkers are not the same.

5% level for EREG and LDH, we may conclude that patients with EREG $\leq 7.43$ or LDH $> 992$ would have more clinical benefits when treated by centuximab and BSC than by BSC alone, while there is no additional or smaller clinical benefit from centuximab and BSC for patients with EREG $> 7.43$ or LDH $\leq 992$. There is only a marginally higher clinical benefit from cetuximab and BSC for patients with ALKPH $> 270$ than those with ALKPH $\leq 270$ because there is only a trend to significance at 5% level for the difference in treatment effect between the subgroups defined by ALKPH from the B-Score tests.

## 7. Discussions

In this section, we discuss potential future work. First, in this work, we consider the use of a single unknown cutpoint on a continuous biomarker to classify patients into two subgroups; however, in many applications, it may be desirable to have more than two subgroups and thus multiple cutpoints. In the recent literature on *prognostic* classification problems, this issue has been studied under various models such as accelerated failure time models [25] and change plane models [28]. The procedures proposed in this work may be extended to *predictive* classification with multiple cutpoints but there are several challenges to implement the procedures and investigate their theoretical properties. Specifically, assume that $\boldsymbol{W} = 1$ and there are two cutpoints. In this case, the response $Y$ has a $\nu$-density given by $\exp(y\iota_0 - \phi(\iota_0))$, where

$$\iota_0 = \alpha_0 + \beta_0 U + (\gamma_{0,1} + \lambda_{0,1} U) X_{c_{0,1}} + (\gamma_{0,2} + \lambda_{0,2} U)(X_{c_{0,2}} - X_{c_{0,1}})$$

with $X_c = I(X \leq c)$ and $c_{0,1} < c_{0,2}$. To test $H_0 : \lambda_{0,1} = \lambda_{0,2} = 0$, which is the goal of *predictive* classification, one approach would be to extend the profile score test statistic $S_n$ in (2.3) as follows:

$$S_n = \|n^{-1/2} \sum_{i=1}^{n} (Y_i - \phi'(\hat{\iota}_i)) U_i [X_{i,\hat{c}_{n,1}}, X_{i,\hat{c}_{n,2}} - X_{\hat{c}_{n,1}}]^T \|^2,$$

where $\hat{\iota}_i = \hat{\alpha}_n + \hat{\beta}_n U_i + \hat{\gamma}_{n,1} X_{i,\hat{c}_{n,1}} + \hat{\gamma}_{n,2}(X_{i,\hat{c}_{n,2}} - X_{i,\hat{c}_{n,1}})$, and $(\hat{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_{n,1}, \hat{\gamma}_{n,2})$ and $(\hat{c}_{n,1}, \hat{c}_{n,2})$ are the *joint* maximum likelihood estimators (MLE) under the null. The critical value for $S_n$ could be obtained by modifying the $m$-out-of-$n$ bootstrap procedure in Section 3 but its computation may be more expensive because of the need to compute joint MLEs. The identifiability issue becomes more complicated, since it could happen that neither or one of $c_{0,1}$ and $c_{0,2}$ is identifiable, which makes theoretical investigations more difficult. The number of cutpoints may also be unknown and need to be estimated. These problems would be interesting topics for future research.

Other interesting directions for future research include the case where there are multiple biomarkers which need to be combined for *predictive* classification, and as discussed in Section 4, the slow convergence rate of the profile Wald-type test statistics.

## Appendix A: Triangular array setup for size analysis

Let $P$ be the joint distribution of $(Y, \boldsymbol{W}, U, X)$ on the observation space $S := S_0 \times \{\boldsymbol{w} \in \mathbb{R}^d : \|\boldsymbol{w}\| \leq C_w\} \times \{0,1\} \times \mathbb{R}$, where $S_0$ is the response space. The conditional density of $Y$ given $(\boldsymbol{W}, U, X)$ is given in (1.1), relative to some $\sigma$-finite measure $\nu$. From [31] it is known that $\mathrm{E}[Y|\boldsymbol{W}, U, X] = \phi'(\boldsymbol{\eta}_0^T \boldsymbol{Z}_{c_0})$, $\mathrm{VAR}[Y|\boldsymbol{W}, U, X] = \phi''(\boldsymbol{\eta}_0^T \boldsymbol{Z}_{c_0})$, where $\boldsymbol{Z}_c = (\boldsymbol{W}^T, U, X_c)^T$ and $X_c = I(X \leq c)$, and $\phi''(t) > 0$ for all $t \in \mathbb{R}$. Further, $\phi(\cdot)$ is an infinitely differentiable convex function on $\mathbb{R}$.

In this section, we consider a triangular array setup that will be applied to both the MLE based on the original data and the bootstrapped data. For each $n \in \mathbb{N}$, let $\boldsymbol{\mathcal{D}}_{n,i} = (Y_{n,i}, \boldsymbol{W}_{n,i}, U_{n,i}, X_{n,i}),\ i \in [m_n]$ be a random sample from a distribution $Q_n$, defined on the common underlying probability space $(\Omega, \mathcal{G}, \mathrm{pr})$, where $m_n \to \infty$. Assume that conditional on $(\boldsymbol{W}_{n,i}, U_{n,i}, X_{n,i}) = (\boldsymbol{w}, u, x)$, the $\nu$-density of $Y_{n,i} = y$ is

$$\exp\left(y\left(\boldsymbol{\eta}_n^T \boldsymbol{z}_{c_n}\right) - \phi(\boldsymbol{\eta}_n^T \boldsymbol{z}_{c_n})\right), \tag{A.1}$$

where $\boldsymbol{\eta}_n^T \boldsymbol{z}_{c_n} = \boldsymbol{\alpha}_n^T \boldsymbol{w} + \beta_n u + \gamma_n x_{c_n}$, $\boldsymbol{z}_c$ is defined after (1.1), $\boldsymbol{\eta}_n = (\boldsymbol{\alpha}_n^T, \beta_n, \gamma_n)^T \in \mathbb{R}^{d+2}$ and $c_n \in [\ell, u]$. Note that compared to (1.1), we set $\lambda_0 = 0$ in (A.1). Also, $\mathrm{E}[Y_{n,1}|\boldsymbol{W}_{n,1}, U_{n,1}, X_{n,1}] = \phi'(\boldsymbol{\eta}_n^T \boldsymbol{Z}_{n,1,c_n})$, where $\boldsymbol{Z}_{n,1,c_n} = (\boldsymbol{W}_{n,1}^T, U_{n,1}, X_{n,1,c_n})^T$ and $X_{n,1,c_n} = I(X_{n,1} \leq c_n)$.

Denote by $\mathbb{Q}_n^* := m_n^{-1} \sum_{i=1}^{m_n} \delta_{\boldsymbol{\mathcal{D}}_{n,i}}$ the empirical measure on $S$ induced by $\boldsymbol{\mathcal{D}}_{n,i}, i \in [m_n]$, where $\delta_{\boldsymbol{\mathcal{D}}_{n,i}}$ is the Dirac measure at $\boldsymbol{\mathcal{D}}_{n,i}$. The MLE $(\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*)$, based on $\boldsymbol{\mathcal{D}}_{n,i}, i \in [m_n]$, is defined to be

$$(\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*) := \underset{(c,\boldsymbol{\eta}) \in [\ell,u] \times \mathbb{R}^{d+2}}{\mathrm{sargmax}} \mathbb{Q}_n^* \varphi_{c,\boldsymbol{\eta}} = \underset{(c,\boldsymbol{\eta}) \in [\ell,u] \times \mathbb{R}^{d+2}}{\mathrm{sargmax}} \frac{1}{m_n} \sum_{i=1}^{m_n} \varphi_{c,\boldsymbol{\eta}}(\boldsymbol{\mathcal{D}}_{n,i}),$$

where we recall $\varphi_{c,\boldsymbol{\eta}}(y, w, u, x) = y(\boldsymbol{\eta}^T \boldsymbol{z}_c) - \phi(\boldsymbol{\eta}^T \boldsymbol{z}_c)$ in (3.1), and for an arbitrary distribution $Q$ on $S$ and a function $f : S \to \mathbb{R}$, denote by $Qf = \int f(x) Q(dx)$ as long as the integral is well defined.

Define two semi-metrics on $[\ell, u] \times \mathbb{R}^{d+2}$, $d_0((c_1, \boldsymbol{\eta}_1), (c_2, \boldsymbol{\eta}_2)) = \sqrt{|c_1 - c_2|} + \|\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2\|$ and $d_1((c_1, \boldsymbol{\eta}_1), (c_2, \boldsymbol{\eta}_2)) = \|\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2\|$, where $\|\cdot\|$ is the Euclidean norm. In this section, we establish the consistency and the convergence rate of $(\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*)$ to $(c_n, \boldsymbol{\eta}_n)$, in terms of $d_0$ for the identifiable case (i.e. $\gamma_0 \neq 0$), and $d_1$ for the non-identifiable case (i.e. $\gamma_0 = 0$). Further, we derive the limiting distribution of $(\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*)$ under the triangular array setup.

**Additional Notations.** For two sequences of random variables $\{A_n : n \in \mathbb{N}\}$ and $\{B_n : n \in \mathbb{N}\}$, we write $A_n = O_{\mathrm{pr}}(B_n)$ (resp. $A_n = o_{\mathrm{pr}}(B_n)$) if $A_n/B_n$ is bounded (resp. converges to zero) in probability. If $\{A_n\}$ and $\{B_n\}$ are in fact deterministic, we omit the subscript pr.

Let $f, F$ be two real-valued functions on $S$, and $\mathcal{F}$ a collection of functions on $S$. $F$ is said to be an envelope function for $\mathcal{F}$ if $|g(s)| \leq F(s)$ for all $g \in \mathcal{F}, s \in S$. Further, define $\mathcal{F} - f := \{g - f : g \in \mathcal{F}\}$, and for any $\delta > 0$,

$$\mathcal{F}_\delta = \{\varphi_{c,\boldsymbol{\eta}} : (c, \boldsymbol{\eta}) \in K_\delta\}, \text{ with } K_\delta = \{(c, \boldsymbol{\eta}) : c \in [\ell, u], \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| \leq \delta\}. \tag{A.2}$$

For a distribution $Q$ on $S$, define $\|Q\|_{\mathcal{F}} = \sup_{f\in\mathcal{F}} |Qf|$, and $\|f\|_{Q,2} = (Q[|f|^2])^{1/2}$.

Let $N(\xi, \mathcal{F}, d)$ denotes the $\xi$ covering number of the set $\mathcal{F}$ under the semi-metric $d$. Following [42, Page 239], for any $\epsilon > 0$, define the uniform entropy $J(\epsilon, \mathcal{F}, F)$ for the class $\mathcal{F}$ with an envelope function $F$ as follows

$$J(\epsilon, \mathcal{F}, F) = \sup_{Q} \int_0^{\epsilon} \sqrt{1 + \log N(\xi\|F\|_{Q,2}, \mathcal{F}, L_2(Q))}d\xi, \tag{A.3}$$

where the supremum is taken over all discrete probability measures $Q$ with $\|F\|_{Q,2} > 0$.

All random variables $\boldsymbol{\mathcal{D}}_{n,i},\ i \in [m_n]$ are defined on the underlying probability space $(\Omega, \mathcal{G}, \mathrm{pr})$. Since we will work with distributions on the space $S$, we introduce the following measurable *coordinate* mappings on $S$:

$$\mathcal{Y}(y, \boldsymbol{w}, u, x) = y,\ \ \boldsymbol{\mathcal{W}}(y, \boldsymbol{w}, u, x) = \boldsymbol{w},\ \ \mathcal{U}(y, \boldsymbol{w}, u, x) = u,\ \ \mathcal{X}(y, \boldsymbol{w}, u, x) = x,$$
$$\mathcal{X}_c = I(\mathcal{X} \le c), \quad \boldsymbol{\mathcal{Z}}_c = (\boldsymbol{\mathcal{W}}^T, \mathcal{U},\ \mathcal{X}_c)^T, \quad \tilde{\boldsymbol{\mathcal{Z}}}_c = (\boldsymbol{\mathcal{Z}}_c^T, \mathcal{U}\mathcal{X}_c)^T,$$

which are random variables (vectors) on the space $S$, and we can use operator notations such as $Q[\mathcal{X}_c]$ for $\int I(x \le c)Q(dx)$, where $Q$ is a distribution on $S$.

Let $T$ be an arbitrary index set, and for a function $f : T \to \mathbb{R}$, denote its $\ell_\infty$ norm by $\|f\|_\infty = \sup_{t\in T} |f(t)|$. Denote by $\ell^\infty(T)$ the space of uniformly bounded, real-valued functions on $T$ equipped with $\ell_\infty$ norm. Let $Z_n = \{Z_{n,t} : t \in T\},\ n \ge 1$ be a sequence of random processes indexed by $T$, for which $\|Z_n\|_\infty < \infty$ for each $n \ge 1$ almost surely, and $Z$ be a tight random element in $\ell^\infty(T)$. Then as in [41], we say $Z_n,\ n \ge 1$ *converges weakly* to $Z$, or $Z_n \rightsquigarrow Z$, in $\ell^\infty(T)$ if $\mathrm{E}[g(Z_n)] \to \mathrm{E}[g(Z)]$, as $n \to \infty$, for any bounded and continuous function $g : \ell^\infty(T) \to \mathbb{R}$. [3]

### *A.1. Assumptions for the triangular array setup*

We make the following assumptions for the triangular array setup.

(A.I) For $j \in \{0, 1\}$,

$$0 < \liminf_n \mathrm{E}[U_{n,1}] \le \limsup_n \mathrm{E}[U_{n,1}] < 1,$$
$$0 < \liminf_n \mathrm{pr}(X_{n,1} \le \ell, U_{n,1} = j) \le \limsup_n \mathrm{pr}(X_{n,1} \le u, U_{n,1} = j) < 1,$$
$$\sup_{|c_1-c_2|\le\delta_n} \mathrm{pr}(c_1 \wedge c_2 < X_{n,1} \le c_1 \vee c_2) \to 0 \quad \text{for every } \delta_n \to 0.$$
$$\liminf_n \lambda_{\min}(\mathrm{E}[\boldsymbol{W}_{n,1}\boldsymbol{W}_{n,1}^T \,|\, U_{n,1} = j, X_{n,1} > u]) > 0,$$
$$\liminf_n \lambda_{\min}(\mathrm{E}[\boldsymbol{W}_{n,1}\boldsymbol{W}_{n,1}^T \,|\, U_{n,1} = j, X_{n,1} \le \ell]) > 0, \quad \sup_n \|\boldsymbol{W}_{n,1}\| \le C_w.$$

[3] The random processes $Z_n,\ n \ge 1$, viewed as maps from the underlying probability spaces to $\ell^\infty(T)$, are usually not Borel measurable, in which case the expectations are with respect to outer-probabilities. For details and a definitive treatment of functional weak convergence, we refer readers to [41].

(A.II) For any $\delta > 0$, $\|Q_n - P\|_{\mathcal{F}_\delta} \to 0$.

(A.III) For any $\delta > 0$, $(c_0, \boldsymbol{\eta}_0)$ is a well-separated maximizer of $P\varphi_{c,\boldsymbol{\eta}}$ in $K_\delta$. That is, for any $\epsilon > 0$, $P\varphi_{c,\boldsymbol{\eta}} < P\varphi_{c_0,\boldsymbol{\eta}_0}$ for all $(c, \boldsymbol{\eta}) \in K_\delta$ such that $d_\iota((c, \boldsymbol{\eta}), (c_0, \boldsymbol{\eta}_0)) > \epsilon$, where $\iota \in \{0, 1\}$ needs to be specified.

(A.IV) $d_\iota((c_n, \boldsymbol{\eta}_n), (c_0, \boldsymbol{\eta}_0)) \to 0$, where $\iota \in \{0, 1\}$ needs to be specified.

*If* $\gamma_0 \neq 0$, we further need the next three conditions:

(A.V) For some $\epsilon \in [0, 1/4)$:

$$\begin{aligned} 0 \ &< \liminf_{n\to\infty} \inf_{n^{-1+2\epsilon} \le |c-c_n| < 1} \left\{ \frac{1}{|c - c_n|} Q_n[I(c \wedge c_n < \mathcal{X} \le c \vee c_n)] \right\} \\ &\le \limsup_{n\to\infty} \sup_{n^{-1+2\epsilon} \le |c-c_n| < 1} \left\{ \frac{1}{|c - c_n|} Q_n[I(c \wedge c_n < \mathcal{X} \le c \vee c_n)] \right\} < \infty. \end{aligned}$$

(A.VI) For any $\delta \in [0, 1/4)$:

$$\sqrt{m_n} Q_n[I(c_n - m_n^{-1+2\delta} < \mathcal{X} < c_n + m_n^{-1+2\delta})] \to 0.$$

(A.VII) For any $(c_1, \boldsymbol{\eta}_1), (c_2, \boldsymbol{\eta}_2) \in [\ell, u] \times \mathbb{R}^{d+2}$, $Q_n[(\mathcal{Y} - \phi'(\boldsymbol{\eta}_1^T \boldsymbol{\mathcal{Z}}_{c_1}))(\mathcal{Y} - \phi'(\boldsymbol{\eta}_2^T \boldsymbol{\mathcal{Z}}_{c_2}))\tilde{\boldsymbol{\mathcal{Z}}}_{c_1} \tilde{\boldsymbol{\mathcal{Z}}}_{c_2}^T]$ converges as $n \to \infty$. Further,

$$\lim_{n\to\infty} Q_n[\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})\tilde{\boldsymbol{\mathcal{Z}}}_{c_n} \tilde{\boldsymbol{\mathcal{Z}}}_{c_n}^T] = P[\phi''(\boldsymbol{\eta}_0^T \boldsymbol{\mathcal{Z}}_{c_0})\tilde{\boldsymbol{\mathcal{Z}}}_{c_0} \tilde{\boldsymbol{\mathcal{Z}}}_{c_0}^T].$$

*If* $\gamma_0 = 0$, we need the following two conditions:

(A.VIII) For some constant $B_1$, $\lim_{n\to\infty} \sqrt{m_n}\gamma_n = B_1$. Further, if $B_1 \neq 0$, assume that $\lim_{n\to\infty} c_n = B_2$ for some $B_2 \in [\ell, u]$.

(A.IX) Uniformly over $(c_1, c_2, c_3) \in [\ell, u]^3$,

$$\lim_{n\to\infty} Q_n[\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_3})\tilde{\boldsymbol{\mathcal{Z}}}_{c_1} \tilde{\boldsymbol{\mathcal{Z}}}_{c_2}^T] = P[\phi''(\boldsymbol{\eta}_0^T \boldsymbol{\mathcal{Z}}_{c_3})\tilde{\boldsymbol{\mathcal{Z}}}_{c_1} \tilde{\boldsymbol{\mathcal{Z}}}_{c_2}^T].$$

### *A.2. Consistency under triangular array setup*

**Theorem A.1.** *Let* $\iota \in \{0, 1\}$*. Assume that* (A.I)-(A.III) *with* $\iota$ *hold, and that* $\sup_n \|\boldsymbol{\eta}_n\| < \infty$*. Then* $d_\iota((\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*), (c_0, \boldsymbol{\eta}_0)) \to 0$ *in probability.*

*Proof.* From Lemma A.2, $(\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*)$ is uniformly tight. Besides, for any $\delta > 0$, $(c_0, \boldsymbol{\eta}_0)$ is a well-separated maximizer of $P\varphi_{c,\boldsymbol{\eta}}$ in $K_\delta$ from condition (A.III). Therefore by [42, Corollary 3.2.3 (ii)], it suffices to show that $\|\mathbb{Q}_n^* - P\|_{\mathcal{F}_\delta} \to 0$ in probability for any $\delta > 0$. Note the following decomposition

$$\|\mathbb{Q}_n^* - P\|_{\mathcal{F}_\delta} \le \|\mathbb{Q}_n^* - Q_n\|_{\mathcal{F}_\delta} + \|Q_n - P\|_{\mathcal{F}_\delta},$$

where the second term converges to zero by condition (A.II). For the first term, by [42, Theorem 2.14.1], there exists an absolute constant $C > 0$, such that

$$\mathrm{E}\|\mathbb{Q}_n^* - Q_n\|_{\mathcal{F}_\delta} \;\le\; \frac{CJ(1, \mathcal{F}_\delta, F_\delta)}{\sqrt{m_n}} \sqrt{Q_n[F_\delta^2]},$$

where $F_\delta(\mathcal{Y}, \boldsymbol{\mathcal{W}}, \mathcal{U}, \mathcal{X}_c) = C_{1,\delta}|\mathcal{Y}| + C_{2,\delta}$ is defined before Lemma A.1. By Lemma A.1, $J(1, \mathcal{F}_\delta, F_\delta) < \infty$. Further, since $\sup_n \|\boldsymbol{\eta}_n\| < \infty$ and $\|\boldsymbol{\mathcal{Z}}_c\| \le 2 + C_w$ for any $c \in [\ell, u]$, we have

$$\sup_n Q_n[\mathcal{Y}^2] = \sup_n \left(\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}) + (\phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}))^2\right) < \infty, \tag{A.4}$$

which implies $\sup_n \sqrt{Q_n[F_\delta^2]} < \infty$. As a result, $\|\mathbb{Q}_n^* - Q_n\|_{\mathcal{F}_\delta} \to 0$ in probability. Then due to [42, Corollay 3.2.3 (ii)], we obtain that $d_\iota((\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*), (c_0, \boldsymbol{\eta}_0)) \to 0$ in probability. □

Recall the function class $\mathcal{F}_\delta$ and the set $K_\delta$ defined in (A.2), and $C_w$ in condition (A.I). For any $\delta > 0$, define

$$C_{1,\delta} = \|\boldsymbol{\alpha}_0\| C_w + |\beta_0| + |\gamma_0| + (2 + C_w)\delta, \quad C_{2,\delta} = \sup_{|t| \le C_{1,\delta}} \left(|\phi(t)| + |\phi'(t)|\right).$$

Since $\phi$ is infinitely differentiable on $\mathbb{R}$, for any $\delta > 0$, we have $C_{2,\delta} < \infty$ and

$$F_\delta(\mathcal{Y}, \boldsymbol{\mathcal{W}}, \mathcal{U}, \mathcal{X}_c) = C_{1,\delta}|\mathcal{Y}| + C_{2,\delta},$$

is an envelope function for $\mathcal{F}_\delta$.

**Lemma A.1.** *Under* (A.I)*,* $J(1, \mathcal{F}_\delta, F_\delta) < \infty$ *for any* $\delta > 0$*.*

*Proof.* For any $\delta > 0$, define a function class $\mathcal{F}_{1,\delta}$ on $S$ as follows:

$$\mathcal{F}_{1,\delta} = \{S \ni (y, \boldsymbol{w}, u, x) \to \boldsymbol{\alpha}^T \boldsymbol{w} + \beta u + \gamma I(x \le c) : (c, \boldsymbol{\eta}) \in K_\delta\}.$$

From [11, Definition 2.1], $\mathcal{F}_{1,\delta}$ is a VC-type class with the constant envelope function $C_{1,\delta}$. Further, the class $\{\mathcal{Y}\}$ is a single function, and thus is a VC type class with the envelope function $|\mathcal{Y}|$.

Now define $g : \mathbb{R}^2 \to \mathbb{R}$ by $g(a_1, a_2) = a_1 a_2 - \phi(a_1)$. Then we have

$$\mathcal{F}_\delta \subset g(\mathcal{F}_{1,\delta}, \{\mathcal{Y}\}) := \{f_{1,\delta}\mathcal{Y} - \phi(f_{1,\delta}) : \; f_{1,\delta} \in \mathcal{F}_{1,\delta}\}.$$

Observe that for any $f_{1,\delta}, f'_{1,\delta} \in \mathcal{F}_{1,\delta}$,

$$\begin{aligned} &\left|g \circ (f_{1,\delta}, \mathcal{Y})(y, \boldsymbol{w}, u, x) - g \circ (f'_{1,\delta}, \mathcal{Y})(y, \boldsymbol{w}, u, x)\right| \\ &\le (|y| + C_{2,\delta})|f_{1,\delta}(y, \boldsymbol{w}, u, x) - f'_{1,\delta}(y, \boldsymbol{w}, u, x)|. \end{aligned}$$

By [11, Lemma A.6], $\mathcal{F}_\delta$ is of VC type with the envelope function $F_\delta$, which completes the proof due to the calculation in the proof of [11, Corollary 5.1]. □

**Lemma A.2.** *Assume that* (A.I) *holds, and* $\sup_n \|\boldsymbol{\eta}_n\| < \infty$*. Then* $(\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*) = O_{\mathrm{pr}}(1)$*.*

*Proof.* The tightness of $\hat{c}_n^*$ holds since it is restricted in $[\ell, u]$, and in what follows we focus on $\hat{\boldsymbol{\eta}}_n^*$. Since $(\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*)$ maximizes $\mathbb{Q}_n^* \varphi_{c,\boldsymbol{\eta}}$, we have that $\mathbb{Q}_n^*[\varphi_{\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*} - \varphi_{c_n, \boldsymbol{\eta}_n}] \geq 0$, or equivalently

$$\mathbb{Q}_n^*[\mathcal{Y}((\hat{\boldsymbol{\eta}}_n^*)^T \boldsymbol{\mathcal{Z}}_{\hat{c}_n^*} - \boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}) - (\phi((\hat{\boldsymbol{\eta}}_n^*)^T \boldsymbol{\mathcal{Z}}_{\hat{c}_n^*}) - \phi(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}))] \geq 0.$$

Therefore we have $II_n \leq |I_n^{(1)}| + |II_n^{(2)}|$, where

$$I_n^{(1)} = (\hat{\boldsymbol{\eta}}_n^*)^T \mathbb{Q}_n^*[(\mathcal{Y} - \phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}))\boldsymbol{\mathcal{Z}}_{\hat{c}_n^*}], \quad I_n^{(2)} = \boldsymbol{\eta}_n^T \mathbb{Q}_n^*[(\mathcal{Y} - \phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}))\boldsymbol{\mathcal{Z}}_{c_n}]$$
$$II_n = \mathbb{Q}_n^*[\phi((\hat{\boldsymbol{\eta}}_n^*)^T \boldsymbol{\mathcal{Z}}_{\hat{c}_n^*}) - \phi(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}) - \phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})((\hat{\boldsymbol{\eta}}_n^*)^T \boldsymbol{\mathcal{Z}}_{\hat{c}_n^*} - \boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})].$$

By Lemma A.3 (below) and since $\sup_n \|\boldsymbol{\eta}_n\| < \infty$, we have

$$|I_n^{(1)}| + |I_n^{(2)}| = (1 + \|\hat{\boldsymbol{\eta}}_n^*\|) o_{\text{pr}}(1). \tag{A.5}$$

Next we consider the term $II_n$. Since $\phi$ is convex,

$$\hat{\phi}_n := \phi((\hat{\boldsymbol{\eta}}_n^*)^T \boldsymbol{\mathcal{Z}}_{\hat{c}_n^*}) - \phi(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}) - \phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})((\hat{\boldsymbol{\eta}}_n^*)^T \boldsymbol{\mathcal{Z}}_{\hat{c}_n^*} - \boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}) \geq 0.$$

which implies that

$$II_n \geq II_n^{(1)} + II_n^{(2)} + II_n^{(3)}, \quad \text{where} \quad II_n^{(1)} = \mathbb{Q}_n^*[\hat{\phi}_n I(\mathcal{U} = 0, \mathcal{X} \leq \ell)],$$
$$II_n^{(2)} = \mathbb{Q}_n^*[\hat{\phi}_n I(\mathcal{U} = 1, \mathcal{X} \leq \ell)], \quad II_n^{(3)} = \mathbb{Q}_n^*[\hat{\phi}_n I(\mathcal{U} = 0, \mathcal{X} \geq u)].$$

Let $C_0, C_1$ be the constants in Lemma A.5 (below). Define an event in the observation space $S$:

$$\mathcal{A} = \left\{ \left| \frac{(\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n)^T}{\|\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n\|} \boldsymbol{\mathcal{W}} \right| \geq C_0^{-1}, \|\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n\| \geq 1, \mathcal{U} = 0, \mathcal{X} \leq \ell \right\}.$$

If $\mathcal{A}$ occurs, then $\mathcal{X}_c = 0$ for $c \in [\ell, u]$, and thus

$$\hat{\phi}_n \geq \kappa(K, C_0^{-1}) \left| (\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n)^T \boldsymbol{\mathcal{W}} \right| \geq \kappa(K, C_0^{-1}) C_0^{-1} \|\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n\|,$$

where $K = C_w \sup_n \|\boldsymbol{\eta}_n\| < \infty$, and $\kappa(K, C_0^{-1}) > 0$ is defined in Lemma A.4 (below). Thus there exists some positive constant $C$ such that

$$\begin{aligned}
&C II_n^{(1)} = C \mathbb{Q}_n^*[\hat{\phi}_n(\mathcal{U} = 0, \mathcal{X} \leq \ell)] \\
&\geq \|\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n\| I(\|\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n\| \geq 1) \mathbb{Q}_n^* \left[ I \left( \left| \frac{(\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n)^T}{\|\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n\|} \boldsymbol{\mathcal{W}} \right| \geq C_0^{-1}, \mathcal{U} = 0, \mathcal{X} \leq \ell \right) \right] \\
&\geq \|\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n\| I(\|\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n\| \geq 1) \inf_{\boldsymbol{\theta} \in \mathcal{S}^{d-1}} \mathbb{Q}_n^*[I(|\boldsymbol{\theta}^T \boldsymbol{\mathcal{W}}| \geq C_0^{-1}, \mathcal{U} = 0, \mathcal{X} \leq \ell)] \\
&\geq \|\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n\| I(\|\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n\| \geq 1)(C_1^{-1} + o_{\text{pr}}(1)),
\end{aligned}$$

where the last inequality is due to Lemma A.5.

Further, if $\|\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n\| C_w \leq 2^{-1} |\hat{\beta}_n^* - \beta_n|$, $\mathcal{U} = 1$, $\mathcal{X} \leq \ell$, then

$$|(\hat{\boldsymbol{\eta}}_n^*)^T \boldsymbol{\mathcal{Z}}_{\hat{c}_n^*} - \boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}| \geq |\hat{\beta}_n^* - \beta_n| - \|\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n\| C_w \geq 2^{-1} |\hat{\beta}_n^* - \beta_n|.$$

If $\|\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n\| C_w \le 2^{-1}|\hat{\gamma}_n^* - \gamma_n|$, $\mathcal{U} = 0$, $\mathcal{X} \ge u$, then

$$|(\hat{\boldsymbol{\eta}}_n^*)^T \boldsymbol{\mathcal{Z}}_{\hat{c}_n^*} - \boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}| \ge |\hat{\gamma}_n^* - \gamma_n| - \|\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n\| C_w \ge 2^{-1}|\hat{\gamma}_n^* - \gamma_n|.$$

Then by a similar argument as for $II_n^{(1)}$, there exist positive constants $C_2$ and $C_3$ such that

$$II_n^{(2)} \ge (C_2^{-1} + o_{\mathrm{pr}}(1))|\hat{\beta}_n^* - \beta_n| I(|\hat{\beta}_n^* - \beta_n| \ge \max\{2\|\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n\| C_w, 2\}),$$
$$II_n^{(3)} \ge (C_3^{-1} + o_{\mathrm{pr}}(1))|\hat{\gamma}_n^* - \gamma_n| I(|\hat{\gamma}_n^* - \gamma_n| \ge \max\{2\|\hat{\boldsymbol{\alpha}}_n^* - \boldsymbol{\alpha}_n\| C_w, 2\}).$$

Combining the above three cases, and due to (A.5), we have that there exists a positive constant $C_4$ such that

$$(C_4^{-1} + o_{\mathrm{pr}}(1))\|\hat{\boldsymbol{\eta}}_n^* - \boldsymbol{\eta}_n\| I(\|\hat{\boldsymbol{\eta}}_n^* - \boldsymbol{\eta}_n\| \ge 8C_w + 6) \le II_n \le (1 + \|\hat{\boldsymbol{\eta}}_n^*\|) o_{\mathrm{pr}}(1).$$

This completes the proof for $\|\hat{\boldsymbol{\eta}}_n^* - \boldsymbol{\eta}_n\| = O_{\mathrm{pr}}(1)$, since $\sup_n \|\boldsymbol{\eta}_n\| < \infty$. □

**Lemma A.3.** *Assume that* (A.I) *holds, and that* $\sup_n \|\boldsymbol{\eta}_n\| < \infty$. *Then*

$$\sup_{c \in [\ell, u]} \|\mathbb{Q}_n^*[(\mathcal{Y} - \phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}))\boldsymbol{\mathcal{Z}}_c]\| \ = \ o_{\mathrm{pr}}(1).$$

*Proof.* Let $\delta := \sup_n \|\boldsymbol{\eta}_n\| < \infty$, and define

$$\mathcal{F}_{2,\delta} = \{S \ni (y, \boldsymbol{w}, u, x) \mapsto (y - \phi'(\boldsymbol{\eta}^T \boldsymbol{z}_{c_1}))\boldsymbol{z}_{c_2} : \ (c_1, c_2) \in [\ell, u]^2, \ \|\boldsymbol{\eta}\| \le \delta\}.$$

By a similar argument as in Lemma A.1, there exists a constant $C$, that depends on $\delta$, such that $F_{2,\delta} = C(|\mathcal{Y}| + 1)$ is an envelope function for $\mathcal{F}_{2,\delta}$, and that $J(1, \mathcal{F}_{2,\delta}, F_{2,\delta}) < \infty$. Then due to [42, Theorem 2.14.1], for some absolute constant $C'$,

$$\begin{aligned}
&\mathrm{E}\left[\sup_{c \in [\ell, u]} \|\mathbb{Q}_n^*[(\mathcal{Y} - \phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}))\boldsymbol{\mathcal{Z}}_c]\|\right] \\
=&\mathrm{E}\left[\sup_{c \in [\ell, u]} \|(\mathbb{Q}_n^* - Q_n)[(\mathcal{Y} - \phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}))\boldsymbol{\mathcal{Z}}_c]\|\right] \\
\le& \mathrm{E}\|\mathbb{Q}_n^* - Q_n\|_{\mathcal{F}_{2,\delta}} \le \sup_n \frac{C' J(1, \mathcal{F}_{2,\delta}, F_{2,\delta})}{\sqrt{m_n}} \sqrt{Q_n F_{2,\delta}^2}.
\end{aligned}$$

Then the proof is complete due to (A.4) and condition (A.I). □

**Lemma A.4.** *For any* $K > 0$ *and* $\delta > 0$,

$$\kappa(K, \delta) := \inf_{|y - x| \ge \delta, |x| \le K} \left| \frac{\phi(y) - \phi(x)}{y - x} - \phi'(x) \right| > 0.$$

*Proof.* Since $\phi$ is convex and infinitely differentiable on $\mathbb{R}$, for any $y \ge x + \delta$, we have

$$\frac{\phi(y) - \phi(x)}{y - x} - \phi'(x) \ge \frac{\phi(x + \delta) - \phi(x)}{\delta} - \phi'(x) = \frac{1}{2}\delta\phi''(x + t\delta),$$

for some $t \in (0,1)$ by the mean value theorem, and thus

$$\inf_{y \geq x+\delta, |x|<K} \frac{\phi(y)-\phi(x)}{y-x} - \phi'(x) \geq \frac{1}{2}\delta \inf_{|x|\leq K+\delta} \phi''(x) > 0,$$

where the last inequality is because $\phi''$ is continuous and $\phi''(t) > 0$ for any $t \in \mathbb{R}$. The case where $y \leq x - \delta$ is similar and omitted. □

Denote by $\mathcal{S}^{d-1}$ the unit sphere in $\mathbb{R}^d$, i.e. $\mathcal{S}^{d-1} = \{\boldsymbol{s} \in \mathbb{R}^d : \ \|\boldsymbol{s}\| = 1\}$.

**Lemma A.5.** *Assume* (A.I) *holds. Then there exist positive constants* $C_0, C_1$ *such that*

$$\inf_{\boldsymbol{\theta} \in \mathcal{S}^{d-1}} \mathbb{Q}_n^*[I(|\boldsymbol{\theta}^T \boldsymbol{\mathcal{W}}| \geq C_0^{-1}, \mathcal{U}=0, \mathcal{X} \leq \ell)] \geq C_1^{-1} + o_{\mathrm{pr}}(1).$$

*Proof.* By a similar argument as in Lemma A.3,

$$\sup_{\theta \in \mathcal{S}^{d-1}} \left| (\mathbb{Q}_n^* - Q_n)[I(|\boldsymbol{\theta}^T \boldsymbol{\mathcal{W}}| \geq C_0^{-1}, \mathcal{U}=0, \mathcal{X} \leq \ell)] \right| = o_{\mathrm{pr}}(1).$$

Then, due to (A.I), it suffices to show that for some constant $C_0$,

$$\liminf_n \inf_{\boldsymbol{\theta} \in \mathcal{S}^{d-1}} \mathrm{pr}(|\boldsymbol{\theta}^T \boldsymbol{W}_{n,1}| \geq C_0^{-1} \,|\, U_{n,1}=0,\ X_{n,1} \leq \ell) > 0.$$

Let $\lambda_n = \lambda_{\min}(\mathrm{E}[\boldsymbol{W}_{n,1}\boldsymbol{W}_{n,1}^T \,|\, U_{n,1}=0,\ X_{n,1} \leq \ell])$. For any $\boldsymbol{\theta} \in \mathcal{S}^{d-1}$,

$$\begin{aligned}
\lambda_n &\leq \mathrm{E}[|\boldsymbol{\theta}^T \boldsymbol{W}_{n,1}|^2 \,|\, U_{n,1}=0,\ X_{n,1} \leq \ell] \\
&= \mathrm{E}[|\boldsymbol{\theta}^T \boldsymbol{W}_{n,1}|^2 I(|\boldsymbol{\theta}^T \boldsymbol{W}_{n,1}| \leq \sqrt{\lambda_n/2}) \,|\, U_{n,1}=0,\ X_{n,1} \leq \ell] \\
&+ \mathrm{E}[|\boldsymbol{\theta}^T \boldsymbol{W}_{n,1}|^2 I(|\boldsymbol{\theta}^T \boldsymbol{W}_{n,1}| > \sqrt{\lambda_n/2}) \,|\, U_{n,1}=0,\ X_{n,1} \leq \ell] \\
&\leq \frac{\lambda_n}{2} + C_w^2 \mathrm{pr}(|\boldsymbol{\theta}^T \boldsymbol{W}_{n,1}| \geq \sqrt{\lambda_n/2} \,|\, U_{n,1}=0,\ X_{n,1} \leq \ell),
\end{aligned}$$

which implies that

$$\inf_{\boldsymbol{\theta} \in \mathcal{S}^{d-1}} \mathrm{pr}(|\boldsymbol{\theta}^T \boldsymbol{W}_{n,1}| \geq \sqrt{\lambda_n/2} \,|\, U_{n,1}=0,\ X_{n,1} \leq \ell) \geq \frac{\lambda_n}{2C_w^2}.$$

Then the proof is complete since $\liminf_n \lambda_n > 0$ due to (A.I). □

### *A.3. The identifiable case*

For the lemmas and theorems in this subsection we assume conditions (A.I)-(A.V) hold with $\iota = 0$.

*A.3.1. Convergence rates in the identifiable case*

**Theorem A.2.** *Consider the identifiable case under the null, i.e.* $\gamma_0 \neq 0$ *and* $\lambda_0 = 0$. *Let* (A.I)-(A.V) *hold. Then* $m_n^{1/2-\epsilon} d_0((\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*), (c_n, \boldsymbol{\eta}_n)) = O_{\text{pr}}(1)$, *where* $\epsilon$ *appears in condition* (A.V).

*Proof.* From conditions (A.I)-(A.IV) with $\iota = 0$ and Theorem A.1, we have that $d_0((\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*), (c_n, \boldsymbol{\eta}_n)) \to 0$ in probability. Then by [42, Theorem 3.4.1], it suffices to verify the following two conditions for some positive constant $C$ and large enough $n$:

$$\sup_{\delta/2 \leq d_0((c,\boldsymbol{\eta}),(c_n,\boldsymbol{\eta}_n)) \leq \delta} Q_n[\varphi_{c,\boldsymbol{\eta}} - \varphi_{c_n,\boldsymbol{\eta}_n}] \leq -C^{-1}\delta^2, \tag{A.6}$$

$$\mathrm{E} \sup_{\delta/2 \leq d_0((c,\boldsymbol{\eta}),(c_n,\boldsymbol{\eta}_n)) \leq \delta} \sqrt{m_n} |(\mathbb{Q}_n^* - Q_n)\varphi_{c,\boldsymbol{\eta}} - (\mathbb{Q}_n^* - Q_n)\varphi_{c_n,\boldsymbol{\eta}_n}| \leq C\delta, \tag{A.7}$$

for any $4m_n^{-1/2+\epsilon} \leq \delta \leq \delta_u$, where $\epsilon$ appears in condition (A.V), and $\delta_u$ is a constant that will be specified. Note that in the proof the value of the constant $C$ may vary from line to line.

Next we verify (A.6) and (A.7) for any fixed $\delta$ which satisfies $4m_n^{-1/2+\epsilon} \leq \delta \leq \delta_u$. We focus on the case $c < c_n$, and the case $c > c_n$ can be verified similarly.

Due to conditions (A.I) and (A.V), and since $\phi$ is infinitely differentiable on $\mathbb{R}$ and $\phi''$ is a positive function on $\mathbb{R}$, there exists some constant $C > 0$ such that for large enough $n$, $i \in \{0, 1\}$, $\|\boldsymbol{\eta} - \boldsymbol{\eta}_n\| \leq \delta$, and $c \in [\ell, u]$, we have

$$\begin{aligned}
&(|\phi'| + |\phi''| + |\phi'''|)(\boldsymbol{\eta}^T \boldsymbol{Z}_c) \leq C, \quad \phi''(\boldsymbol{\eta}^T \boldsymbol{Z}_c) \geq C^{-1}, \\
&C^{-1} \leq \frac{Q_n[I(c \wedge c_n < \mathcal{X} \leq c \vee c_n)]}{|c - c_n|} \leq C, \text{ if } m_n^{-1+2\epsilon} \leq |c - c_n| < 1 \\
&\lambda_{\min}(Q_n[\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^T | \mathcal{U} = i, \mathcal{X} \leq \ell]) > C^{-1}, \quad Q_n[I(\mathcal{U} = i, \mathcal{X} \leq \ell)] > C^{-1}, \\
&\lambda_{\min}(Q_n[\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{W}}^T | \mathcal{U} = i, \mathcal{X} > u]) > C^{-1}, \quad Q_n[I(\mathcal{U} = i, \mathcal{X} > u)] > C^{-1}.
\end{aligned} \tag{A.8}$$

Also, note that due to (A.1)

$$\mathrm{E}[Y_{n,1} \mid \boldsymbol{W}_{n,1}, U_{n,1}, X_{n,1}] = \phi'(\boldsymbol{\eta}_n^T \boldsymbol{Z}_{1,c_n}).$$

**Verifying (A.6).** Fix some $(c, \boldsymbol{\eta})$ such that $\delta/2 \leq d_0((c, \boldsymbol{\eta}), (c_n, \boldsymbol{\eta}_n)) \leq \delta$. From (A.8), and by the mean value theorem, there exists a constant $C_0$ such that

$$\begin{aligned}
&Q_n[\varphi_{c,\boldsymbol{\eta}} - \varphi_{c_n,\boldsymbol{\eta}_n}] = Q_n[\varphi_{c,\boldsymbol{\eta}_n} - \varphi_{c_n,\boldsymbol{\eta}_n} + \varphi_{c,\boldsymbol{\eta}} - \varphi_{c,\boldsymbol{\eta}_n}] \\
\leq &Q_n[\varphi_{c,\boldsymbol{\eta}_n} - \varphi_{c_n,\boldsymbol{\eta}_n}] - 2^{-1} Q_n[\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_c)(\boldsymbol{\eta} - \boldsymbol{\eta}_n)^T \boldsymbol{\mathcal{Z}}_c \boldsymbol{\mathcal{Z}}_c^T (\boldsymbol{\eta} - \boldsymbol{\eta}_n)] \\
&+ Q_n[(\phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}) - \phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_c))(\boldsymbol{\eta} - \boldsymbol{\eta}_n)^T \boldsymbol{\mathcal{Z}}_c] + C_0\delta^3 \\
= &D_1 + D_2 + D_3 + C_0\delta^3,
\end{aligned} \tag{A.9}$$

where $D_1$, $D_2$ and $D_3$ will be upper bounded separately as follows.

Upper bounding $D_1$. Due to (A.8), from the mean value theorem, $D_1$ can be upper bounded by

$$\begin{aligned} D_1 &= Q_n[\varphi_{c,\boldsymbol{\eta}_n} - \varphi_{c_n,\boldsymbol{\eta}_n}] \\ &= - Q_n\left[\left(\gamma_n \mathcal{Y} - \left(\phi(\boldsymbol{\alpha}_n^T \boldsymbol{\mathcal{W}} + \beta_n \mathcal{U} + \gamma_n) - \phi(\boldsymbol{\alpha}_n^T \boldsymbol{\mathcal{W}} + \beta_n \mathcal{U})\right)\right) I(c < \mathcal{X} \le c_n)\right] \\ &= - Q_n\left[\left(\gamma_n \phi'(\boldsymbol{\alpha}_n^T \boldsymbol{\mathcal{W}} + \beta_n \mathcal{U} + \gamma_n) - \left(\phi(\boldsymbol{\alpha}_n^T \boldsymbol{\mathcal{W}} + \beta_n \mathcal{U} + \gamma_n)\right.\right.\right. \\ &\quad - \left.\left.\left.\phi(\boldsymbol{\alpha}_n^T \boldsymbol{\mathcal{W}} + \beta_n \mathcal{U})\right)\right) I(c < \mathcal{X} \le c_n)\right] \\ &= - Q_n\left[2^{-1}\gamma_n^2 \phi''(\boldsymbol{\alpha}_n^T \boldsymbol{\mathcal{W}} + \beta_n \mathcal{U} + \tilde{\gamma}) I(c < \mathcal{X} \le c_n)\right] \\ &\le - C^{-1}\gamma_n^2 Q_n\left[I(c < \mathcal{X} \le c_n)\right], \end{aligned}$$

where $\tilde{\gamma}$ is between $\gamma_n$ and 0.

By (A.8), $Q_n[I(c < \mathcal{X} \le c_n)] \ge C^{-1}|c - c_n|$, for any $c \in [\ell, u]$, if $1 > |c - c_n| \ge m_n^{-1+2\epsilon} \ge n^{-1+2\epsilon}$. By condition (A.IV), $\gamma_n \to \gamma_0 \ne 0$. Therefore there exists a constant $C_1 > 0$ such that for large enough $n$,

$$D_1 = Q_n[\varphi_{c,\boldsymbol{\eta}_n} - \varphi_{c_n,\boldsymbol{\eta}_n}] \le -C_1^{-1}|c - c_n|, \quad \text{if } |c_n - c| \ge m_n^{-1+2\epsilon}.$$

Upper bounding $D_2$. For $D_2$ in (A.9), note that,

$$\begin{aligned} \tilde{D}_2 =& Q_n[(\boldsymbol{\eta} - \boldsymbol{\eta}_n)^T \boldsymbol{\mathcal{Z}}_c \boldsymbol{\mathcal{Z}}_c^T (\boldsymbol{\eta} - \boldsymbol{\eta}_n)] \\ =& Q_n[(\boldsymbol{\alpha} - \boldsymbol{\alpha}_n)^T \boldsymbol{\mathcal{W}} + (\beta - \beta_n)\mathcal{U} + (\gamma - \gamma_n)\mathcal{X}_c]^2 \\ \ge& Q_n\left[I(\mathcal{U} = 0, \mathcal{X}_u = 0)((\boldsymbol{\alpha} - \boldsymbol{\alpha}_n)^T \boldsymbol{\mathcal{W}})^2\right] \\ &+ Q_n\left[I(\mathcal{U} = 1, \mathcal{X}_u = 0)((\boldsymbol{\alpha} - \boldsymbol{\alpha}_n)^T \boldsymbol{\mathcal{W}} + (\beta - \beta_n))^2\right] \\ &+ Q_n\left[I(\mathcal{U} = 0, \mathcal{X}_\ell = 1)((\boldsymbol{\alpha} - \boldsymbol{\alpha}_n)^T \boldsymbol{\mathcal{W}} + (\gamma - \gamma_n))^2\right] \\ =& Q_n[I(\mathcal{U} = 0, \mathcal{X}_u = 0)] Q_n[((\boldsymbol{\alpha} - \boldsymbol{\alpha}_n)^T \boldsymbol{\mathcal{W}})^2 | \mathcal{U} = 0, \mathcal{X}_u = 0] \\ &+ Q_n[I(\mathcal{U} = 1, \mathcal{X}_u = 0)] Q_n[((\boldsymbol{\alpha} - \boldsymbol{\alpha}_n)^T \boldsymbol{\mathcal{W}} + (\beta - \beta_n))^2 | \mathcal{U} = 1, \mathcal{X}_u = 0] \\ &+ Q_n[I(\mathcal{U} = 0, \mathcal{X}_\ell = 1)] Q_n[((\boldsymbol{\alpha} - \boldsymbol{\alpha}_n)^T \boldsymbol{\mathcal{W}} + (\gamma - \gamma_n))^2 | \mathcal{U} = 0, \mathcal{X}_\ell = 1]. \end{aligned}$$

If $\|\boldsymbol{\eta} - \boldsymbol{\eta}_n\| > \delta/4$, then one of the following cases holds: (i) $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_n\| \ge \min\{\delta/12, \delta/(17C_w)\}$, (ii) $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_n\| < \delta/(17C_w)$ and $\|\beta - \beta_n\| \ge \delta/12$, (iii) $\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_n\| < \delta/(17C_w)$ and $\|\gamma - \gamma_n\| \ge \delta/12$. As $2a^2 + 2(a+b)^2 \ge b^2$ for any reals $a, b$, due to (A.8), we have

$$\begin{aligned} \tilde{D}_2 &\ge C^{-1}\|\boldsymbol{\alpha} - \boldsymbol{\alpha}_n\|^2, \quad \text{if case (i)}, \\ \tilde{D}_2 &\ge C^{-1}|\beta - \beta_n|^2, \quad \text{if case (ii)}, \\ \tilde{D}_2 &\ge C^{-1}|\gamma - \gamma_n|^2, \quad \text{if case (iii)}. \end{aligned}$$

Due to (A.8), there exists a constant $C_2 > 0$ such that

$$D_2 \le -C_2^{-1}\|\boldsymbol{\eta} - \boldsymbol{\eta}_n\|^2 = -C_2^{-1}\delta^2, \quad \text{if } \|\boldsymbol{\eta} - \boldsymbol{\eta}_n\| > \delta/4.$$

Upper bounding $D_3$. The last term $D_3$ in (A.9), as $\phi''(\cdot)$ is upper bounded from (A.8), by the mean value theorem, can be upper bounded by

$$D_3 = Q_n\left[((\boldsymbol{\eta} - \boldsymbol{\eta}_n)^T \boldsymbol{\mathcal{Z}}_c)(\phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}) - \phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_c))\right]$$

$$
\begin{aligned}
=&Q_n\left[((\boldsymbol{\alpha}-\boldsymbol{\alpha}_n)^T\boldsymbol{\mathcal{W}}+(\beta-\beta_n)\mathcal{U})\left(\gamma_n\phi''(\boldsymbol{\alpha}_n^T\boldsymbol{\mathcal{W}}+\beta_n\mathcal{U}+\tilde{\gamma})\right)I(c<\mathcal{X}\le c_n)\right]\\
\le&C|\gamma_n|\|\boldsymbol{\eta}-\boldsymbol{\eta}_n\|Q_n[I(c<\mathcal{X}\le c_n)],
\end{aligned}
$$

where $\tilde{\gamma}$ is between 0 and $\gamma_n$. By condition (A.IV), $\gamma_n\to\gamma_0\neq 0$, and $\|\boldsymbol{\eta}-\boldsymbol{\eta}_n\|\le\delta,\ |c-c_n|\le\delta^2$, then from (A.8), there exists a constant $C_3>0$ such that for large enough $n$,

$$
D_3\le C|\gamma_n||c-c_n|\|\boldsymbol{\eta}-\boldsymbol{\eta}_n\|\le C_3\delta^3.
$$

Since $d_0((c,\boldsymbol{\eta}),(c_n,\boldsymbol{\eta}_n))>\delta/2$, either $\|\boldsymbol{\eta}-\boldsymbol{\eta}_n\|^2>\delta^2/16$ or $|c-c_n|>\delta^2/16\ge m_n^{-1+2\epsilon}$, therefore $D_1+D_2\le -C_4^{-1}\delta^2$. Further, there exists a small enough $\delta_u>0$ such that $-C_4^{-1}\delta^2+C_3\delta^3+C_0\delta^3\le -C^{-1}\delta^2$ for all $\delta\le\delta_u$. Thus, for any $4m_n^{-1/2+\epsilon}\le\delta\le\delta_u$,

$$
\sup_{\delta/2\le d_0((c,\boldsymbol{\eta}),(c_n,\boldsymbol{\eta}_n))\le\delta} Q_n[\varphi_{c,\boldsymbol{\eta}}-\varphi_{c_n,\boldsymbol{\eta}_n}]\le -C^{-1}\delta^2.
$$

**Verifying** (A.7). We next verify (A.7) using [42, Theorem 2.14.1]. Consider the function class

$$
\mathcal{F}_{n,\delta}=\left\{\varphi_{c,\boldsymbol{\eta}}-\varphi_{c_n,\boldsymbol{\eta}_n}: d_0((c,\boldsymbol{\eta}),(c_n,\boldsymbol{\eta}_n)\le\delta\right\}.
$$

For any $(c,\boldsymbol{\eta})$ satisfying $d_0((c,\boldsymbol{\eta}),(c_n,\boldsymbol{\eta}_n))\le\delta$ and $4m_n^{-1/2+\epsilon}\le\delta\le\delta_u$, due to (A.8), by the mean value theorem, $|\varphi_{c,\boldsymbol{\eta}}-\varphi_{c_n,\boldsymbol{\eta}_n}|$ can be upper bounded by

$$
\begin{aligned}
&|y(\boldsymbol{\eta}^T\boldsymbol{z}_c-\boldsymbol{\eta}_n^T\boldsymbol{z}_{c_n})|+|\phi(\boldsymbol{\eta}^T\boldsymbol{z}_c)-\phi(\boldsymbol{\eta}_n^T\boldsymbol{z}_c)|+|\phi(\boldsymbol{\eta}_n^T\boldsymbol{z}_c)-\phi(\boldsymbol{\eta}_n^T\boldsymbol{z}_{c_n})|\\
=&\left|((\boldsymbol{\alpha}-\boldsymbol{\alpha}_n)^T\boldsymbol{w}+(\beta-\beta_n)u+(\gamma-\gamma_n)I(x\le c)-\gamma_nI(c<x\le c_n))y\right|\\
+&\left|\left((\boldsymbol{\alpha}-\boldsymbol{\alpha}_n)^T\boldsymbol{w}+(\beta-\beta_n)u+(\gamma-\gamma_n)I(x\le c)\right)\phi'(a)\right|\\
+&\left|\left(\phi(\boldsymbol{\alpha}_n^T\boldsymbol{w}+\beta_nu+\gamma_n)-\phi(\boldsymbol{\alpha}_n^T\boldsymbol{w}+\beta_nu)\right)I(c<x\le c_n)\right|\\
\le&|y|(\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_n\|C_w+|\beta-\beta_n|+|\gamma-\gamma_n|+|\gamma_n|I(c<x<c_n))\\
+&C(\|\boldsymbol{\alpha}-\boldsymbol{\alpha}_n\|C_w+|\beta-\beta_n|+|\gamma-\gamma_n|+|\gamma_n|I(c<x\le c_n))\\
\le&C(|y|+1)(\delta+|\gamma_n|I(c_n-\delta^2<x<c_n)):=F_{n,\delta},
\end{aligned}
\tag{A.10}
$$

where $a$ is between $\boldsymbol{\eta}^T\boldsymbol{z}_c$ and $\boldsymbol{\eta}_n^T\boldsymbol{z}_c$. From (A.8) and conditions (A.I), (A.IV) and (A.V), we have $\sup_n(Q_n[F_{n,\delta}^2])^{1/2}\le C\delta$.

Similar to Lemma A.1, $\sup_n J(1,\mathcal{F}_{n,\delta},F_{n,\delta})\le C<\infty$ with the envelope function $F_{n,\delta}$. Then, from [42, Theorem 2.14.1] and (A.4), for $4m_n^{-1/2+\epsilon}\le\delta\le\delta_u$, we have

$$
\begin{aligned}
&\mathrm{E}\sup_{\delta/2\le d_0((c,\boldsymbol{\eta}),(c_n,\boldsymbol{\eta}_n))\le\delta}\sqrt{m_n}\left|(\mathbb{Q}_n^*\varphi_{c,\boldsymbol{\eta}}-Q_n\varphi_{c,\boldsymbol{\eta}})-(\mathbb{Q}_n^*\varphi_{c_n,\boldsymbol{\eta}_n}-Q_n\varphi_{c_n,\boldsymbol{\eta}_n})\right|\\
\le&\sup_n J(1,\mathcal{F}_{n,\delta},F_{n,\delta})\sqrt{\sup_n Q_nF_{n,\delta}^2}\le C\delta.
\end{aligned}
$$

□

*A.3.2. Limiting distributions in the identifiable case*

Next we derive the asymptotic distributions of the score test statistic and the MLE under the triangular array setup.

Let $p = E[U]$ and recall $F_0(\cdot)$, $F_1(\cdot)$ in condition (C.I), and $\boldsymbol{V}^{(1)}_{c,\boldsymbol{\eta}}, \boldsymbol{V}^{(2)}_{c,\boldsymbol{\eta}}, V^{(3)}_{c,\boldsymbol{\eta}}$, $\sigma^2_{c,\boldsymbol{\eta}}$ defined in (2.5). Recall from the Section 2.1 that $\mathbb{Z}_{\boldsymbol{\eta}}$ is a random vector of length $(d+2)$ that has the multivariate normal distribution with zero mean and covariance matrix $(\boldsymbol{V}^{(1)}_{c_0,\boldsymbol{\eta}_0})^{-1}$, $\mathbb{Z}_S$ is a random variable that has the normal distribution with zero mean and variance $\sigma^2_{c_0,\boldsymbol{\eta}_0}$, and $\mathbb{Z}_{\boldsymbol{\eta}}$ and $\mathbb{Z}_S$ are independent.

Let $\boldsymbol{g}_{c,\boldsymbol{\eta}}(y, \boldsymbol{w}, x, u) := (y - \phi'(\boldsymbol{\eta}^T \boldsymbol{z}_c))\tilde{\boldsymbol{z}}_c = (\boldsymbol{g}^{(1)}_{c,\boldsymbol{\eta}}, g^{(2)}_{c,\boldsymbol{\eta}})$, where

$$\begin{aligned} \boldsymbol{g}^{(1)}_{c,\boldsymbol{\eta}}(y, \boldsymbol{w}, x, u) &:= (y - \phi'(\boldsymbol{\eta}^T \boldsymbol{z}_c))\boldsymbol{z}_c, \\ g^{(2)}_{c,\boldsymbol{\eta}}(y, \boldsymbol{w}, x, u) &:= (y - \phi'(\boldsymbol{\eta}^T \boldsymbol{z}_c))ux_c. \end{aligned} \tag{A.11}$$

Denote $\mathbb{G}^*_n := \sqrt{m_n}(\mathbb{Q}^*_n - Q_n)$.

**Theorem A.3.** *Consider the identifiable case under the null, i.e.* $\gamma_0 \neq 0$ *and* $\lambda_0 = 0$*. If* (C.I) *and* (A.I)-(A.VII) *hold, then* $(\sqrt{m_n}(\hat{\boldsymbol{\eta}}^*_n - \boldsymbol{\eta}_n), S^*_n) \rightsquigarrow (\mathbb{Z}_{\boldsymbol{\eta}}, \mathbb{Z}_S)$.

*Proof.* By condition (A.VII), we have $\boldsymbol{V}^{(i)}_n \to \boldsymbol{V}^{(i)}_{c_0,\boldsymbol{\eta}_0}$ for $i = 1, 2, 3$, where

$$\begin{aligned} &\boldsymbol{V}^{(1)}_n := Q_n[\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})\boldsymbol{\mathcal{Z}}_{c_n}\boldsymbol{\mathcal{Z}}^T_{c_n}], \quad \boldsymbol{V}^{(2)}_n := Q_n[\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})\mathcal{X}_{c_n}\mathcal{U}\boldsymbol{\mathcal{Z}}^T_{c_n}] \\ &V^{(3)}_n := Q_n[\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})\mathcal{X}_{c_n}\mathcal{U}]. \end{aligned}$$

First we derive an asymptotic linear expansion of $\sqrt{m_n}(\hat{\boldsymbol{\eta}}^*_n - \boldsymbol{\eta}_n)$. Since $(\hat{c}^*_n, \hat{\boldsymbol{\eta}}^*_n)$ is the maximum likelihood estimator, we have

$$0 = \frac{1}{\sqrt{m_n}} \sum_{i=1}^{m_n} \boldsymbol{Z}_{n,i,\hat{c}^*_n}(Y_{n,i} - \phi'(\boldsymbol{Z}^T_{n,i,\hat{c}^*_n}\hat{\boldsymbol{\eta}}^*_n)) = \mathbb{G}^*_n \boldsymbol{g}^{(1)}_{\hat{c}^*_n,\hat{\boldsymbol{\eta}}^*_n} + \sqrt{m_n} Q_n \boldsymbol{g}^{(1)}_{\hat{c}^*_n,\hat{\boldsymbol{\eta}}^*_n}.$$

By Lemma A.6 (ahead), $\mathbb{G}^*_n \boldsymbol{g}^{(1)}_{\hat{c}^*_n,\hat{\boldsymbol{\eta}}^*_n} = \mathbb{G}^*_n \boldsymbol{g}^{(1)}_{c_n,\boldsymbol{\eta}_n} + o_{\mathrm{pr}}(1)$. Further,

$$\begin{aligned} \sqrt{m_n} Q_n \boldsymbol{g}^{(1)}_{\hat{c}^*_n,\hat{\boldsymbol{\eta}}^*_n} &= \sqrt{m_n} Q_n \boldsymbol{g}^{(1)}_{c_n,\hat{\boldsymbol{\eta}}^*_n} + \sqrt{m_n} Q_n [\boldsymbol{g}^{(1)}_{\hat{c}^*_n,\hat{\boldsymbol{\eta}}^*_n} - \boldsymbol{g}^{(1)}_{c_n,\hat{\boldsymbol{\eta}}^*_n}] \\ &=_{(1)} \sqrt{m_n} Q_n \boldsymbol{g}^{(1)}_{c_n,\hat{\boldsymbol{\eta}}^*_n} + o_{\mathrm{pr}}(1), \\ &=_{(2)} -\boldsymbol{V}^{(1)}_n \sqrt{m_n}(\hat{\boldsymbol{\eta}}^*_n - \boldsymbol{\eta}_n) + o_{\mathrm{pr}}(1), \end{aligned}$$

where (1) holds by Lemma A.7 (ahead) and (2) is due to the Taylor Theorem and that $\|\hat{\boldsymbol{\eta}}^*_n - \boldsymbol{\eta}_n\|^2 = o_{\mathrm{pr}}(n^{-1/2})$ by Theorem A.2 (note that $\epsilon < 1/4$ in condition (A.V)). From condition (C.I), $\boldsymbol{V}^{(1)}_{c_0,\boldsymbol{\eta}_0}$ is invertible, and thus we have

$$\sqrt{m_n}(\hat{\boldsymbol{\eta}}^*_n - \boldsymbol{\eta}_n) = (\boldsymbol{V}^{(1)}_n)^{-1} \mathbb{G}^*_n \boldsymbol{g}^{(1)}_{c_n,\boldsymbol{\eta}_n} + o_{\mathrm{pr}}(1). \tag{A.12}$$

Now consider the score test statistic. By similar arguments and by the mean value form of the Taylor Theorem, Lemma A.7 and (A.12), we have

$$S^*_n = \mathbb{G}^*_n g^{(2)}_{\hat{c}^*_n,\hat{\boldsymbol{\eta}}^*_n} + \sqrt{m_n} Q_n g^{(2)}_{\hat{c}^*_n,\hat{\boldsymbol{\eta}}^*_n} = \mathbb{G}^*_n g^{(2)}_{c_n,\boldsymbol{\eta}_n} + \sqrt{m_n} Q_n g^{(2)}_{\hat{c}^*_n,\hat{\boldsymbol{\eta}}^*_n} + o_{\mathrm{pr}}(1)$$

$$
\begin{aligned}
&= \mathbb{G}_n^* g^{(2)}_{c_n,\boldsymbol{\eta}_n} + \sqrt{m_n} Q_n g^{(2)}_{c_n,\hat{\boldsymbol{\eta}}_n^*} + \sqrt{m_n} Q_n [g^{(2)}_{\hat{c}_n^*,\hat{\boldsymbol{\eta}}_n^*} - g^{(2)}_{c_n,\hat{\boldsymbol{\eta}}_n^*}] + o_{\mathrm{pr}}(1) \\
&= \mathbb{G}_n^* g^{(2)}_{c_n,\boldsymbol{\eta}_n} - \boldsymbol{V}_n^{(2)} \sqrt{m_n}(\hat{\boldsymbol{\eta}}_n^* - \boldsymbol{\eta}_n) + o_{\mathrm{pr}}(1) \\
&= \mathbb{G}_n^* g^{(2)}_{c_n,\boldsymbol{\eta}_n} - \boldsymbol{V}_n^{(2)} (\boldsymbol{V}_n^{(1)})^{-1} \mathbb{G}_n^* \boldsymbol{g}^{(1)}_{c_n,\boldsymbol{\eta}_n} + o_{\mathrm{pr}}(1)
\end{aligned}
$$

Thus, we have $(\sqrt{m_n}(\hat{\boldsymbol{\eta}}_n^* - \boldsymbol{\eta}_n), S_n^*) = \mathbb{G}_n^* [\hat{\boldsymbol{g}}_n] + o_{\mathrm{pr}}(1)$, where

$$
\hat{\boldsymbol{g}}_n = \Big( (\boldsymbol{V}_n^{(1)})^{-1} \boldsymbol{g}^{(1)}_{c_n,\boldsymbol{\eta}_n}, \;\; g^{(2)}_{c_n,\boldsymbol{\eta}_n} - \boldsymbol{V}_n^{(2)} (\boldsymbol{V}_n^{(1)})^{-1} \boldsymbol{g}^{(1)}_{c_n,\boldsymbol{\eta}_n} \Big).
$$

Finally, note that

$$
Q_n[\hat{\boldsymbol{g}}_n \hat{\boldsymbol{g}}_n^T] = \begin{bmatrix} (\boldsymbol{V}_n^{(1)})^{-1} & 0 \\ 0 & V_n^{(3)} - \boldsymbol{V}_n^{(2)} (\boldsymbol{V}_n^{(1)})^{-1} (\boldsymbol{V}_n^{(2)})^T \end{bmatrix} \to \begin{bmatrix} (\boldsymbol{V}^{(1)}_{c_0,\boldsymbol{\eta}_0})^{-1} & 0 \\ 0 & \sigma^2_{c_0,\boldsymbol{\eta}_0} \end{bmatrix}.
$$

Then the proof is complete by the Lindeberg-Feller central limit theorem. □

**Lemma A.6.** *Consider the identifiable case under the null, i.e.* $\gamma_0 \neq 0$ *and* $\lambda_0 = 0$. *Under conditions* (A.I)- (A.VII), $\left\| \mathbb{G}_n^* [\boldsymbol{g}_{\hat{c}_n^*,\hat{\boldsymbol{\eta}}_n^*} - \boldsymbol{g}_{c_n,\boldsymbol{\eta}_n}] \right\| = o_{\mathrm{pr}}(1)$.

*Proof.* Fix some $\delta > 0$. Due to Theorem A.2 and the asymptotically uniformly equicontinuity property [42, Page 37], it suffices to show that there exists a tight, uniformly $d_0$-continuous, Gaussian process $\mathbb{G}$ such that $\{\mathbb{G}_n^* \boldsymbol{g}_{c,\boldsymbol{\eta}} : \; (c,\boldsymbol{\eta}) \in K_\delta\}$ converges weakly to $\mathbb{G}$ in $(\ell^\infty(K_\delta))^{d+3}$. In turn, by [42, Theorem 2.11.1], it suffices to show the following conditions hold:

1. There exists an envelope function $G_\delta$ on $S$ such that for any $(c,\boldsymbol{\eta}) \in K_\delta$, $G_\delta(y, \boldsymbol{w}, u, x) \geq |y - \phi'(\boldsymbol{\eta}^T \boldsymbol{z}_c)| \|\boldsymbol{z}_c\|$, and
$$
Q_n[G_\delta^2 \{G_\delta > \epsilon \sqrt{m_n}\}] \to 0, \quad \text{for every } \epsilon > 0;
$$
2. For every positive sequence $\epsilon_n \to 0$,
$$
\sup_{d_0((c_1,\boldsymbol{\eta}_1),(c_2,\boldsymbol{\eta}_2)) \leq \epsilon_n} Q_n \|\boldsymbol{g}_{c_1,\boldsymbol{\eta}_1} - \boldsymbol{g}_{c_2,\boldsymbol{\eta}_2}\|^2 \to 0;
$$
3. For every positive sequence $\epsilon_n \to 0$ and $j \in [d+3]$,
$$
\int_0^{\epsilon_n} \sqrt{\log N(\xi, K_\delta, d_n^{(j)})} \to 0, \qquad \text{in probability},
$$
where $d_n^{(j)}((c_1,\boldsymbol{\eta}_1),(c_2,\boldsymbol{\eta}_2)) = \|\boldsymbol{g}^{(j)}_{c_1,\boldsymbol{\eta}_1} - \boldsymbol{g}^{(j)}_{c_2,\boldsymbol{\eta}_2}\|_{\mathbb{Q}_n^*,2}$, and $\boldsymbol{g}^{(j)}_{c,\boldsymbol{\eta}}$ is the $j$-th element in $\boldsymbol{g}_{c,\boldsymbol{\eta}}$.
4. For any $(c_1,\boldsymbol{\eta}_1),(c_2,\boldsymbol{\eta}_2) \in K_\delta$, $\mathrm{COV}(\mathbb{G}_n^* \boldsymbol{g}_{c_1,\boldsymbol{\eta}_1}, \mathbb{G}_n^* \boldsymbol{g}_{c_2,\boldsymbol{\eta}_2})$ converges.

Verify condition 1. Since $\boldsymbol{\mathcal{W}}, \mathcal{U}, \mathcal{X}_c$ are bounded, $(c_n,\boldsymbol{\eta}_n) \to (c_0,\boldsymbol{\eta}_0)$ by condition (A.IV), and $\phi(\cdot)$ is infinitely differentiable in $\mathbb{R}$, there exists a positive constant $C \geq 1$ such that $G_\delta = C(|\mathcal{Y}| + 1)$ is an envelope function for

$\{g_{c,\boldsymbol{\eta}}: \ (c,\boldsymbol{\eta}) \in K_\delta\}$. Since the distribution of $Y_{n,1}$ belongs to the exponential family (A.1), and $\boldsymbol{\eta}_n \to \boldsymbol{\eta}_0$, for every $\epsilon > 0$,

$$Q_n\left[G_\delta^2\{G_\delta > \epsilon\sqrt{m_n}\}\right] \le \frac{1}{\epsilon\sqrt{m_n}} Q_n\left[G_\delta^3\right] \to 0.$$

Verify condition 2. Note that

$$\|\tilde{\boldsymbol{\mathcal{Z}}}_{c_1} - \tilde{\boldsymbol{\mathcal{Z}}}_{c_2}\|^2 = |\mathcal{X}_{c_1} - \mathcal{X}_{c_2}|^2 + |\mathcal{X}_{c_1} - \mathcal{X}_{c_2}|^2\mathcal{U}^2 \le 2|\mathcal{X}_{c_1} - \mathcal{X}_{c_2}|^2.$$

Further, since $\phi$ is smooth on $\mathbb{R}$ and $\|\boldsymbol{\mathcal{Z}}_c\| \le C_w + 2$, by the mean value theorem, there exists a positive constant $C$ such that for any $((c_1,\boldsymbol{\eta}_1),(c_2,\boldsymbol{\eta}_2)) \in K_\delta^2$,

$$\|\phi'(\boldsymbol{\eta}_1^T\boldsymbol{\mathcal{Z}}_{c_1}) - \phi'(\boldsymbol{\eta}_2^T\boldsymbol{\mathcal{Z}}_{c_2})\| \le C\left(\|\tilde{\mathcal{X}}_{c_1} - \tilde{\mathcal{X}}_{c_2}\| + \|\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2\|\right),$$

and, as a result,

$$\|\boldsymbol{g}_{c_1,\boldsymbol{\eta}_1} - \boldsymbol{g}_{c_2,\boldsymbol{\eta}_2}\| \le C\left(|\mathcal{Y}||\mathcal{X}_{c_1} - \mathcal{X}_{c_2}| + \|\boldsymbol{\eta}_1 - \boldsymbol{\eta}_2\|\right).$$

Then by a similar argument to (A.4), and due to conditions (A.v) and (A.vi), for every $\epsilon_n \to 0$, we have

$$\sup_{d_0((c_1,\boldsymbol{\eta}_1),(c_2,\boldsymbol{\eta}_2)) \le \epsilon_n} Q_n\|\boldsymbol{g}_{c_1,\boldsymbol{\eta}_1} - \boldsymbol{g}_{c_2,\boldsymbol{\eta}_2}\|^2 \to 0.$$

Verify condition 3. We will show condition 3 for $j = d+2$; the other cases can be shown similarly and omitted here. Let $\mathcal{F}_{3,\delta} = \{S \ni (y,\boldsymbol{w},u,x) \to (y - \phi'(\boldsymbol{\eta}^T\boldsymbol{z}_c))x_c : \ (c,\boldsymbol{\eta}) \in K_\delta\}$. Similar to Lemma A.1, there exists a positive constant $C$ such that for each $n$ and $\epsilon > 0$,

$$\sup_Q \log N(\epsilon\|G_\delta\|_{Q,2}, \mathcal{F}_{3,\delta}, L_2(Q)) \le C\left(1 + \log\left(\frac{1}{\epsilon}\right)\right),$$

where $Q$ is any discrete measure on $S$, which indicates $J(\epsilon_n, \mathcal{F}_{3,\delta}, G_\delta) \to 0$ for any $\epsilon_n \to 0$.

Further, since $\mathrm{E}[\|G_\delta\|_{\mathbb{Q}_n^*,2}] = \|G_\delta\|_{Q_n,2}$, due to the definition of $G_\delta$ and (A.4), we have that $1 \le \|G_\delta\|_{\mathbb{Q}_n^*,2} = O_{\mathrm{pr}}(1)$. Then, for any $\epsilon_n \to 0$,

$$\begin{aligned}
&\int_0^{\epsilon_n} N(\xi, K_\delta, d_n^{(j)})d\xi = \int_0^{\epsilon_n} \sqrt{\log N(\xi, \mathcal{F}_{3,\delta}, L_2(\mathbb{Q}_n^*))}d\xi \\
=&\|G_\delta\|_{\mathbb{Q}_n^*,2}\int_0^{\epsilon_n/\|G_\delta\|_{\mathbb{Q}_n^*,2}} \sqrt{\log N(\xi\|G_\delta\|_{\mathbb{Q}_n^*,2}, \mathcal{F}_{3,\delta}, L_2(\mathbb{Q}_n^*))}d\xi \\
\le&\|G_\delta\|_{\mathbb{Q}_n^*,2} J(\epsilon_n/\|G_\delta\|_{\mathbb{Q}_n^*,2}, \mathcal{F}_{3,\delta}, G_\delta) = o_{\mathrm{pr}}(1).
\end{aligned}$$

Verify condition 4. Note that for any $(c_1,\boldsymbol{\eta}_1),(c_2,\boldsymbol{\eta}_2) \in K_\delta$,

$$\mathrm{cov}(\mathbb{G}_n^*\boldsymbol{g}_{c_1,\boldsymbol{\eta}_1}, \mathbb{G}_n^*\boldsymbol{g}_{c_2,\boldsymbol{\eta}_2}) = Q_n[(\mathcal{Y} - \phi'(\boldsymbol{\eta}_1^T\boldsymbol{\mathcal{Z}}_{c_1}))(\mathcal{Y} - \phi'(\boldsymbol{\eta}_2^T\boldsymbol{\mathcal{Z}}_{c_2}))\tilde{\boldsymbol{\mathcal{Z}}}_{c_1}\tilde{\boldsymbol{\mathcal{Z}}}_{c_2}^T].$$

Then condition 4 is verified due to (A.vii). □

**Lemma A.7.** *Consider the identifiable case under the null, i.e.* $\gamma_0 \neq 0$ *and* $\lambda_0 = 0$. *Assume* (A.I)- (A.VI) *hold. Then* $\sqrt{m_n} Q_n[\boldsymbol{g}_{\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*} - \boldsymbol{g}_{c_n, \hat{\boldsymbol{\eta}}_n^*}] = o_{\text{pr}}(1)$.

*Proof.* Due to Theorem A.1 and by the mean value theorem, we have

$$\begin{aligned}
&\sqrt{m_n} Q_n[\boldsymbol{g}_{\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*} - \boldsymbol{g}_{c_n, \hat{\boldsymbol{\eta}}_n^*}] = I(\|\hat{\boldsymbol{\eta}}_n^* - \boldsymbol{\eta}_0\| \leq 1) \sqrt{m_n} Q_n[\boldsymbol{g}_{\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*} - \boldsymbol{g}_{c_n, \hat{\boldsymbol{\eta}}_n^*}] + o_{\text{pr}}(1) \\
&= I(\|\hat{\boldsymbol{\eta}}_n^* - \boldsymbol{\eta}_0\| \leq 1) \sqrt{m_n} Q_n[(\phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}) - \phi'((\hat{\boldsymbol{\eta}}_n^*)^T \boldsymbol{\mathcal{Z}}_{\hat{c}_n^*}))(\tilde{\boldsymbol{\mathcal{Z}}}_{\hat{c}_n^*} - \tilde{\boldsymbol{\mathcal{Z}}}_{c_n})] \\
&\quad - I(\|\hat{\boldsymbol{\eta}}_n^* - \boldsymbol{\eta}_0\| \leq 1) \sqrt{m_n} Q_n[(\phi'((\hat{\boldsymbol{\eta}}_n^*)^T \boldsymbol{\mathcal{Z}}_{\hat{c}_n^*}) - \phi'((\hat{\boldsymbol{\eta}}_n^*)^T \boldsymbol{\mathcal{Z}}_{c_n}) \tilde{\boldsymbol{\mathcal{Z}}}_{c_n}] + o_{\text{pr}}(1) \\
&= I(\|\hat{\boldsymbol{\eta}}_n^* - \boldsymbol{\eta}_0\| \leq 1) \sqrt{m_n} Q_n[(\phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}) - \phi'((\hat{\boldsymbol{\eta}}_n^*)^T \boldsymbol{\mathcal{Z}}_{\hat{c}_n^*}))(\tilde{\boldsymbol{\mathcal{Z}}}_{\hat{c}_n^*} - \tilde{\boldsymbol{\mathcal{Z}}}_{c_n})] \\
&\quad - I(\|\hat{\boldsymbol{\eta}}_n^* - \boldsymbol{\eta}_0\| \leq 1) \sqrt{m_n} Q_n[\phi''(a)((\boldsymbol{\eta}_n^*)^T (\boldsymbol{\mathcal{Z}}_{\hat{c}_n^*} - \boldsymbol{\mathcal{Z}}_{c_n})) \tilde{\boldsymbol{\mathcal{Z}}}_{c_n}] + o_{\text{pr}}(1),
\end{aligned}$$

where $a$ is between $(\hat{\boldsymbol{\eta}}_n^*)^T \boldsymbol{\mathcal{Z}}_{\hat{c}_n^*}$ and $(\hat{\boldsymbol{\eta}}_n^*)^T \boldsymbol{\mathcal{Z}}_{c_n^*}$. Now each function inside $[\cdot]$ is zero unless $\mathcal{X}_{c_n} \neq \mathcal{X}_{c_n^*}$. Since $\phi$ is smooth, and $\sup_{c \in [\ell, u]} \|\tilde{\boldsymbol{\mathcal{Z}}}_c\|$ and $\sup_n \|\boldsymbol{\eta}_n\|$ are both fine, there exists a positive constant $C$ such that

$$\left| \sqrt{m_n} Q_n[\boldsymbol{g}_{\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*} - \boldsymbol{g}_{c_n, \hat{\boldsymbol{\eta}}_n^*}] \right| \leq C \sqrt{m_n} Q_n[I(\hat{c}_n^* \wedge c_n < \mathcal{X} \leq \hat{c}_n^* \vee c_n)] + o_{\text{pr}}(1).$$

Let $\delta \in (\epsilon, 1/4)$, where $\epsilon < 1/4$ appears in condition (A.V). Then

$$\begin{aligned}
&\sqrt{m_n} Q_n[I(\hat{c}_n^* \wedge c_n < \mathcal{X} \leq \hat{c}_n^* \vee c_n)] \\
=& I(|\hat{c}_n^* - c_n| \leq m_n^{-1+2\delta}) \sqrt{m_n} Q_n[I(\hat{c}_n^* \wedge c_n < \mathcal{X} \leq \hat{c}_n^* \vee c_n)] \\
+& I(|\hat{c}_n^* - c_n| > m_n^{-1+2\delta}) \sqrt{m_n} Q_n[I(\hat{c}_n^* \wedge c_n < \mathcal{X} \leq \hat{c}_n^* \vee c_n)] \\
=& o_{\text{pr}}(1),
\end{aligned}$$

where, in the second to the last equality, the first term is $o(1)$ by condition (A.VI), and the second term is $o_{\text{pr}}(1)$ since $m_n^{1-2\epsilon} |\hat{c}_n^* - c_n| = O_{\text{pr}}(1)$ due to Theorem A.2. Then the proof is complete. □

### *A.4. The non-identifiable case*

For the lemmas and theorems in this subsection we assume conditions (A.III) and (A.IV) hold with $\iota = 1$.

#### *A.4.1. Convergence rates in the non-identifiable case*

**Theorem A.4.** *Consider the non-identifiable case under the null, i.e.* $\gamma_0 = \lambda_0 = 0$. *Let* (A.I)-(A.IV) *and* (A.VIII) *hold. Then* $\sqrt{m_n} d_1((\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*), (c_n, \boldsymbol{\eta}_n)) = O_{\text{pr}}(1)$.

*Proof.* The proof of Theorem A.4 is similar to the proof of Theorem A.2. First from conditions (A.I)-(A.IV) with $\iota = 1$ and Theorem A.1, $d_1((\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*), (c_n, \boldsymbol{\eta}_n))$ converges to 0 in probability. Then by [42, Theorem 3.4.1], to achieve the rate $\sqrt{m_n}$, we need to verify (A.6) and (A.7) with $d_0$ replaced by $d_1$, for any fixed $\delta$ which satisfies $\kappa m_n^{-1/2} \leq \delta \leq \delta_u$, where $\kappa, \delta_u$ will be specified below. We also focus on the case $c < c_n$, and omit the case $c > c_n$.

**Verifying (A.6)** Fix some $(c, \boldsymbol{\eta})$ such that $\delta/2 \leq d_1((c, \boldsymbol{\eta}), (c_n, \boldsymbol{\eta}_n)) \leq \delta$. The same decomposition of $Q_n[\varphi_{c,\boldsymbol{\eta}} - \varphi_{c_n,\boldsymbol{\eta}_n}]$ continues to hold in the non-identifiable case:

$$Q_n[\varphi_{c,\boldsymbol{\eta}} - \varphi_{c_n,\boldsymbol{\eta}_n}] \leq D_1 + D_2 + D_3 + C_0\delta^3,$$

where $D_1$, $D_2$, $D_3$ are defined in (A.9). By definition of $d_1(\cdot, \cdot)$, $\|\boldsymbol{\eta} - \boldsymbol{\eta}_n\| \leq \delta$. Further, from the proof of Theorem A.2 and by the Taylor Theorem, for some constant $C > 0$ that does not depend on $\delta, \kappa, \delta_u$, we have

$$\begin{aligned}
D_1 \leq& - C^{-1}\gamma_n^2 Q_n \left[I(c < \mathcal{X} \leq c_n)\right] \leq 0,\\
D_2 \leq& - C^{-1}\delta^2,\\
D_3 \leq& C|\gamma_n||c - c_n|\|\boldsymbol{\eta} - \boldsymbol{\eta}_n\| \leq C(u-\ell)\delta^2/\kappa \times (\sqrt{m_n}|\gamma_n|).
\end{aligned}$$

Recall the limit $B_1$ in condition (A.VIII), and let $\kappa = \max\{3C^2(u-\ell)|B_1|, 1\}$. Then for large enough $n$, $D_3 \leq (2C)^{-1}\delta^2$, which implies that there exists a small $\delta_u > 0$ such that for $\delta \leq \delta_u$ and large enough $n$,

$$\sup_{\delta/2 \leq d_1((c,\boldsymbol{\eta}),(c_n,\boldsymbol{\eta}_n)) \leq \delta} Q_n\left[\varphi_{c,\boldsymbol{\eta}} - \varphi_{c_n,\boldsymbol{\eta}_n}\right] \leq -(3C)^{-1}\delta^2.$$

**Verifying (A.7)** Define

$$\tilde{\mathcal{F}}_{n,\delta} = \left\{\varphi_{c,\boldsymbol{\eta}} - \varphi_{c_n,\boldsymbol{\eta}_n} : d_1((c, \boldsymbol{\eta}), (c_n, \boldsymbol{\eta}_n) \leq \delta\right\}.$$

From (A.10), there exists a positive constant $C$ such that for any $(c, \boldsymbol{\eta})$ satisfying $d_1((c, \boldsymbol{\eta}), (c_n, \boldsymbol{\eta}_n)) \leq \delta$, we have

$$|\varphi_{c,\boldsymbol{\eta}} - \varphi_{c_n,\boldsymbol{\eta}_n}| \leq \tilde{F}_{n,\delta} := C(|y| + 1)(\delta + |\gamma_n|).$$

As $\gamma_n = o(\delta)$, by (A.4), there exists some constant $C$ such that $(Q_n[\tilde{F}^2_{n,\delta}])^{1/2} \leq C\delta$ for large enough $n$.

Similar to Lemma A.1, $\sup_n J(1, \mathcal{F}_{n,\delta}, \tilde{F}_{n,\delta}) < \infty$ with the envelope function $\tilde{F}_{n,\delta}$. Then by [42, Theorem 2.14.1], for any $\delta_u \geq \delta \geq m_n^{-1/2}$ and large enough $n$, we have

$$\mathrm{E} \sup_{\delta/2 \leq d_1((c,\boldsymbol{\eta}),(c_n,\boldsymbol{\eta}_n)) \leq \delta} \sqrt{m_n}|(\mathbb{Q}_n^* - Q_n)\varphi_{c,\boldsymbol{\eta}} - (\mathbb{Q}_n^* - Q_n)\varphi_{c_n,\boldsymbol{\eta}_n})| \leq C\delta.$$

□

### *A.4.2. Limiting distributions in non-identifiable case*

We recall and introduce a few notations. For any $c_1, c_2 \in \mathbb{R}$, define $\rho(c_1, c_2) = |c_1 - c_2|$. Recall the definitions of $(\mathbb{C}, \mathbb{H}, \mathbb{S})$ in (2.7), and $\boldsymbol{V}^{(1)}_{c,\boldsymbol{\eta}}, \boldsymbol{V}^{(2)}_{c,\boldsymbol{\eta}}$ in (2.5). Let $B_1, B_2$ be the constants appearing in condition (A.VIII).

In Lemma A.10, we show that there exists a zero mean Gaussian process $\{((\boldsymbol{\Delta}_c^{(1)})^T, \Delta_c^{(2)}, \Delta_c^{(3)})^T : c \in [\ell, u]\}$, that is tight in $(\ell^\infty([\ell, u]))^{d+4}$, is uniformly $\rho$-continuous, and has the following covariance function: for $c_1, c_2 \in [\ell, u]$,

$$\begin{aligned}&\mathrm{Cov}\left(((\boldsymbol{\Delta}_{c_1}^{(1)})^T, \Delta_{c_1}^{(2)}, \Delta_{c_1}^{(3)})^T, ((\boldsymbol{\Delta}_{c_2}^{(1)})^T, \Delta_{c_2}^{(2)}, \Delta_{c_2}^{(3)})^T\right) \qquad \text{(A.13)}\\ =&P\begin{bmatrix} \phi''(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_0})\tilde{\boldsymbol{\mathcal{Z}}}_{c_1}\tilde{\boldsymbol{\mathcal{Z}}}_{c_2}^T & \phi''(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_0})\tilde{\boldsymbol{\mathcal{Z}}}_{c_1}(\mathcal{X}_{c_2} - \mathcal{X}_{B_2}) \\ \phi''(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_0})\tilde{\boldsymbol{\mathcal{Z}}}_{c_2}^T(\mathcal{X}_{c_1} - \mathcal{X}_{B_2}) & \phi''(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_0})(\mathcal{X}_{c_1} - \mathcal{X}_{B_2})(\mathcal{X}_{c_2} - \mathcal{X}_{B_2})\end{bmatrix}.\end{aligned}$$

Note that for each $c$, $\boldsymbol{\Delta}_c^{(1)}$ is a random vector of length $d+2$ and $\Delta_c^{(2)}$, $\Delta_c^{(3)}$ are real valued random variables. Further, $\{((\boldsymbol{\Delta}_c^{(1)})^T, \Delta_c^{(2)})^T : \ c \in [\ell, u]\}$ does not depend on $B_2$, and has the same distribution as the random process appear in Subsection 2.1.

Further, for each $c \in [\ell, u]$, define

$$\mu_c^{(1)} = P[\phi''(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_0})\mathcal{U}I(B_2 < \mathcal{X} \le c)], \quad \mu_c^{(2)} = P[\phi''(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_0})(\mathcal{X}_c - \mathcal{X}_{B_2})^2],$$
$$\boldsymbol{\mu}_c^{(3)} = P[\phi''(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_0})(\mathcal{X}_c - \mathcal{X}_{B_2})\boldsymbol{\mathcal{Z}}_c],$$

and

$$\begin{aligned}\widetilde{\mathbb{C}} &= \underset{c\in[\ell,u]}{\operatorname{argmax}}\ \frac{1}{2}\left(\boldsymbol{\Delta}_c^{(1)} - B_1\boldsymbol{\mu}_c^{(3)}\right)^T (\boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(1)})^{-1}\left(\boldsymbol{\Delta}_c^{(1)} - B_1\boldsymbol{\mu}_c^{(3)}\right)\\ &\qquad + B_1\Delta_c^{(3)} - \frac{1}{2}B_1^2\mu_c^{(2)},\\ \widetilde{\mathbb{H}} &= (\boldsymbol{V}_{\widetilde{\mathbb{C}},\boldsymbol{\eta}_0}^{(1)})^{-1}\left(\boldsymbol{\Delta}_{\widetilde{\mathbb{C}}}^{(1)} - B_1\boldsymbol{\mu}_{\widetilde{\mathbb{C}}}^{(3)}\right),\\ \widetilde{\mathbb{S}} &= \Delta_{\widetilde{\mathbb{C}}}^{(2)} - \boldsymbol{V}_{\widetilde{\mathbb{C}},\boldsymbol{\eta}_0}^{(2)}\widetilde{\mathbb{H}} - B_1\mu_{\widetilde{\mathbb{C}}}^{(1)}.\end{aligned} \tag{A.14}$$

Note that when $B_1 = 0$, the distribution of $(\widetilde{\mathbb{C}}, \widetilde{\mathbb{H}}, \widetilde{\mathbb{S}})$ does not depend on $B_2$, and is the same as $(\mathbb{C}, \mathbb{H}, \mathbb{S})$ in (2.7).

Further, for any $\delta > 0$ let $\tilde{K}_\delta = \{(c, \boldsymbol{h}) \in [\ell, u] \times \mathbb{R}^{d+2} : c \in [\ell, u], \|\boldsymbol{h}\| \le \delta\}$. For any $(c, \boldsymbol{h}) \in [\ell, u] \times \mathbb{R}^{d+2}$, define the following functions on the observation space $S$:

$$\begin{aligned}&\tilde{\varphi}_{c,\boldsymbol{h},n} := \sqrt{m_n}(\varphi_{c,\boldsymbol{\eta}_n+\boldsymbol{h}/\sqrt{m_n}} - \varphi_{c_n,\boldsymbol{\eta}_n}),\\ &\boldsymbol{f}_{c,n}^{(1)} = (\mathcal{Y} - \phi'(\boldsymbol{\eta}_n^T\boldsymbol{\mathcal{Z}}_{c_n}))\boldsymbol{\mathcal{Z}}_c, \quad f_{c,n}^{(2)} = (\mathcal{Y} - \phi'(\boldsymbol{\eta}_n^T\boldsymbol{\mathcal{Z}}_{c_n}))\mathcal{U}\mathcal{X}_c,\\ &\boldsymbol{f}_{c,n}^{(3)} = \phi''(\boldsymbol{\eta}_n^T\boldsymbol{\mathcal{Z}}_c)\mathcal{U}\mathcal{X}_c\boldsymbol{\mathcal{Z}}_c, \quad f_{c,n}^{(4)} = \phi''(\boldsymbol{\eta}_n^T\boldsymbol{\mathcal{Z}}_{c_n})\mathcal{U}I(c_n < \mathcal{X} \le c),\\ &f_{c,n}^{(5)} = \left(\mathcal{Y} - \phi'(\boldsymbol{\eta}_n^T\boldsymbol{\mathcal{Z}}_{c_n})\right)(\mathcal{X}_c - \mathcal{X}_{c_n}).\end{aligned} \tag{A.15}$$

Finally, denote $\mathbb{G}_n^* = \sqrt{m_n}(\mathbb{Q}_n^* - Q_n)$.

The following theorem establishes the limiting distribution of the MLE and the score test statistics under the triangular array setup.

**Theorem A.5.** *Consider the non-identifiable case under the null, i.e.* $\gamma_0 = \lambda_0 = 0$. *If* (C.I), (A.I)-(A.IV) *with* $\tau = 1$, (A.VIII) *and* (A.IX) *hold, then*

$$(\hat{c}_n^*, \sqrt{m_n}(\hat{\boldsymbol{\eta}}_n^* - \boldsymbol{\eta}_n), S_n^*) \rightsquigarrow (\widetilde{\mathbb{C}}, \widetilde{\mathbb{H}}, \widetilde{\mathbb{S}}).$$

*Proof.* Let $\hat{\boldsymbol{h}}_n^* = \sqrt{m_n}(\hat{\boldsymbol{\eta}}_n^* - \boldsymbol{\eta}_n)$. By Theorem A.4, $\hat{\boldsymbol{h}}_n^*$ is bounded in probability. Note that by definition,

$$(\hat{c}_n^*, \hat{\boldsymbol{h}}_n^*) = \underset{(c,\boldsymbol{h})\in[\ell,u]\times\mathbb{R}^{d+2}}{\text{sargmax}} \mathbb{Q}_n^*[\varphi_{c,\boldsymbol{\eta}_n+\boldsymbol{h}/\sqrt{m_n}} - \varphi_{c_n,\boldsymbol{\eta}_n}]$$
$$= \underset{(c,\boldsymbol{h})\in[\ell,u]\times\mathbb{R}^{d+2}}{\text{sargmax}} \mathbb{G}_n^*\tilde{\varphi}_{c,\boldsymbol{h},n} + \sqrt{m_n}Q_n\tilde{\varphi}_{c,\boldsymbol{h},n}.$$

Further, observe that

$$\begin{aligned} S_n^* &= \sqrt{m_n}\mathbb{G}_n^*[(\mathcal{Y} - \phi'(\boldsymbol{\eta}_n^T\boldsymbol{\mathcal{Z}}_{c_n}))\mathcal{U}\mathcal{X}_{\hat{c}_n^*}] \\ &\quad - \sqrt{m_n}\mathbb{Q}_n^*[(\phi'((\hat{\boldsymbol{\eta}}_n^*)^T\boldsymbol{\mathcal{Z}}_{\hat{c}_n^*}) - \phi'(\boldsymbol{\eta}_n^T\boldsymbol{\mathcal{Z}}_{\hat{c}_n^*}))\mathcal{U}\mathcal{X}_{\hat{c}_n^*}] \\ &\quad + \sqrt{m_n}\mathbb{Q}_n^*[(\phi'(\boldsymbol{\eta}_n^T\boldsymbol{\mathcal{Z}}_{c_n}) - \phi'(\boldsymbol{\eta}_n^T\boldsymbol{\mathcal{Z}}_{\hat{c}_n^*}))\mathcal{U}I(c_n < \mathcal{X} \le \hat{c}_n^*)] \\ &= \sqrt{m_n}\mathbb{G}_n^*[f_{\hat{c}_n^*,n}^{(2)}] - \mathbb{Q}_n^*[(\boldsymbol{f}_{\hat{c}_n^*,n}^{(3)})^T]\hat{\boldsymbol{h}}_n^* - (\sqrt{m_n}\gamma_n)\mathbb{Q}_n^*[f_{\hat{c}_n^*,n}^{(4)}] + o_{\text{pr}}(1), \end{aligned}$$

where in the last equality, we applied the Taylor Theorem, and used the fact that $\hat{\boldsymbol{h}}_n^* = O_{\text{pr}}(1)$ and $\sqrt{m_n}\gamma_n \to B_1$ by (A.VIII).

By Lemmas A.8, A.10 (both ahead) and the Slutsky's theorem, for any $\delta > 0$, in the space of $(\ell^\infty(\tilde{K}_\delta))^{d+6}$, we have for $(c, \boldsymbol{h}) \in \tilde{K}_\delta$,

$$\begin{bmatrix} \sqrt{m_n}Q_n\tilde{\varphi}_{c,\boldsymbol{h},n} \\ \mathbb{Q}_n^*[(\boldsymbol{f}_{c,n}^{(3)})^T] \\ \mathbb{Q}_n^*[f_{c,n}^{(4)}] \\ \mathbb{G}_n^*\tilde{\varphi}_{c,\boldsymbol{h},n} \\ \mathbb{G}_n^* f_{c,n}^{(2)} \end{bmatrix} \rightsquigarrow \begin{bmatrix} -\frac{1}{2}\boldsymbol{h}^T\boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(1)}\boldsymbol{h} - \frac{B_1^2}{2}\mu_c^{(2)} - B_1\boldsymbol{h}^T\boldsymbol{\mu}_c^{(3)} \\ \boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(2)} \\ \mu_c^{(1)} \\ \boldsymbol{h}^T\boldsymbol{\Delta}_c^{(1)} + B_1\Delta_c^{(3)} \\ \Delta_c^{(2)} \end{bmatrix}.$$

For each $c \in [\ell, u]$, the maximizer and the maximum value of the function $\mathbb{R}^{d+2} \ni \boldsymbol{h} \mapsto \boldsymbol{h}^T\boldsymbol{\Delta}_c^{(1)} + B_1\Delta_c^{(3)} - \frac{1}{2}\boldsymbol{h}^T\boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(1)}\boldsymbol{h} - \frac{B_1^2}{2}\mu_c^{(2)} - B_1\boldsymbol{h}^T\boldsymbol{\mu}_c^{(3)} \in \mathbb{R}$ are respectively:

$$\begin{aligned} &(\boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(1)})^{-1}(\boldsymbol{\Delta}_c^{(1)} - B_1\boldsymbol{\mu}_c^{(3)}), \qquad \text{and} \\ &\frac{1}{2}\left(\boldsymbol{\Delta}_c^{(1)} - B_1\boldsymbol{\mu}_c^{(3)}\right)^T(\boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(1)})^{-1}\left(\boldsymbol{\Delta}_c^{(1)} - B_1\boldsymbol{\mu}_c^{(3)}\right) + B_1\Delta_c^{(3)} - \frac{B_1^2}{2}\mu_c^{(2)}. \end{aligned}$$

Further, note that $\{((\boldsymbol{\Delta}_c^{(1)})^T, \Delta_c^{(2)}, \Delta_c^{(3)})^T : c \in [\ell, u]\}$ is uniformly $\rho$-continuous, and so are $\boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(1)}, \boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(2)}, \mu_c^{(1)}, \mu_c^{(2)}, \boldsymbol{\mu}_c^{(3)}$ due to (C.I), and that $\hat{\boldsymbol{h}}_n^* = O_{\text{pr}}(1)$. Then by the continuous mapping theorem, we have

$$\left(\hat{c}_n^*, \hat{\boldsymbol{h}}_n^*, \mathbb{G}_n^* f_{\hat{c}_n^*,n}^{(2)}, \mathbb{Q}_n^*[\boldsymbol{f}_{\hat{c}_n^*,n}^{(3)}]^T, \mathbb{Q}_n^*[f_{\hat{c}_n^*,n}^{(4)}]\right) \rightsquigarrow (\widetilde{\mathbb{C}}, \widetilde{\mathbb{H}}, \Delta_{\widetilde{\mathbb{C}}}^{(2)}, \boldsymbol{V}_{\widetilde{\mathbb{C}},\boldsymbol{\eta}_0}^{(2)}, \mu_{\widetilde{\mathbb{C}}}^{(1)}).$$

Finally, the proof is complete by another application of the continuous mapping theorem. □

Recall the definitions of $\tilde{\varphi}_{c,\boldsymbol{h},n}$, $\boldsymbol{f}_{c,n}^{(3)}$ and $f_{c,n}^{(4)}$ above.

**Lemma A.8.** *Consider the non-identifiable case under the null, i.e.* $\gamma_0 = \lambda_0 = 0$*. Assume* (C.I)*,* (A.I)*,* (A.IV)*,* (A.VIII) *and* (A.IX) *hold. Then for any* $\delta > 0$

$$\sup_{(c,\boldsymbol{h})\in\tilde{K}_\delta} |\sqrt{m_n} Q_n \tilde{\varphi}_{c,\boldsymbol{h},n} + \frac{1}{2}\boldsymbol{h}^T \boldsymbol{V}^{(1)}_{c,\boldsymbol{\eta}_0}\boldsymbol{h} + \frac{B_1^2}{2}\mu_c^{(2)} + B_1 \boldsymbol{h}^T \boldsymbol{\mu}_c^{(3)}| \to 0.$$

$$\sup_{c\in[\ell,u]} \|\mathbb{Q}_n^*[\boldsymbol{f}^{(3)}_{c,n}]^T - \boldsymbol{V}^{(2)}_{c,\boldsymbol{\eta}_0}\| \to 0, \quad \textit{in probability.}$$

$$\sup_{c\in[\ell,u]} |\mathbb{Q}_n^*[f^{(4)}_{c,n}] - \mu_c^{(1)}| \to 0, \quad \textit{in probability.}$$

*Proof.* We start with the first claim. Note that for any $(c, \boldsymbol{h}) \in \tilde{K}_\delta$,

$$\begin{aligned}
\sqrt{m_n} Q_n \tilde{\varphi}_{c,\boldsymbol{h},n} &= m_n Q_n[\varphi_{c,\boldsymbol{\eta}_n + \boldsymbol{h}/\sqrt{m_n}} - \varphi_{c_n,\boldsymbol{\eta}_n}] \\
=& m_n Q_n \left[\phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})((\boldsymbol{\eta}_n + \boldsymbol{h}/\sqrt{m_n})^T \boldsymbol{\mathcal{Z}}_c - \boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})\right] \\
&- m_n Q_n \left[(\phi((\boldsymbol{\eta}_n + \boldsymbol{h}/\sqrt{m_n})^T \boldsymbol{\mathcal{Z}}_c) - \phi(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}))\right] \\
=& m_n Q_n[g(r_n, s_n, t_n, \boldsymbol{q}_n)],
\end{aligned}$$

where $r_n = \boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}$, $s_n = \gamma_n \mathcal{X}_{c_n}$, $t_n = \gamma_n \mathcal{X}_c$, $\boldsymbol{q}_n = \frac{\boldsymbol{h}}{\sqrt{m_n}}$, and

$$g(r, s, t, \boldsymbol{q}) = \phi'(r)(t - s + \boldsymbol{q}^T \boldsymbol{\mathcal{Z}}_c) - (\phi(r - s + t + \boldsymbol{q}^T \boldsymbol{\mathcal{Z}}_c) - \phi(r)).$$

Let $\boldsymbol{v} = (s, t, \boldsymbol{q})^T$. Elementary calculation shows that

$$\begin{aligned}
&g(r_n, 0, 0, 0) = 0, \quad \left.\frac{\partial g(r_n, \boldsymbol{v}^T)}{\partial \boldsymbol{v}}\right|_{(r_n,0,0,0)} = (0, 0, 0)^T, \\
&\left.\frac{\partial^2 g(r_n, \boldsymbol{v}^T)}{\partial \boldsymbol{v}\partial \boldsymbol{v}^T}\right|_{(r_n,0,0,0)} = -\phi''(r_n) \begin{bmatrix} 1 \\ -1 \\ -\boldsymbol{\mathcal{Z}}_c \end{bmatrix} \times \begin{bmatrix} 1 & -1 & -\boldsymbol{\mathcal{Z}}_c \end{bmatrix}.
\end{aligned}$$

From conditions (A.IV) and (A.VIII), $\boldsymbol{\eta}_n \to \boldsymbol{\eta}_0$ and $\sqrt{m_n}\gamma_n \to B_1$. Since $\|\boldsymbol{\mathcal{Z}}_c\|$ is bounded,

$$\sup_{(c,\boldsymbol{h})\in\tilde{K}_\delta,\, n\in\mathbb{N}} \left(|\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})| + |\phi'''((\boldsymbol{\eta}_n + \boldsymbol{h}/\sqrt{m_n})^T \boldsymbol{\mathcal{Z}}_c)|\right) < \infty,$$

by the mean value theorem,

$$\begin{aligned}
&\sqrt{m_n} Q_n \tilde{\varphi}_{c,\boldsymbol{h},n} = -2^{-1} Q_n[\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})(\sqrt{m_n}\gamma_n(\mathcal{X}_c - \mathcal{X}_{c_n}) + \boldsymbol{h}^T \boldsymbol{\mathcal{Z}}_c)^2] + o(1) \\
&= -2^{-1}\boldsymbol{h}^T Q_n \left[\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}) \boldsymbol{\mathcal{Z}}_c \boldsymbol{\mathcal{Z}}_c^T\right] \boldsymbol{h} \\
&\quad - 2^{-1}(\sqrt{m_n}\gamma_n)^2 Q_n[\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})(\mathcal{X}_c - \mathcal{X}_{c_n})^2] \\
&\quad - (\sqrt{m_n}\gamma_n)\boldsymbol{h}^T Q_n[\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})(\mathcal{X}_c - \mathcal{X}_{c_n})\boldsymbol{\mathcal{Z}}_c] + o(1),
\end{aligned}$$

where $o(1)$ is uniform over $(c, \boldsymbol{h}) \in \tilde{K}_\delta$. Then the proof of the first claim is complete due to condition (A.IX) and (A.VIII).

Now we consider the second claim. Note that $\sup_{c\in[\ell,u]}\|\mathbb{Q}_n^*[\boldsymbol{f}_{c,n}^{(3)}]^T - V_{c,\boldsymbol{\eta}_0}^{(2)}\|$ is upper bounded by

$$\sup_{c\in[\ell,u]}\|(\mathbb{Q}_n^* - Q_n)[\boldsymbol{f}_{c,n}^{(3)}]^T\| + \sup_{c\in[\ell,u]}\|Q_n[\phi''(\boldsymbol{\eta}_n^T\boldsymbol{\mathcal{Z}}_c)\mathcal{U}\mathcal{X}_c\boldsymbol{\mathcal{Z}}_c^T] - P[\phi''(\boldsymbol{\alpha}_0^T\boldsymbol{\mathcal{W}} + \beta_0\mathcal{U})\mathcal{U}\mathcal{X}_c\boldsymbol{\mathcal{Z}}_c^T]\|.$$

By a similar argument to Lemma A.3, the first term is $o_{\mathrm{pr}}(1)$. Further, the second term is $o(1)$ by condition (A.IX).

The proof for the last claim is similar, and thus omitted. Then the proof is complete. □

Recall that $\tilde{\varphi}_{c,\boldsymbol{h},n}$ and $\boldsymbol{f}_{c,n}^{(1)}, f_{c,n}^{(2)}, f_{c,n}^{(5)}$ are defined (A.15), and that $\mathbb{G}_n^* = \sqrt{m_n}(\mathbb{Q}_n^* - Q_n)$. Next we derive the limiting process for $\{\mathbb{G}_n^*[\tilde{\varphi}_{c,\boldsymbol{h},n}, f_{c,n}^{(2)}] : (c,h) \in \tilde{K}_\delta\}$ in $(\ell^\infty(\tilde{K}_\delta))^2$ for any $\delta > 0$. The key step is to approximate $\mathbb{G}_n^*[\tilde{\varphi}_{c,\boldsymbol{h},n}]$ by $\mathbb{G}_n^*[\boldsymbol{h}^T\boldsymbol{f}_{c,n}^{(1)} + B_1 f_{c,n}^{(5)}]$ uniformly over $(c,\boldsymbol{h}) \in \tilde{K}_\delta$, which is established in the following lemma.

**Lemma A.9.** *Consider the non-identifiable case under the null, i.e.* $\gamma_0 = \lambda_0 = 0$. *Suppose* (A.I), (A.IV) *and* (A.VIII) *hold, then for any* $\delta > 0$*,*

$$\sup_{(c,\boldsymbol{h})\in\tilde{K}_\delta}\left|\mathbb{G}_n^*[\tilde{\varphi}_{c,\boldsymbol{h},n}] - \mathbb{G}_n^*[\boldsymbol{h}^T\boldsymbol{f}_{c,n}^{(1)} + B_1 f_{c,n}^{(5)}]\right| = o_{\mathrm{pr}}(1).$$

*Proof.* Fix some $\delta > 0$. Due to (A.VIII) and (A.I), $\gamma_n = O(1/\sqrt{m_n})$ and $\sup_{c\in[\ell,u]}\|\boldsymbol{\mathcal{Z}}_c\| < \infty$, thus uniformly over $(c,\boldsymbol{h}) \in \tilde{K}_\delta$,

$$\left(\boldsymbol{\eta}_n + \frac{\boldsymbol{h}}{\sqrt{m_n}}\right)^T\boldsymbol{\mathcal{Z}}_c - \boldsymbol{\eta}_n^T\boldsymbol{\mathcal{Z}}_{c_n} = \gamma_n\left(\mathcal{X}_c - \mathcal{X}_{c_n}\right) + \frac{1}{\sqrt{m_n}}\boldsymbol{h}^T\boldsymbol{\mathcal{Z}}_c = O\left(\frac{1}{\sqrt{m_n}}\right).$$

Since $\phi(\cdot)$ is infinitely differentiable on $\mathbb{R}$, $\sup_n\|\boldsymbol{\eta}_n\| < \infty$ due to (A.IV), and $\sup_{c\in[\ell,u]}\|\boldsymbol{\mathcal{Z}}_c\| < \infty$, by the Taylor expansion to the third order, uniformly over $(c,\boldsymbol{h}) \in \tilde{K}_\delta$,

$$\begin{aligned}&\tilde{\varphi}_{c,\boldsymbol{h},n} - \boldsymbol{h}^T\boldsymbol{f}_{c,n}^{(1)} - B_1 f_{c,n}^{(5)} = (\sqrt{m_n}\gamma_n - B_1)f_{c,n}^{(5)}\\ &\quad - \frac{1}{2\sqrt{m_n}}\phi''(\boldsymbol{\eta}_n^T\boldsymbol{\mathcal{Z}}_{c_n})(\sqrt{m_n}\gamma_n\left(\mathcal{X}_c - \mathcal{X}_{c_n}\right) + \boldsymbol{h}^T\boldsymbol{\mathcal{Z}}_c)^2 + o\left(\frac{1}{\sqrt{m_n}}\right).\end{aligned}$$

Thus, $\sup_{(c,\boldsymbol{h})\in\tilde{K}_\delta}\left|\mathbb{G}_n^*[\tilde{\varphi}_{c,\boldsymbol{h},n}] - \mathbb{G}_n^*[\boldsymbol{h}^T\boldsymbol{f}_{c,n}^{(1)} + B_1 f_{c,n}^{(5)}]\right| \le |\sqrt{m_n}\gamma_n - B_1| \times I_n + 2^{-1}II_n + o(1)$, where

$$\begin{aligned}I_n &:= \sup_{(c,\boldsymbol{h})\in\tilde{K}_\delta}\left|\mathbb{G}_n^*[\left(\mathcal{Y} - \phi'(\boldsymbol{\eta}_n^T\boldsymbol{\mathcal{Z}}_{c_n})\right)\left(\mathcal{X}_c - \mathcal{X}_{c_n}\right)]\right|,\\ II_n &:= \sup_{(c,\boldsymbol{h})\in\tilde{K}_\delta}\left|(\mathbb{Q}_n^* - Q_n)[\phi''(\boldsymbol{\eta}_n^T\boldsymbol{\mathcal{Z}}_{c_n})(\sqrt{m_n}\gamma_n\left(\mathcal{X}_c - \mathcal{X}_{c_n}\right) + \boldsymbol{h}^T\boldsymbol{\mathcal{Z}}_c)^2]\right|.\end{aligned}$$

By [42, Theorem 2.14.1] and a similar argument as Lemma A.3, we have $I_n = O_{\mathrm{pr}}(1)$ and $II_n = o_{\mathrm{pr}}(1)$, which complete the proof due to (A.VIII). □

Recall the distance function $\rho(c_1, c_2) = |c_1 - c_2|$ for $c_1, c_2 \in [\ell, u]$.

**Lemma A.10.** *Consider the non-identifiable case under the null, i.e.* $\gamma_0 = \lambda_0 = 0$*. Suppose* (A.I), (A.IV), (A.VIII) *and* (A.IX) *hold. Then there exists a zero mean Gaussian process* $\{((\boldsymbol{\Delta}_c^{(1)})^T, \Delta_c^{(2)}, \Delta_c^{(3)})^T : \ c \in [\ell, u]\}$*, that is tight in* $(\ell^\infty([\ell, u]))^{d+4}$*, that is uniformly* $\rho$*-continuous, whose covariance function is given by* (A.13)*, and for which* $\boldsymbol{\Delta}_c^{(1)}$ *is of length* $d+2$ *and* $\Delta_c^{(2)}, \Delta_c^{(3)}$ *both of length* 1 *for each* $c \in [\ell, u]$*. Further, for any* $\delta > 0$*, in* $(\ell^\infty(\tilde{K}_\delta))^2$*,*

$$\{\mathbb{G}_n^*(\tilde{\varphi}_{c,\boldsymbol{h},n}, f_{c,n}^{(2)}) : \ (c, \boldsymbol{h}) \in \tilde{K}_\delta\} \rightsquigarrow \left\{(\boldsymbol{h}^T \boldsymbol{\Delta}_c^{(1)} + B_1 \Delta_c^{(3)}, \Delta_c^{(2)}) : \ (c, \boldsymbol{h}) \in \tilde{K}_\delta\right\}.$$

*Proof.* Fix some $\delta > 0$. By Lemma A.9 and the continuous mapping theorem, it suffices to show that there exists a Gaussian process $\{((\boldsymbol{\Delta}_c^{(1)})^T, \Delta_c^{(2)}, \Delta_c^{(3)})^T :$ $c \in [\ell, u]\}$ with the above prescribed conditions such that in $(\ell^\infty([\ell, u]))^{d+4}$

$$\{\mathbb{G}_n^*((\boldsymbol{f}_{c,n}^{(1)})^T, f_{c,n}^{(2)}, f_{c,n}^{(5)})^T) : \ c \in [\ell, u]\} \rightsquigarrow \{((\boldsymbol{\Delta}_c^{(1)})^T, \Delta_c^{(2)}, \Delta_c^{(3)})^T : \ c \in [\ell, u]\}.$$

In turn, by [42, Theorem 2.11.1], it suffices to verify the following conditions:

1. There exists a function $G_\delta$ on $S$ such that uniformly over $c \in [\ell, u]$, $G_\delta(y, \boldsymbol{w}, u, x) \geq |y - \phi'(\boldsymbol{\eta}_n^T \boldsymbol{z}_{c_n})| \|\tilde{\boldsymbol{z}}_c\|$, and
$$Q_n[G_{n,\delta}^2 \{G_{n,\delta} > \epsilon \sqrt{m_n}\}] \to 0, \quad \text{for every } \epsilon > 0.$$
2. For every positive sequence $\epsilon_n \to 0$,
$$\sup_{|c_1 - c_2| < \epsilon_n} Q_n[(\mathcal{Y} - \phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}))^2 \|\tilde{\boldsymbol{\mathcal{Z}}}_{c_1} - \tilde{\boldsymbol{\mathcal{Z}}}_{c_2}\|^2] \to 0.$$
3. For every positive sequence $\epsilon_n \to 0$ and $j \in [d+4]$,
$$\int_0^{\epsilon_n} \sqrt{\log N(\xi, [\ell, u], d_n^{(j)})} \to 0, \quad \text{in probability,}$$
where $d_n^{(j)}(c_1, c_2) = \|(\mathcal{Y} - \phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})(\tilde{\boldsymbol{\mathcal{Z}}}_{c_1}^{(j)} - \tilde{\boldsymbol{\mathcal{Z}}}_{c_2}^{(j)}))\|_{\mathbb{Q}_n^*, 2}$, and $\tilde{\boldsymbol{\mathcal{Z}}}_c^{(j)}$ is the $j$-th element in $\tilde{\boldsymbol{\mathcal{Z}}}_c$.
4. For any $(c_1, c_2) \in [\ell, u]^2$,
$$\text{cov}\left(\mathbb{G}_n^*[(\boldsymbol{f}_{c_1,n}^{(1)})^T, f_{c_1,n}^{(2)}, f_{c_1,n}^{(5)}]^T, \mathbb{G}_n^*[(\boldsymbol{f}_{c_2,n}^{(1)})^T, f_{c_2,n}^{(2)}, f_{c_2,n}^{(5)}]^T\right)$$
converges to the right hand side of (A.13).

The verification for the first three conditions is almost identical to those arguments in the proof of Lemma A.6, and thus omitted. Further, the condition 4 is due to assumption (A.IX), which completes the proof. ∎

## Appendix B: Some useful lemmas and proofs

**Lemma B.1.** *Consider the case under* $H_0 : \lambda_0 = 0$*. If* (C.I) *holds, then condition* (A.III) *holds with* $\iota = 0$ *for the identifiable case, i.e.,* $\gamma_0 \neq 0$*, and with* $\iota = 1$ *for the non-identifiable case, i.e.,* $\gamma_0 = 0$*.*

*Proof.* Denote $M(c, \boldsymbol{\eta}) = P\varphi_{c,\boldsymbol{\eta}}$. We consider two cases separately.

**Identifiable case.** First we show that $(c_0, \boldsymbol{\eta}_0)$ is the unique maximizer of $M(c, \boldsymbol{\eta})$. By [41, Lemma 5.35], it is sufficient to show that if $d_0((c, \boldsymbol{\eta}), (c_0, \boldsymbol{\eta}_0)) \neq 0$, then $\mathrm{pr}(\phi'(\boldsymbol{\eta}^T \boldsymbol{Z}_c) \neq \phi'(\boldsymbol{\eta}_0^T \boldsymbol{Z}_{c_0})) > 0$, which is equivalent to $\mathrm{pr}(\boldsymbol{\eta}^T \boldsymbol{Z}_c \neq \boldsymbol{\eta}_0^T \boldsymbol{Z}_{c_0}) > 0$, since $\phi'(\cdot)$ is strictly increasing. Note that

$$\{\boldsymbol{\eta}^T \boldsymbol{Z}_c \neq \boldsymbol{\eta}_0^T \boldsymbol{Z}_{c_0}\} = \{(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T \boldsymbol{W} + (\beta - \beta_0)U + (\gamma - \gamma_0)X_c + \gamma_0(X_c - X_{c_0}) \neq 0\}.$$

From the definition, $d_0((c, \boldsymbol{\eta}), (c_0, \boldsymbol{\eta}_0)) \neq 0$ indicates one of the following cases holds: (i) $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}_0$, (ii) $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$, $\beta \neq \beta_0$, (iii) $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$, $\beta = \beta_0$, $\gamma \neq \gamma_0$, and (iv) $\boldsymbol{\alpha} = \boldsymbol{\alpha}_0$, $\beta = \beta_0$, $\gamma = \gamma_0$, $c \neq c_0$.

If $(c, \boldsymbol{\eta})$ belongs to case (i), we have

$$\mathrm{pr}(\boldsymbol{\eta}^T \boldsymbol{Z}_c \neq \boldsymbol{\eta}_0 \boldsymbol{Z}_{c_0}) \geq \mathrm{pr}((\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T \boldsymbol{W} \neq 0, U = X_u = 0).$$

From Assumption (C.I), $\lambda_{\min}(\mathrm{E}(\boldsymbol{W}\boldsymbol{W}^T | U = 0, X_u = 0)) > 0$ and thus $(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T \mathrm{E}(\boldsymbol{W}\boldsymbol{W}^T | U = 0, X_u = 0)(\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) \geq \lambda_{\min}(\mathrm{E}(\boldsymbol{W}\boldsymbol{W}^T | U = 0, X_u = 0)) \|\boldsymbol{\alpha} - \boldsymbol{\alpha}_0\|^2 > 0$ if $\boldsymbol{\alpha} \neq \boldsymbol{\alpha}_0$. On the other hand, if $\mathrm{pr}((\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T \boldsymbol{W} = 0 | U = X_u = 0) = 1$, it is clear that $\mathrm{E}((\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T \boldsymbol{W}\boldsymbol{W}^T (\boldsymbol{\alpha} - \boldsymbol{\alpha}_0) | U = 0, X_u = 0) = 0$, which is a contradiction. Therefore $\mathrm{pr}((\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T \boldsymbol{W} = 0 | U = X_u = 0) < 1$ and thus $\mathrm{pr}((\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T \boldsymbol{W} \neq 0 | U = X_u = 0) > 0$. Since $0 < F_0(u) < 1$, $0 < \mathrm{E}[U] < 1$ from Assumption (C.I), we have $\mathrm{pr}((\boldsymbol{\alpha} - \boldsymbol{\alpha}_0)^T \boldsymbol{W} \neq 0, U = X_u = 0) > 0$.

Similarly, for cases (ii), (iii) and (iv), by Assumption (C.I),

$$\begin{aligned}
\mathrm{pr}(\boldsymbol{\eta}^T \boldsymbol{Z}_c \neq \boldsymbol{\eta}_0 \boldsymbol{Z}_{c_0}) &\geq \mathrm{pr}((\beta - \beta_0)U \neq 0, X_u = 0) \\
&\geq \mathrm{pr}(U = 1, X_u = 0) > 0, && \text{case (ii)}, \\
\mathrm{pr}(\boldsymbol{\eta}^T \boldsymbol{Z}_c \neq \boldsymbol{\eta}_0 \boldsymbol{Z}_{c_0}) &\geq \mathrm{pr}((\gamma - \gamma_0)X_c \neq 0, \gamma_0(X_c - X_{c_0}) = 0) \\
&\geq \mathrm{pr}(X_\ell = 1) > 0, && \text{case (iii)}, \\
\mathrm{pr}(\boldsymbol{\eta}^T \boldsymbol{Z}_c \neq \boldsymbol{\eta}_0 \boldsymbol{Z}_{c_0}) &\geq \mathrm{pr}(\gamma_0(X_c - X_{c_0}) \neq 0) \\
&\geq \mathrm{pr}(X_c = 0, X_{c_0} = 1) + \mathrm{pr}(X_{c_0} = 0, X_c = 1) > 0, && \text{case (iv)},
\end{aligned}$$

where the last inequality holds, since $F(\cdot)$ is continuous and differentiable with a positive derivative at a small neighbor of $c_0$ from Assumption (C.I). Then by [41, lemma 5.35], $(c_0, \boldsymbol{\eta}_0)$ is the unique maximizer of $M(c, \boldsymbol{\eta})$.

As $M(c, \boldsymbol{\eta})$ is continuous in $(c, \boldsymbol{\eta})$ due to Assumption (C.I), the unique maximizer $(c_0, \boldsymbol{\eta}_0)$ must be well separated over any compact set, i.e. condition (A.III) holds with $\iota = 0$.

**Non-identifiable case.** Similar to the identifiable case, $(c_0, \boldsymbol{\eta}_0)$ is the unique maximizer of $M(c, \boldsymbol{\eta})$ in the non-identifiable case with respect to $d_1$. Next we show the maximizer $(c_0, \boldsymbol{\eta}_0)$ is well-separated, with respect to $d_1$, in the compact set $K_\delta$, for any $\delta > 0$.

Fix some $\delta > 0$. Suppose there exists $\epsilon > 0$ such that

$$M(c_0, \boldsymbol{\eta}_0) = \sup_{\epsilon \le d_1((c,\boldsymbol{\eta}),(c_0,\boldsymbol{\eta}_0)) \le \delta} M(c, \boldsymbol{\eta}).$$

Define $K_{\epsilon,\delta} := \{(c, \boldsymbol{\eta}) \in K_\delta : d_1((c, \boldsymbol{\eta}), (c_0, \boldsymbol{\eta}_0)) \ge \epsilon\}$. Then by definition, there exists a sequence $(\tilde{c}_n, \tilde{\boldsymbol{\eta}}_n),\ n \ge 1$, in $K_{\epsilon,\delta}$, such that $M(\tilde{c}_n, \tilde{\boldsymbol{\eta}}_n) \to M(c_0, \boldsymbol{\eta}_0)$. As $K_{\epsilon,\delta}$ is a compact set, there exists a sub-sequence $(\tilde{c}_{n_k}, \tilde{\boldsymbol{\eta}}_{n_k}),\ k \ge 1$ and $(\tilde{\tilde{c}}_0, \tilde{\tilde{\boldsymbol{\eta}}}_0) \in K_{\epsilon,\delta}$ such that $d_1((\tilde{c}_{n_k}, \tilde{\boldsymbol{\eta}}_{n_k}), (\tilde{\tilde{c}}_0, \tilde{\tilde{\boldsymbol{\eta}}}_0)) \to 0$. By the continuity of $M(c, \boldsymbol{\eta})$, we have $M(\tilde{\tilde{c}}_0, \tilde{\tilde{\boldsymbol{\eta}}}_0) = \lim_{k\to\infty} M(\tilde{c}_{n_k}, \tilde{\boldsymbol{\eta}}_{n_k}) = M(c_0, \boldsymbol{\eta}_0)$, which contradicts with the fact that $(c_0, \boldsymbol{\eta}_0)$ is the unique maximizer. Therefore, for all $0 < \epsilon < \delta$,

$$M(c_0, \boldsymbol{\eta}_0) > \sup_{\epsilon \le d_1((c,\boldsymbol{\eta}),(c_0,\boldsymbol{\eta}_0)) \le \delta} M(c, \boldsymbol{\eta}),$$

i.e. condition (A.III) holds with $\iota = 1$. □

Next we provide a few lemmas under the bootstrap setup, under which $Q_n$ takes the form as follows:

$$\begin{aligned}(\boldsymbol{W}_{n,1}, U_{n,1}, X_{n,1}) &\sim \frac{1}{n}\sum_{i=1}^{n} \delta_{(\boldsymbol{W}_i, U_i, X_i)}, \qquad \text{(B.1)}\\ Y_{n,1} | \boldsymbol{W}_{n,1}, U_{n,1}, X_{n,1} &\sim \exp(Y_{n,1}(\boldsymbol{\eta}_n^T \boldsymbol{Z}_{n,1,c_n}) - \phi(\boldsymbol{\eta}_n^T \boldsymbol{Z}_{n,1,c_n})),\end{aligned}$$

with respect to the measure $\nu$, where $(c_n, \boldsymbol{\eta}_n), n \in \mathbb{N}$ is a sequence in $[\ell, u] \times \mathbb{R}^{d+2}$.

In what follows in this subsection, for the identifiable case, we establish that given almost every sequence of $\boldsymbol{\mathcal{D}}_i = (\boldsymbol{\mathcal{W}}_i, U_i, X_i),\ i \ge 1$, for every sequence $(c_n, \boldsymbol{\eta}_n) \to (c_0, \boldsymbol{\eta}_0)$, certain condition holds for the $Q_n$ above; in other words, the null set is common to every sequence $(c_n, \boldsymbol{\eta}_n), n \in \mathbb{N}$ with the property that $(c_n, \boldsymbol{\eta}_n) \to (c_0, \boldsymbol{\eta}_0)$. Similar comment applies to the non-identifiable case. Further, note that *the distribution of* $(\boldsymbol{W}_{n,1}, U_{n,1}, X_{n,1})$ *does not depend on* $(c_n, \boldsymbol{\eta}_n), n \in \mathbb{N}$.

The following Lemma verifies condition (A.I) under the bootstrap setup, which only concerns *covariates*, and does not depend on $(c_n, \boldsymbol{\eta}_n), n \in \mathbb{N}$.

**Lemma B.2.** *Suppose* (C.I) *holds. Then condition* (A.I) *holds for* $Q_n$ *in* (B.1)*, given almost every sequence of* $\boldsymbol{\mathcal{D}}_i,\ i \ge 1$.

*Proof.* The condition $\sup_n \|\boldsymbol{W}_{n,1}\| \le C_w$ follows from Assumption (C.I). For $j \in \{0, 1\}$, by the strong law of large numbers, we have that almost surely,

$$\mathrm{E}[U_{n,1}] = \frac{1}{n}\sum_{i=1}^{n} U_i \to \mathrm{E}[U]$$

$$\mathrm{pr}(X_{n,1} \le \ell, U_{n,1} = j) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \le \ell, U_i = j) \to \mathrm{pr}(X \le \ell, U = j)$$

$$\mathrm{pr}(X_{n,1} \leq u, U_{n,1} = j) = \frac{1}{n}\sum_{i=1}^{n} I(X_i \leq u, U_i = j) \to \mathrm{pr}(X \leq u, U = j).$$

By Assumption (C.I), the first two conditions in (A.I) hold almost surely.

Next we prove the condition on $\lambda_{\min}(\mathrm{E}[\boldsymbol{W}_{n,1}\boldsymbol{W}_{n,1}^T|U_{n,1} = j, X_{n,1} \leq \ell])$, and the other case on $\lambda_{\min}(\mathrm{E}[\boldsymbol{W}_{n,1}\boldsymbol{W}_{n,1}^T|U_{n,1} = j, X_{n,1} > u])$ can be proved in a similar way. For $Q_n$ in (B.1) and $j \in \{0, 1\}$, by the strong law of large numbers, almost surely,

$$\mathrm{E}[\boldsymbol{W}_{n,1}\boldsymbol{W}_{n,1}^T|U_{n,1} = j, X_{n,1} \leq \ell] = \frac{\sum_{i=1}^{n} I(U_i = j, X_i \leq \ell)\boldsymbol{W}_i\boldsymbol{W}_i^T}{\sum_{i=1}^{n} I(U_i = j, X_i \leq \ell)}$$
$$\to \mathrm{E}[\boldsymbol{W}\boldsymbol{W}^T|U = j, X \leq \ell], \quad \text{almost surely.}$$

As $\liminf_n \lambda_{\min}(\cdot)$ is a continuous function in $(\ell^\infty(\mathbb{R}))^{d\times d}$, from Assumption (C.I), we have $\liminf_n \lambda_{\min}(\mathrm{E}[\boldsymbol{W}_{n,1}\boldsymbol{W}_{n,1}^T|U_{n,1} = j, X_{n,1} \leq \ell]) > 0$ almost surely.

Finally, by [42, Theorem 2.4.3], uniformly over $c_1, c_2 \in [\ell, u]$, given almost sure all sequence of $\boldsymbol{\mathcal{D}}_i,\ i \geq 1$,

$$\begin{aligned}&\mathrm{pr}(c_1 \wedge c_2 < X_{n,1} \leq c_1 \vee c_2)\\ =&\frac{1}{n}\sum_{i=1}^{n} I(c_1 \wedge c_2 < X_i \leq c_1 \wedge c_2) \to \mathrm{pr}(c_1 \wedge c_2 < X \leq c_1 \wedge c_2).\end{aligned}$$

From Assumption (C.I) and the mean value theorem, the last condition in (A.I) holds almost surely. □

**Lemma B.3.** *Consider the case under the null $H_0 : \lambda_0 = 0$, and $Q_n$ in* (B.1). *Assume* (C.I) *holds.*

1. *For the identifiable case, i.e., $\gamma_0 \neq 0$, condition* (A.II) *holds for every sequence $(c_n, \boldsymbol{\eta}_n) \to (c_0, \boldsymbol{\eta}_0)$, given almost every sequence of $\boldsymbol{\mathcal{D}}_i,\ i \geq 1$.*
2. *For the non-identifiable case, i.e., $\gamma_0 = 0$, condition* (A.II) *holds for every sequence $(c_n, \boldsymbol{\eta}_n)$ such that $\boldsymbol{\eta}_n \to \boldsymbol{\eta}_0$, given almost every sequence of $\boldsymbol{\mathcal{D}}_i,\ i \geq 1$.*

*Proof.* We start with the first claim. Fix some $\delta > 0$, and define the following function class on $S$:

$$\{\tilde{\phi}_{c,\boldsymbol{\eta}} : (c, \boldsymbol{\eta}) \in K_\delta\}, \quad \text{where } \tilde{\phi}_{c,\boldsymbol{\eta}}(y, \boldsymbol{w}, u, x) = \boldsymbol{\eta}^T\boldsymbol{z}_c\phi'(\boldsymbol{\eta}_0^T\boldsymbol{z}_{c_0}) - \phi(\boldsymbol{\eta}^T\boldsymbol{z}_c).$$

For any $(c, \boldsymbol{\eta}) \in K_\delta$, by the definition of $Q_n$ in (B.1), we have the following decomposition for $Q_n\varphi_{c,\boldsymbol{\eta}} - P\varphi_{c,\boldsymbol{\eta}}$:

$$\begin{aligned}&Q_n\left[\phi'(\boldsymbol{\eta}_n^T\boldsymbol{\mathcal{Z}}_{c_n})\boldsymbol{\eta}^T\boldsymbol{\mathcal{Z}}_c - \phi(\boldsymbol{\eta}^T\boldsymbol{\mathcal{Z}}_c))\right] - P\left[\phi'(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_0})\boldsymbol{\eta}^T\boldsymbol{\mathcal{Z}}_c - \phi(\boldsymbol{\eta}^T\boldsymbol{\mathcal{Z}}_c)\right]\\ =&Q_n[(\phi'(\boldsymbol{\eta}_n^T\boldsymbol{\mathcal{Z}}_{c_n}) - \phi'(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_n}))\boldsymbol{\eta}^T\boldsymbol{\mathcal{Z}}_c] + Q_n[(\phi'(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_n}) - \phi'(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_0}))\boldsymbol{\eta}^T\boldsymbol{\mathcal{Z}}_c]\\ &+ (Q_n - P)[\tilde{\phi}_{c,\boldsymbol{\eta}}]\end{aligned}$$

$$= E^{(1)}_{n,c,\boldsymbol{\eta}} + E^{(2)}_{n,c,\boldsymbol{\eta}} + E^{(3)}_{n,c,\boldsymbol{\eta}}.$$

Thus it suffices to show that $\sup_{(c,\boldsymbol{\eta})\in K_\delta} |E^{(k)}_{n,c,\boldsymbol{\eta}}| \to 0$ for every sequence $(c_n, \boldsymbol{\eta}_n) \to (c_0, \boldsymbol{\eta}_0)$, almost surely, for $k = 1, 2, 3$. We first observe that for any $(c, \boldsymbol{\eta}) \in K_\delta$, $\|\boldsymbol{\eta}\| \leq \|\boldsymbol{\eta}_0\| + \delta$ and $\|\boldsymbol{\mathcal{Z}}_c\| \leq C_w + 2$, and that $\phi$ is infinitely differentiable on $\mathbb{R}$.

$\underline{E^{(1)}_{n,c,\boldsymbol{\eta}}}$ For every sequence $(c_n, \boldsymbol{\eta}_n) \to (c_0, \boldsymbol{\eta}_0)$, by the mean value form of the Taylor Theorem, there exist some constants $C, C'$ that may depend on the sequence $(c_n, \boldsymbol{\eta}_n),\ n \geq 1$, but not on $n$ such that

$$\begin{aligned}
\sup_{(c,\boldsymbol{\eta})\in K_\delta} |E^{(1)}_{n,c,\boldsymbol{\eta}}| &= \sup_{(c,\boldsymbol{\eta})\in K_\delta} |Q_n[(\phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}) - \phi'(\boldsymbol{\eta}_0^T \boldsymbol{\mathcal{Z}}_{c_n}))\boldsymbol{\eta}^T \boldsymbol{\mathcal{Z}}_c]| \\
&\leq C \sup_{(c,\boldsymbol{\eta})\in K_\delta} |Q_n[\boldsymbol{\eta}^T \boldsymbol{\mathcal{Z}}_c \boldsymbol{\mathcal{Z}}_{c_n}^T (\boldsymbol{\eta}_n - \boldsymbol{\eta}_0)]| \\
&\leq C' \|\boldsymbol{\eta}_n - \boldsymbol{\eta}_0\| \to 0.
\end{aligned}$$

This holds surely, not just almost surely.

$\underline{E^{(2)}_{n,c,\boldsymbol{\eta}}}$ Note that $\boldsymbol{\eta}_0^T \boldsymbol{\mathcal{Z}}_{c_n} - \boldsymbol{\eta}_0^T \boldsymbol{\mathcal{Z}}_{c_0} = \gamma_0(\mathcal{X}_{c_n} - \mathcal{X}_{c_0})$. Then there exists a constant $C > 0$, that does not depend on $(c_n, \boldsymbol{\eta}_n),\ n \geq 1$, such that

$$\begin{aligned}
\sup_{(c,\boldsymbol{\eta})\in K_\delta} |E^{(2)}_{n,c,\boldsymbol{\eta}}| &= \sup_{(c,\boldsymbol{\eta})\in K_\delta} |Q_n[(\phi'(\boldsymbol{\eta}_0^T \boldsymbol{\mathcal{Z}}_{c_n}) - \phi'(\boldsymbol{\eta}_0^T \boldsymbol{\mathcal{Z}}_{c_0}))\boldsymbol{\eta}^T \boldsymbol{\mathcal{Z}}_c]| \\
&\leq C \sup_{(c,\boldsymbol{\eta})\in K_\delta} Q_n[I(c_n \wedge c_0 < \mathcal{X} \leq c_n \vee c_0)] \\
&\leq C \sup_{(c_1,c_2)\in[\ell,u]^2 : |c_1 - c_2| \leq |c_n - c_0|} Q_n[I(c_1 \wedge c_2 < \mathcal{X} \leq c_1 \vee c_2)].
\end{aligned}$$

In Lemma B.2, we showed that if $c_n \to c_0$, then the last term converges to zero almost surely, where the null set does not depend on $(c_n, \boldsymbol{\eta}_n),\ n \geq 1$.

$\underline{E^{(3)}_{n,c,\boldsymbol{\eta}}}$ Note that $\tilde{\phi}_{c,\boldsymbol{\eta}}$ only depends on $(\boldsymbol{w}, u, x)$, and that under $Q_n$ in (B.1), $(\boldsymbol{W}_{n,1}, U_{n,1}, X_{n,1})$ is a random pair from the empirical measure induced by $(\boldsymbol{W}_i, U_i, X_i),\ i \in [n]$. Similar to Lemma A.1, $\{\tilde{\phi}_{c,\boldsymbol{\eta}} : (c, \boldsymbol{\eta}) \in K_\delta\}$ is a strongly Glivenko-Cantelli class. Therefore $\sup_{(c,\boldsymbol{\eta})\in K_\delta} |E^{(3)}_{n,c,\boldsymbol{\eta}}| \to 0$ almost surely, where the null set does not depend on $(c_n, \boldsymbol{\eta}_n),\ n \geq 1$.

Next, we consider the second claim. The decomposition continues to hold, and the exact same argument shows that $\sup_{(c,\boldsymbol{\eta})\in K_\delta} |E^{(k)}_{n,c,\boldsymbol{\eta}}| \to 0$ for every sequence $(c_n, \boldsymbol{\eta}_n)$ such that $\boldsymbol{\eta}_n \to \boldsymbol{\eta}_0$, almost surely, for $k = 1, 3$. If $\gamma_0 = 0$, then by definition $\sup_{(c,\boldsymbol{\eta})\in K_\delta} |E^{(2)}_{n,c,\boldsymbol{\eta}}| = 0$. Thus the proof is complete. □

**Lemma B.4.** *Consider the identifiable case under the null, i.e.* $\gamma_0 \neq 0$ *and* $\lambda_0 = 0$*, and* $Q_n$ *in* (B.1)*. Assume* (C.I) *holds. Then for any* $\epsilon \in (0, 1/4)$*, condition* (A.V) *holds for every sequence* $(c_n, \boldsymbol{\eta}_n)$ *such that* $c_n \to c_0$*, given almost all sequence of* $\boldsymbol{\mathcal{D}}_i,\ i \geq 1$*.*

*Proof.* Fix any $\epsilon \in (0, 1/4)$ and a sequence sequence $(c_n, \boldsymbol{\eta}_n)$ such that $c_n \to c_0$. We first focus on those $c \in (c_n + n^{-1+2\epsilon}, c_n + 1]$.

Define $\mathcal{K}(\cdot) = I(\cdot \in [-1, 0))$, and observe that for $Q_n$ defined in (B.1), almost surely,

$$Q_n[I(c \wedge c_n < \mathcal{X} \le c \vee c_n)] = \frac{1}{n}\sum_{i=1}^{n} I(c_n < X_i \le c) = \frac{1}{n}\sum_{i=1}^{n} \mathcal{K}\left(\frac{c_n - X_i}{|c - c_n|}\right)$$
$$\in (F(c) - F(c_n)) \mp \left|\frac{1}{n}\sum_{i=1}^{n} \mathcal{K}\left(\frac{c_n - X_i}{|c - c_n|}\right) - (F(c) - F(c_n))\right|.$$

Note that

$$\sup_{n^{-1+2\epsilon} < |c - c_n| \le 1} \frac{1}{|c - c_n|}\left|\frac{1}{n}\sum_{i=1}^{n} \mathcal{K}\left(\frac{c_n - X_i}{|c - c_n|}\right) - (F(c) - F(c_n))\right|$$
$$\le \sup_{n^{-1+2\epsilon} < h \le 1} \sup_{c \in [\ell, u]} \left|\frac{1}{n}\sum_{i=1}^{n} \frac{1}{h}\mathcal{K}\left(\frac{c - X_i}{h}\right) - \mathrm{E}\left[\frac{1}{h}\mathcal{K}\left(\frac{c - X_i}{h}\right)\right]\right|.$$

By [16, Theorem 1], the last term converges to zero almost surely, where the null set does not depend on $(c_n, \boldsymbol{\eta}_n)$, $n \ge 1$.

Further, due to the mean value theorem and Assumption (C.I),

$$0 < \inf_{n^{-1+2\epsilon} < |c - c_n| \le 1} \frac{1}{|c - c_n|}|F(c) - F(c_n)|$$
$$\le \sup_{n^{-1+2\epsilon} < |c - c_n| \le 1} \frac{1}{|c - c_n|}|F(c) - F(c_n)| < \infty.$$

As a result, we have that for those $c \in (c_n + n^{-1+2\epsilon}, c_n + 1]$, condition (A.V) holds for every $c_n \to c_0$, almost surely. The proof for the case where $c \in [c_n - 1, c_n - n^{-1+2\epsilon})$ is similar if we define $\mathcal{K}(\cdot) = I(\cdot \in [0, 1))$. □

**Lemma B.5.** *Consider the identifiable case under the null, i.e. $\gamma_0 \neq 0$ and $\lambda_0 = 0$, and $Q_n$ in* (B.1)*. Assume* (C.I) *holds. Then condition* (A.VI) *holds for every sequence $(c_n, \boldsymbol{\eta}_n)$ such that $c_n \to c_0$, given almost all sequence of $\boldsymbol{\mathcal{D}}_i$, $i \ge 1$.*

*Proof.* Fix any $\delta \in [0, 1/4)$, and any sequence $(c_n, \boldsymbol{\eta}_n)$ such that $c_n \to c_0$. By condition (C.I), there exists a small neighbourhood $\mathcal{U}_0$ around $c_0$ on which $X$ has a density. Since $c_n \to c_0$, for large $n$ and $c_n \in \mathcal{U}_0$.

Define $\mathcal{F}_{4,\epsilon} = \{\mathbb{R} \ni x \to I(c_1 < x < c_2) : (c_1, c_2) \in \mathcal{U}_0^2, |c_1 - c_2| \le \epsilon\}$. Note that for large $n$,

$$\sqrt{m_n} Q_n[I(c_n - m_n^{-1+2\delta} < \mathcal{X} < c_n + m_n^{-1+2\delta})]$$
$$\le \sup_{|c_1 - c_2| \le 2m_n^{-1+2\delta}} \sqrt{m_n} Q_n[I(c_1 \wedge c_2 < \mathcal{X} \le c_1 \vee c_2)]$$
$$\le \sup_{|c_1 - c_2| \le 2m_n^{-1+2\delta}} \sqrt{m_n} P[I(c_1 \wedge c_2 < \mathcal{X} \le c_1 \vee c_2)] + \sqrt{m_n}\|Q_n - P\|_{\mathcal{F}_{4, 2m_n^{-1+2\delta}}},$$

where the first term converges to 0 by the mean value theorem, and the second term converges to zero almost surely by a similar argument as in [27, Proof of

Lemma 10], where the null set does not depend on $(c_n, \boldsymbol{\eta}_n), n \geq 1$. Then the proof is complete. □

**Lemma B.6.** *Consider the identifiable case under the null, i.e.* $\gamma_0 \neq 0$ *and* $\lambda_0 = 0$*, and* $Q_n$ *in* (B.1)*. Assume* (C.I) *holds. Then condition* (A.VII) *holds for every sequence* $(c_n, \boldsymbol{\eta}_n) \to (c_0, \boldsymbol{\eta}_0)$*, given almost all sequence of* $\boldsymbol{\mathcal{D}}_i,\ i \geq 1$*.*

*Proof.* We first consider the second statement in condition (A.VII). Fix a sequence $(c_n, \boldsymbol{\eta}_n) \to (c_0, \boldsymbol{\eta}_0)$. Then for large $n$, $\|\boldsymbol{\eta}_n - \boldsymbol{\eta}_0\| \leq 1$. Let $\mathcal{F}_5 = \{S \ni (y, \boldsymbol{w}, u, x) \to \phi''(\boldsymbol{\eta}^T z_c)\tilde{\boldsymbol{z}}_c\tilde{\boldsymbol{z}}_c^T : (c, \boldsymbol{\eta}) \in K_1\}$. Note that for large $n$,

$$
\begin{aligned}
&\|Q_n[\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})\tilde{\boldsymbol{\mathcal{Z}}}_{c_n}\tilde{\boldsymbol{\mathcal{Z}}}_{c_n}^T] - P[\phi''(\boldsymbol{\eta}_0^T \boldsymbol{\mathcal{Z}}_{c_0})\tilde{\boldsymbol{\mathcal{Z}}}_{c_0}\tilde{\boldsymbol{\mathcal{Z}}}_{c_0}^T]\| \\
\leq &\|Q_n - P\|_{\mathcal{F}_5} + \|P[\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})\tilde{\boldsymbol{\mathcal{Z}}}_{c_n}\tilde{\boldsymbol{\mathcal{Z}}}_{c_n}^T] - P[\phi''(\boldsymbol{\eta}_0^T \boldsymbol{\mathcal{Z}}_{c_0})\tilde{\boldsymbol{\mathcal{Z}}}_{c_0}\tilde{\boldsymbol{\mathcal{Z}}}_{c_0}^T]\|.
\end{aligned}
$$

Note that functions in $\mathcal{F}_5$ only depends on $(\boldsymbol{w}, u, x)$, and that under $Q_n$ in (B.1), $(\boldsymbol{W}_{n,1}, U_{n,1}, X_{n,1})$ is a random pair from the empirical measure induced by $(\boldsymbol{W}_i, U_i, X_i),\ i \in [n]$. Similar to Lemma A.1, $\mathcal{F}_5$ is a strongly Glivenko-Cantelli class, and thus $\|Q_n - P\|_{\mathcal{F}_5}$ converges to zero almost surely, where the null set does not depend on $(c_n, \boldsymbol{\eta}_n),\ n \geq 1$. Further, the last term converges to 0 due to (C.I). Then the proof for the second statement in condition (A.VII) is complete.

Now we consider the first statement in condition (A.VII). Note that for any $(c_1, \boldsymbol{\eta}_1), (c_2, \boldsymbol{\eta}_2) \in [\ell, u] \times \mathbb{R}^{d+2}$, we have

$$
\begin{aligned}
Q_n[(\mathcal{Y} - \phi'(\boldsymbol{\eta}_1^T \boldsymbol{\mathcal{Z}}_{c_1}))(\mathcal{Y} - \phi'(\boldsymbol{\eta}_2^T \boldsymbol{\mathcal{Z}}_{c_2}))\tilde{\boldsymbol{\mathcal{Z}}}_{c_1}\tilde{\boldsymbol{\mathcal{Z}}}_{c_2}^T] &= Q_n[\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n})\tilde{\boldsymbol{\mathcal{Z}}}_{c_1}\tilde{\boldsymbol{\mathcal{Z}}}_{c_2}^T] \\
+ Q_n[\big((\phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}) - \phi'(\boldsymbol{\eta}_1^T \boldsymbol{\mathcal{Z}}_{c_1}))(\phi'(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_n}) - \phi'(\boldsymbol{\eta}_2^T \boldsymbol{\mathcal{Z}}_{c_2}))\big)\, \tilde{\boldsymbol{\mathcal{Z}}}_{c_1}\tilde{\boldsymbol{\mathcal{Z}}}_{c_2}^T].
\end{aligned}
$$

For those functions inside $[\cdot]$ on the right hand side, they only depend on $(\boldsymbol{w}, u, x)$. Then by a similar argument as above, almost surely,

$$
\begin{aligned}
\lim_{n\to\infty} Q_n[(\mathcal{Y} - \phi'(\boldsymbol{\eta}_1^T \boldsymbol{\mathcal{Z}}_{c_1}))(\mathcal{Y} - \phi'(\boldsymbol{\eta}_2^T \boldsymbol{\mathcal{Z}}_{c_2}))\tilde{\boldsymbol{\mathcal{Z}}}_{c_1}\tilde{\boldsymbol{\mathcal{Z}}}_{c_2}^T] \ &= \ P[\phi''(\boldsymbol{\eta}_0^T \boldsymbol{\mathcal{Z}}_{c_0})\tilde{\boldsymbol{\mathcal{Z}}}_{c_1}\tilde{\boldsymbol{\mathcal{Z}}}_{c_2}^T] \\
+ P[\big((\phi'(\boldsymbol{\eta}_0^T \boldsymbol{\mathcal{Z}}_{c_0}) - \phi'(\boldsymbol{\eta}_1^T \boldsymbol{\mathcal{Z}}_{c_1}))(\phi'(\boldsymbol{\eta}_0^T \boldsymbol{\mathcal{Z}}_{c_0}) - \phi'(\boldsymbol{\eta}_2^T \boldsymbol{\mathcal{Z}}_{c_2}))\big)\, \tilde{\boldsymbol{\mathcal{Z}}}_{c_1}\tilde{\boldsymbol{\mathcal{Z}}}_{c_2}^T],
\end{aligned}
$$

where the null set does not depend on $(c_n, \boldsymbol{\eta}_n),\ n \geq 1$. Then the proof is complete. □

**Lemma B.7.** *Consider the non-identifiable case under the null, i.e.* $\gamma_0 = \lambda_0 = 0$*, and* $Q_n$ *in* (B.1)*. Assume* (C.I) *holds. Then condition* (A.IX) *holds for every sequence* $(c_n, \boldsymbol{\eta}_n)$ *such that* $\boldsymbol{\eta}_n \to \boldsymbol{\eta}_0$*, given almost all sequence of* $\boldsymbol{\mathcal{D}}_i,\ i \geq 1$*.*

*Proof.* Fix a sequence $(c_n, \boldsymbol{\eta}_n)$ such that $\boldsymbol{\eta}_n \to \boldsymbol{\eta}_0$. Then for large $n$, $\|\boldsymbol{\eta}_n - \boldsymbol{\eta}_0\| \leq 1$. Consider the class $\mathcal{F}_6 = \{S \ni (y, \boldsymbol{w}, u, x) \to \phi''(\boldsymbol{\eta}^T \boldsymbol{z}_{c_3})\tilde{\boldsymbol{z}}_{c_1}\tilde{\boldsymbol{z}}_{c_2} : (c_1, c_2, c_3) \in [\ell, u]^3, \|\boldsymbol{\eta} - \boldsymbol{\eta}_0\| \leq 1\}$. Note that for large $n$,

$$
\sup_{(c_1,c_2,c_3)\in[\ell,u]^3} |Q_n[\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_3})\tilde{\boldsymbol{\mathcal{Z}}}_{c_1}\tilde{\boldsymbol{\mathcal{Z}}}_{c_2}^T] - P[\phi''(\boldsymbol{\eta}_0^T \boldsymbol{\mathcal{Z}}_{c_3})\tilde{\boldsymbol{\mathcal{Z}}}_{c_1}\tilde{\boldsymbol{\mathcal{Z}}}_{c_2}^T]|
$$

$$\leq \|Q_n - P\|_{\mathcal{F}_6} + \sup_{(c_1,c_2,c_3)\in[\ell,u]^3} \left| P\left[ (\phi''(\boldsymbol{\eta}_n^T \boldsymbol{\mathcal{Z}}_{c_3}) - \phi''(\boldsymbol{\eta}_0^T \boldsymbol{\mathcal{Z}}_{c_3})) \tilde{\boldsymbol{\mathcal{Z}}}_{c_1} \tilde{\boldsymbol{\mathcal{Z}}}_{c_2}^T \right] \right|.$$

Note that functions in $\mathcal{F}_6$ only depends on $(\boldsymbol{w}, u, x)$, and that under $Q_n$ in (B.1), $(\boldsymbol{W}_{n,1}, U_{n,1}, X_{n,1})$ is a random pair from the empirical measure induced by $(\boldsymbol{W}_i, U_i, X_i),\ i \in [n]$. Similar to Lemma A.1, $\mathcal{F}_6$ is a strongly Glivenko-Cantelli class, and thus $\|Q_n - P\|_{\mathcal{F}_6}$ converges to zero almost surely, where the null set does not depend on $(c_n, \boldsymbol{\eta}_n),\ n \geq 1$. Further, since $\phi$ is infinitely differentiable on $\mathbb{R}$, the last term converges to 0 since $\sup_{c\in[\ell,u]} \|\boldsymbol{\mathcal{Z}}_c\| \leq C_w + 2 < \infty$ due to (C.I). Then the proof is complete. □

## Appendix C: Proofs regarding size analyais

Consider the cases under the null, i.e. $\lambda_0 = 0$. For $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n)$, the MLE estimator based on the original data, we consider the triangular setup, where $Q_n = P$, $m_n = n$ and $\boldsymbol{\mathcal{D}}_{n,i} = \boldsymbol{\mathcal{D}}_i,\ i \in [n]$. For $(\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*)$, the MLE estimator based on the bootstrapped data, $Q_n$ is defined in (B.1) with $(c_n, \boldsymbol{\eta}_n) = (\hat{c}_n, \hat{\boldsymbol{\eta}}_n)$, and $\boldsymbol{\mathcal{D}}_{n,i} = (Y_i^*, \boldsymbol{W}_i^*, U_i^*, X_i^*),\ i \in [m_n]$.

All results, except for the limiting distribution of $n(\hat{c}_n - c_0)$ in the identifiable case, follow immediately from the results in the triangular array setup in Section A and the verification of conditions in Section B. Thus we defer the proof for the second claim in Theorem 2.1 to Section C.3.

### *C.1. Identifiable case - size analysis*

*Proof of the first claim in Theorem 2.1.* We consider the identifiable case under the null, i.e. $\gamma_0 \neq 0$ and $\lambda_0 = 0$, and the triangular array setup, where $m_n = n$, $\boldsymbol{\mathcal{D}}_{n,i} = \boldsymbol{\mathcal{D}}_i,\ i \in [n]$, and $Q_n = P$ with $(c_n, \boldsymbol{\eta}_n) = (c_0, \boldsymbol{\eta}_0)$. Then the MLE estimator $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n)$ and the score-type test statistic $S_n$ in (2.3), based on the original data, correspond to $(\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*)$ and $S_n^*$ in Theorem A.3.

Note that (A.I), (A.II), (A.IV) with $\iota = 0$, and (A.VII) trivially hold, due to (C.I). Further, by the mean value theorem, and due to (C.I), (A.V) with $\epsilon = 0$ and (A.VI) hold. Finally, (A.III) with $\iota = 0$ holds by Lemma B.1. Then the claim follows immediately from Theorem A.2 and A.3. □

*Proof of Theorem 3.1(i).* We consider the identifiable case under the null, i.e., $\gamma_0 \neq 0$ and $\lambda_0 = 0$, and the triangular array setup: $\boldsymbol{\mathcal{D}}_{n,i} = (Y_i^*, \boldsymbol{W}_i^*, U_i^*, X_i^*), i \in [m_n]$, and $Q_n$ is defined in (B.1) with $(c_n, \boldsymbol{\eta}_n) = (\hat{c}_n, \hat{\boldsymbol{\eta}}_n)$. Then the MLE estimator $(\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*)$ and the score-type test statistic $S_n^*$ in (3.1), based on the bootstrapped data, correspond to $(\hat{c}_n^*, \hat{\boldsymbol{\eta}}_n^*)$ and $S_n^*$ in Theorem A.3.

From Theorem 2.1, $d_0((\hat{c}_n, \hat{\boldsymbol{\eta}}_n), (c_0, \boldsymbol{\eta}_0)) \to 0$ in probability. Then for each sub-sequence $(\hat{c}_{n_k}, \hat{\boldsymbol{\eta}}_{n_k})$, there exists a further sub-sequence $(\hat{c}_{n_{k_\ell}}, \hat{\boldsymbol{\eta}}_{n_{k_\ell}})$ such that $d_0((\hat{c}_{n_{k_\ell}}, \hat{\boldsymbol{\eta}}_{n_{k_\ell}}), (c_0, \boldsymbol{\eta}_0)) \to 0$ almost surely.

We apply Theorem A.3 to $Q_{n_{k_\ell}}$ in (B.1) associated with this sub-sub-sequence $(\hat{c}_{n_{k_\ell}}, \hat{\boldsymbol{\eta}}_{n_{k_\ell}})$. First, by construction, condition (A.IV) holds almost surely. Further, conditions (A.I)-(A.III) and (A.V)- (A.VII) with $\iota = 0$ are verified by

Lemmas B.1-B.6 given almost all sequence of $\boldsymbol{\mathcal{D}}_i,\ i \geq 1$. Thus by Theorem A.3, almost surely,

$$\sup_{\boldsymbol{t}\in\mathbb{R}^{d+3}} |\mathrm{pr}_{|\boldsymbol{\mathcal{D}}}((\sqrt{m_{n_{k_\ell}}}(\hat{\boldsymbol{\eta}}^*_{n_{k_\ell}} - \hat{\boldsymbol{\eta}}_{n_{k_\ell}}), S^*_{n_{k_\ell}}) \leq \boldsymbol{t}) - \mathrm{pr}((\mathbb{Z}_{\boldsymbol{\eta}}, \mathbb{Z}_S) \leq \boldsymbol{t})| \to 0,$$

which completes the proof. □

### C.2. Non-identifiable case - size analysis

Note that for $B_1 = 0$, $(\widetilde{\mathbb{C}}, \widetilde{\mathbb{H}}, \widetilde{\mathbb{S}})$ in (A.14) has the same distribution as $(\mathbb{C}, \mathbb{H}, \mathbb{S})$ in (2.7).

*Proof of Theorem 2.2.* We consider the non-identifiable case under the null, i.e. $\gamma_0 = \lambda_0 = 0$, and the triangular array setup, where $m_n = n$, $\boldsymbol{\mathcal{D}}_{n,i} = \boldsymbol{\mathcal{D}}_i,\ i \in [n]$, and $Q_n = P$ with $(c_n, \boldsymbol{\eta}_n) = (c_0, \boldsymbol{\eta}_0)$. Then the MLE estimator $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n)$ and the score-type test statistic $S_n$ in (2.3), based on the original data, correspond to $(\hat{c}^*_n, \hat{\boldsymbol{\eta}}^*_n)$ and $S^*_n$ in Theorem A.5.

Note that (A.i), (A.ii), (A.iv) with $\iota = 1$, (A.viii), and (A.ix) trivially hold, due to (C.i). Finally, (A.iii) with $\iota = 1$ holds by Lemma B.1. Then the first claim follows immediately from Theorem A.5. □

*Proof of Theorem 3.1(ii).* We consider the non-identifiable case under the null, i.e. $\gamma_0 = \lambda_0 = 0$, and the triangular array setup, where $\boldsymbol{\mathcal{D}}_{n,i} = (Y^*_i, \boldsymbol{W}^*_i, U^*_i, X^*_i)$, $i \in [m_n]$, and $Q_n$ is defined in (B.1) with $(c_n, \boldsymbol{\eta}_n) = (\hat{c}_n, \hat{\boldsymbol{\eta}}_n)$. Then the MLE estimator $(\hat{c}^*_n, \hat{\boldsymbol{\eta}}^*_n)$ and the score-type test statistic $S^*_n$ in (3.1), based on the bootstrapped data, correspond to $(\hat{c}^*_n, \hat{\boldsymbol{\eta}}^*_n)$ and $S^*_n$ in Theorem A.5.

By Theorem 2.2, since $\gamma_0 = 0$, $\sqrt{n}\hat{\gamma}_n = O_{\mathrm{pr}}(1)$ and $m_n/n \to 0$, we have

$$\sqrt{m_n}\hat{\gamma}_n = o_{\mathrm{pr}}(1), \quad \|\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\| = o_{\mathrm{pr}}(1),$$

which implies that for each sub-sequence $(\hat{c}_{n_k}, \hat{\boldsymbol{\eta}}_{n_k})$, there exists a further sub-sequence $(\hat{c}_{n_{k_\ell}}, \hat{\boldsymbol{\eta}}_{n_{k_\ell}})$ such that

$$\sqrt{m_{n_{k_\ell}}}\hat{\gamma}_{n_{k_\ell}} \to 0, \quad \text{ and } \|\hat{\boldsymbol{\eta}}_{n_{k_\ell}} - \boldsymbol{\eta}_0\| \to 0, \text{ almost surely.}$$

We apply Theorem A.5 to $Q_{n_{k_\ell}}$ in (B.1) associated with this sub-sub-sequence $(\hat{c}_{n_{k_\ell}}, \hat{\boldsymbol{\eta}}_{n_{k_\ell}})$. First, by construction, conditions (A.viii) with $B_1 = 0$ and (A.iv) with $\tau = 1$ hold. Second, conditions (A.i)-(A.iii) and (A.ix) hold due to Lemmas B.1-B.3 and B.7. Then by Theorem A.5, almost surely,

$$\sup_{\boldsymbol{t}\in\mathbb{R}^{d+4}} |\mathrm{pr}_{|\boldsymbol{\mathcal{D}}}((\hat{c}^*_{n_{k_\ell}}, \sqrt{m_{n_{k_\ell}}}(\hat{\boldsymbol{\eta}}^*_{n_{k_\ell}} - \hat{\boldsymbol{\eta}}_{n_{k_\ell}}), S^*_{n_{k_\ell}}) \leq \boldsymbol{t}) - \mathrm{pr}((\mathbb{C}, \mathbb{H}, \mathbb{S}) \leq \boldsymbol{t})| \to 0,$$

which completes the proof. □

### *C.3. The weak limit of the cutpoint MLE in the identifiable case*

In this subsection, we prove the second claim in Theorem 2.1 regarding the limiting distribution of $n(\hat{c}_n - c_0)$ in the identifiable case.

Before the formal proof we review some definitions. For a closed interval $I \subset \mathbb{R}$, the space $\tilde{\mathcal{D}}_I$ is the collection of all functions on $I$, that are right-continuous with left limits, endowed with a metric $\tilde{d}_I$ below [21, 36, 44]. Let $\Lambda_I = \{\lambda : I \mapsto I | \lambda$ is strictly increasing, surjective and continuous$\}$ and write $\|\lambda\| := \sup_{s \neq t \in I} |\log \frac{\lambda(s)-\lambda(t)}{s-t}|$. Then for any $f_1, f_2 \in \tilde{\mathcal{D}}_I$:

$$\tilde{d}_I(f_1, f_2) := \inf_{\lambda \in \Lambda_I} \left\{ \sup_{c \in I} \{|f_1(c) - f_2(\lambda(c))|\} + \|\lambda\| \right\}.$$

Note that, endowed with the metric $\tilde{d}_I$, the space $\tilde{\boldsymbol{\mathcal{D}}}_I$ is Polish [37, 8, 21].

For $(\tau, \boldsymbol{h}) \in \mathbb{R} \times \mathbb{R}^{d+2}$, define

$$\bar{\varphi}_{\tau,\boldsymbol{h},n} := \sqrt{n}(\varphi_{c_0+\tau/n, \boldsymbol{\eta}_0 + \boldsymbol{h}/\sqrt{n}} - \varphi_{c_0,\boldsymbol{\eta}_0}).$$

Further, recall from Section A.3.2 that $\boldsymbol{g}^{(1)}_{c,\boldsymbol{\eta}} = (\mathcal{Y} - \phi'(\boldsymbol{\eta}^T \boldsymbol{\mathcal{Z}}_c))\boldsymbol{\mathcal{Z}}_c$ and $g^{(2)}_{c,\boldsymbol{\eta}} = (\mathcal{Y} - \phi'(\boldsymbol{\eta}^T \boldsymbol{\mathcal{Z}}_c))\mathcal{U}\mathcal{X}_c$, and that $\boldsymbol{V}^{(1)}_{c,\boldsymbol{\eta}}$ and $\boldsymbol{V}^{(2)}_{c,\boldsymbol{\eta}}$ are introduced in (2.5). Define

$$g^{(3)}_{c,\boldsymbol{\eta}} := g^{(2)}_{c,\boldsymbol{\eta}} - \boldsymbol{V}^{(2)}_{c,\boldsymbol{\eta}} (\boldsymbol{V}^{(1)}_{c,\boldsymbol{\eta}})^{-1} \boldsymbol{g}^{(1)}_{c,\boldsymbol{\eta}}.$$

Denote by $\mathbb{P}_n = n^{-1} \sum_{i=1}^n \delta_{\boldsymbol{\mathcal{D}}_i}$, the empirical measure on $S$ induced by $\boldsymbol{\mathcal{D}}_i, i \in [n]$, where $\delta_{\boldsymbol{\mathcal{D}}_i}$ is the Dirac measure at $\boldsymbol{\mathcal{D}}_i$, and define $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P)$.

*Proof of the second claim in Theorem 2.1.* By definition,

$$\begin{aligned} (n(\hat{c}_n - c_0), \sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)) = (\hat{\tau}_n, \hat{\boldsymbol{h}}_n) &:= \operatorname*{sargmax}_{(\tau,\boldsymbol{h}) \in [\ell_n, u_n] \times \mathbb{R}^{d+2}} \sqrt{n} \mathbb{P}_n \bar{\varphi}_{\tau,\boldsymbol{h},n} \\ = \operatorname*{sargmax}_{(\tau,\boldsymbol{h}) \in [\ell_n, u_n] \times \mathbb{R}^{d+2}} &\sqrt{n}\mathbb{P}_n[\bar{\varphi}_{\tau,\boldsymbol{h},n} - \bar{\varphi}_{\tau,0,n}] + \sqrt{n}\mathbb{P}_n[\bar{\varphi}_{\tau,0,n}]. \end{aligned}$$

where $\ell_n := n(\ell - c_0)$ and $u_n := n(u - c_0)$.

In the proof of Theorem A.3, whose conditions are verified in Section C.1, we showed that

$$S_n = \mathbb{G}_n \left[ g^{(3)}_{c_0,\boldsymbol{\eta}_0} \right] + o_{\mathrm{pr}}(1).$$

Further, for any compact hyper-rectangle $K \subset \mathbb{R} \times \mathbb{R}^{d+2}$, by the same asymptotic expansion argument as in Section A.3.2, we have

$$\sup_{(\tau,\boldsymbol{h}) \in K} \left| \sqrt{n}\mathbb{P}_n[\bar{\varphi}_{\tau,\boldsymbol{h},n} - \bar{\varphi}_{\tau,0,n}] - \left( \boldsymbol{h}^T \mathbb{G}_n \left[ \boldsymbol{g}^{(1)}_{c_0,\boldsymbol{\eta}_0} \right] - \frac{1}{2} \boldsymbol{h}^T \boldsymbol{V}^{(1)}_{c_0,\boldsymbol{\eta}_0} \boldsymbol{h} \right) \right| = o_{\mathrm{pr}}(1).$$

By a tedious but now standard argument as in [21, 36, 44] (see some discussions below), for any closed interval $I \subset \mathbb{R}$,

$$\begin{bmatrix} \mathbb{G}_n \left[ \boldsymbol{g}^{(1)}_{c_0,\boldsymbol{\eta}_0} \right] \\ \mathbb{G}_n \left[ g^{(3)}_{c_0,\boldsymbol{\eta}_0} \right] \\ \{\sqrt{n}\mathbb{P}_n[\bar{\varphi}_{\tau,0,n}] : \tau \in I\} \end{bmatrix} \rightsquigarrow \begin{bmatrix} \boldsymbol{V}^{(1)}_{c_0,\boldsymbol{\eta}_0} \mathbb{Z}_{\boldsymbol{\eta}} \\ \mathbb{Z}_S \\ \{\mathbb{D}(\tau) : \tau \in I\} \end{bmatrix}, \quad \text{in } \mathbb{R}^2 \times \tilde{\boldsymbol{\mathcal{D}}}_I, \tag{C.1}$$

where recall that $\mathbb{Z}_{\boldsymbol{\eta}}$, $\mathbb{Z}_S$ and $\mathbb{D}(\cdot)$ are defined before Theorem 2.1. Then the proof is complete by the continuous mapping theorem [36, Lemma A.3], and due to the fact that $(\hat{\tau}_n, \hat{\boldsymbol{h}}_n) = O_{\text{pr}}(1)$, which is the first claim in Theorem 2.1, and is proved in Section C.1. □

To show (C.1), as in [21, 36, 44], it involves two steps: (1). establishing the finite-dimensional distribution convergence using the characteristic function method; (2) establishing that $\{\sqrt{n}\mathbb{P}_n[\bar{\varphi}_{\tau,0,n}] : \tau \in I\}$ is uniformaly tight in $\tilde{\boldsymbol{\mathcal{D}}}_I$. The detailed arguments are similar to those in [21, 36, 44], and thus not repeated here.

Below we present an important calculation in order to show why we have the two-sided, compound Poisson process $\mathbb{D}(\cdot)$ in the limit, and why we assume the following for the second part of Theorem 2.1.

(C.II) the conditional distribution of $(\boldsymbol{W}, U)$ given $X = c$ is continuous in a neighbourhood of $c_0$ with respect to the weak convergence.

**Lemma C.1.** *Assume* (C.I) *and* (C.II) *hold. Then for two real numbers* $0 < \tau_1 < \tau_2$*,*

$$\sqrt{n}\mathbb{P}_n\left[\bar{\varphi}_{\tau_2,0,n} - \bar{\varphi}_{\tau_1,0,n}\right] \;\rightsquigarrow\; \mathbb{D}(\tau_2 - \tau_1).$$

*Proof.* Note that by definition,

$$\begin{aligned}&\sqrt{n}\mathbb{P}_n\left[\bar{\varphi}_{\tau_2,0,n} - \bar{\varphi}_{\tau_1,0,n}\right]\\ =&\sum_{i=1}^{n}\left(Y_i\gamma_0 - (\phi(\Theta_{i,+}) - \phi(\Theta_{i,-}))\right) I\left(\frac{\tau_1}{n} < X_i - c_0 \le \frac{\tau_2}{n}\right),\end{aligned}$$

where $\Theta_{i,+} := \boldsymbol{\alpha}_0^T\boldsymbol{W}_i + \beta_0 U_i + \gamma_0$, and $\Theta_{i,-} := \boldsymbol{\alpha}_0^T\boldsymbol{W}_i + \beta_0 U_i$. Then denoting $\Xi_+ := Y_1\gamma_0 - (\phi(\Theta_{1,+}) - \phi(\Theta_{1,-}))$, for any $t \in \mathbb{R}$, we have

$$\begin{aligned}&\mathrm{E}\left[\exp\left\{\sqrt{-1}t\sqrt{n}\mathbb{P}_n\left[\bar{\varphi}_{\tau_2,0,n} - \bar{\varphi}_{\tau_1,0,n}\right]\right\}\right]\\ =&\left(\mathrm{E}\left[\exp\left\{\sqrt{-1}t\Xi_+ I\left(\frac{\tau_1}{n} < X_1 - c_0 \le \frac{\tau_2}{n}\right)\right\}\right]\right)^n\\ =&\left(1 + \mathrm{E}\left[\exp\left\{\sqrt{-1}t\Xi_+\right\} - 1 \,\middle|\, \frac{\tau_1}{n} < X_1 - c_0 \le \frac{\tau_2}{n}\right]\mathrm{pr}\left(\frac{\tau_1}{n} < X_1 - c_0 \le \frac{\tau_2}{n}\right)\right)^n.\end{aligned}$$

Due to (C.I),

$$\lim_n n \times \mathrm{pr}\left(\frac{\tau_1}{n} < X_1 - c_0 \le \frac{\tau_2}{n}\right) = F_X'(c_0)(\tau_2 - \tau_1).$$

Further, due to (C.II) and since $Y$, given $(\boldsymbol{W}, U, X)$, belongs to the exponential family distribution (1.1), the conditional distribution of $(Y, \boldsymbol{W}, U)$ given $X = c$ is also continuous in a neighbourhood of $c_0$ with respect to the weak convergence, due to the (generalized) dominated convergence theorem. As a result, for any $t \in \mathbb{R}$,

$$\lim_n \mathrm{E}\left[\exp\left\{\sqrt{-1}t\Xi_+\right\} \,\middle|\, \frac{\tau_1}{n} < X_1 - c_0 \le \frac{\tau_2}{n}\right] = \mathrm{E}\left[\exp\left\{\sqrt{-1}t\xi_{1,+}\right\}\right].$$

Combining these two parts, for any $t \in \mathbb{R}$, we have

$$\begin{aligned}&\lim_{n\to\infty} \mathrm{E}\left[\exp\left\{\sqrt{-1}t\sqrt{n}\mathbb{P}_n\left[\bar{\varphi}_{\tau_2,0,n} - \bar{\varphi}_{\tau_1,0,n}\right]\right\}\right]\\ &= \exp\left(F_X'(c_0)(\tau_2-\tau_1)\left(\mathrm{E}\left[\exp\left\{\sqrt{-1}t\xi_{1,+}\right\}-1\right)\right]\right).\end{aligned}$$

Since the right hand side is the characteristic function for $\mathbb{D}(\tau_2 - \tau_1)$ for $\tau_2 > \tau_1$, the proof is complete. □

### *C.4. A special case in the non-identifiable case*

In this section we consider a special case where $\boldsymbol{W} = 1$ and $U, X$ are independent, and $X$ follows uniform distribution over $(0, 1)$. Then, in the non-identifiable case,

$$\mathrm{E}[\phi''(\boldsymbol{\alpha}_0^T\boldsymbol{W} + \beta_0 U)\tilde{\boldsymbol{Z}}_{c_1}\tilde{\boldsymbol{Z}}_{c_2}^T] = \begin{bmatrix} t_1 & t_2 & t_1c_2 & t_2c_2 \\ t_2 & t_2 & t_2c_2 & t_2c_2 \\ t_1c_1 & t_2c_1 & t_1(c_1\wedge c_2) & t_2(c_1\wedge c_2) \\ t_2c_1 & t_2c_1 & t_2(c_1\wedge c_2) & t_2(c_1\wedge c_2)\end{bmatrix},$$

which indicates $\{((\boldsymbol{\Delta}_c^{(1)})^T, \Delta_c^{(2)})^T : c \in [\ell, u]\}$ in Theorem 2.2 has the same distribution as

$$\left\{\begin{bmatrix} \sqrt{t_1}B(1) \\ \frac{t_2}{\sqrt{t_1}}B(1) + \sqrt{t_2 - \frac{t_2^2}{t_1}}\tilde{B}(1) \\ \sqrt{t_1}B(c) \\ \frac{t_2}{\sqrt{t_1}}B(c) + \sqrt{t_2 - t_2^2/t_1}\tilde{B}(c)\end{bmatrix} : c \in [\ell, u]\right\},$$

where $B(\cdot)$, $\tilde{B}(\cdot)$ are two independent Brownian motions, $t_1 = \mathrm{E}[\sigma''(\boldsymbol{\alpha}_0 + \beta_0 U)]$ and $t_2 = \mathrm{E}[\sigma''(\boldsymbol{\alpha}_0 + \beta_0 U)U]$.

Then $\mathbb{C}, \mathbb{H}, \mathbb{S}$ in (2.7) have the following representation:

$$\begin{aligned}&\mathbb{C} = \sup_{t\in[\ell,u]} \frac{(B(t) - tB(1))^2}{t(1-t)}, \qquad \mathbb{S} = \sqrt{t_2 - t_2^2/t_1}(\tilde{B}(\mathbb{C}) - \mathbb{C}\tilde{B}(1)),\\ &\mathbb{H} = (\boldsymbol{V}_{\mathbb{C},\boldsymbol{\eta}_0}^{(1)})^{-1}\begin{bmatrix}\sqrt{t_1}B(1) \\ \frac{t_2}{\sqrt{t_1}}B(1) - \sqrt{t_2 - t_2^2/t_1}\tilde{B}(1) \\ \sqrt{t_1}B(\mathbb{C})\end{bmatrix}.\end{aligned} \tag{C.2}$$

In other words, $\hat{c}_n$ converges to the maximizer $\mathbb{C}$ of a weighted Brownian bridge, and $S_n$ to the value of an independent Brownian bridge evaluated at $\mathbb{C}$, up to a multiplicative constant.

### *C.5. Proof for the inconsistency of standard bootstrap in the non-identifiable case*

Recall from Subsection 3.2 that $\mathcal{M}(\mathbb{R}^k)$ denotes the space of Borel probability measures on $\mathbb{R}^k$ where $k \geq 1$ is some integer, and we equip $\mathcal{M}(\mathbb{R}^k)$ with the

the Prokhorov metric $d_{\mathsf{Prok}}(\cdot,\cdot)$ [8, Section 6.5], which characterizes the weak convergence and under which $\mathcal{M}(\mathbb{R}^k)$ is a complete and separable metric space.

Recall from Subsection 3.2 that $\hat{\mathcal{R}}_n$ denotes the empirical distribution of the covariates $\{(\boldsymbol{W}_i, U_i, X_i) : i \in [n]\}$, that is, $\hat{\mathcal{R}}_n = n^{-1}\sum_{i=1}^n \delta_{(\boldsymbol{W}_i,U_i,X_i)}$, which is a random element in $\mathcal{M}(\mathbb{R}^{d+2})$. Further, recall that $\mathcal{R}_\infty$ denotes the population distribution of $(\boldsymbol{W}, U, X)$.

Recall from Subsection 3.2 that $\mathcal{L}_n(c, \boldsymbol{\eta}, \mathcal{R})$ denotes the distribution of the bootstrap test statistic $S_n^*$ when $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n, \hat{\mathcal{R}}_n)$ takes the value $(c, \boldsymbol{\eta}, \mathcal{R})$; that is, $\mathcal{L}_n$ is a measurable mapping from $[\ell, u] \times \mathbb{R}^{d+2} \times \mathcal{M}(\mathbb{R}^{d+2})$ to $\mathcal{M}(\mathbb{R})$, and $\mathcal{L}_n(\hat{c}_n, \hat{\boldsymbol{\eta}}_n, \hat{\mathcal{R}}_n)$ is the bootstrap distribution of $S_n^*$ given the data, which is a random element in $\mathcal{M}(\mathbb{R})$.

Recall the definition of $(\widetilde{\mathbb{C}}, \widetilde{\mathbb{H}}, \widetilde{\mathbb{S}})$ in (A.14). Note that the distribution of $\widetilde{\mathbb{S}}$ depends on the value of $B_1$ and $B_2$ in condition (A.VIII), and thus we denote its law by $\mathcal{L}_\infty(B_2, B_1)$, where $\mathcal{L}_\infty$ may be viewed as a measurable mapping from $[\ell, u] \times \mathbb{R}$ to $\mathcal{M}(\mathbb{R})$.

**Theorem C.1.** *Consider the null, i.e., $\lambda_0 = 0$, and the non-identifiable case, i.e., $\gamma_0 = 0$. Assume that (C.I) holds. Further, consider the standard bootstrap with $m_n = n$. There exist a sequence of random variables $\{(c_n^\dagger, \boldsymbol{\eta}_n^\dagger, \mathcal{R}_n^\dagger) : n \geq 1\}$ and $(\mathbb{C}^\dagger, \mathbb{H}^\dagger)$ such that $(c_n^\dagger, \boldsymbol{\eta}_n^\dagger, \mathcal{R}_n^\dagger)$ has the same distribution as $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n, \hat{\mathcal{R}}_n)$ for each $n \geq 1$, $(\mathbb{C}^\dagger, \mathbb{H}^\dagger)$ as $(\mathbb{C}, \mathbb{H})$, and as $n \to \infty$*

$$d_{\mathsf{Prok}}\left(\mathcal{L}_n(c_n^\dagger, \boldsymbol{\eta}_n^\dagger, \mathcal{R}_n^\dagger),\ \mathcal{L}_\infty(\mathbb{C}^\dagger, \mathbb{H}_\gamma^\dagger)\right) = o_{\mathrm{pr}}(1),$$

*where $\mathbb{H}_\gamma^\dagger$ is the $(d+2)$-th component of $\mathbb{H}^\dagger$.*

*Proof.* By Theorem 2.2, $(\hat{c}_n, \sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0)) \rightsquigarrow (\mathbb{C}, \mathbb{H})$, and the empirical distribution $\hat{\mathcal{R}}_n$ converges weakly (i.e., in terms of $d_{\mathsf{Prok}}$) to the population distribution $\mathcal{R}_\infty$ of the covariates $(\boldsymbol{W}, U, X)$ almost surely [14, Theorem 11.4.1]. Due to Skorohod's representation theorem [8, Theorem 6.7], there exist a sequence of random variables $\{(c_n^\dagger, \boldsymbol{\eta}_n^\dagger, \mathcal{R}_n^\dagger) : n \geq 1\}$ and $(\mathbb{C}^\dagger, \mathbb{H}^\dagger)$ such that $(c_n^\dagger, \boldsymbol{\eta}_n^\dagger, \mathcal{R}_n^\dagger)$ has the same distribution as $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n, \hat{\mathcal{R}}_n)$ for each $n \geq 1$, $(\mathbb{C}^\dagger, \mathbb{H}^\dagger)$ as $(\mathbb{C}, \mathbb{H})$, and for each $\omega \in \Omega$,

$$\lim_{n\to\infty} (c_n^\dagger(\omega), \sqrt{n}(\boldsymbol{\eta}_n^\dagger(\omega) - \boldsymbol{\eta}_0), \mathcal{R}_n^\dagger(\omega)) = (\mathbb{C}^\dagger(\omega), \mathbb{H}^\dagger(\omega), \mathcal{R}_\infty). \tag{C.3}$$

Denote by $\gamma_n^\dagger(\omega)$ the last component of $\boldsymbol{\eta}_n^\dagger(\omega)$; since $\gamma_0 = 0$, we have $\sqrt{n}\gamma_n^\dagger(\omega) \to \mathbb{H}_\gamma^\dagger(\omega)$. Recall the triangle array setup in Appendix A. By Lemma B.2, B.3 and B.7 respectively, conditions (A.I), (A.II) and (A.IX) hold almost surely, when $Q_n$ takes the following form:

$$\begin{aligned}&(\boldsymbol{W}_{n,1}, U_{n,1}, X_{n,1}) \sim \hat{\mathcal{R}}_n,\\ &Y_{n,1}|\boldsymbol{W}_{n,1}, U_{n,1}, X_{n,1} \ \sim\ \exp(Y_{n,1}(\hat{\boldsymbol{\eta}}_n^T \boldsymbol{Z}_{n,1,\hat{c}_n}) - \phi(\hat{\boldsymbol{\eta}}_n^T \boldsymbol{Z}_{n,1,\hat{c}_n})),\end{aligned}$$

with respect to the measure $\nu$. Since $(c_n^\dagger, \boldsymbol{\eta}_n^\dagger, \mathcal{R}_n^\dagger)$ has the same distribution as $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n, \hat{\mathcal{R}}_n)$ for each $n \geq 1$, *by arguing along sub-sequences, without loss* *of generality*, we may assume conditions (A.I), (A.II) and (A.IX) hold almost surely, when $Q_n$ takes the following form:

$$(\boldsymbol{W}_{n,1}, U_{n,1}, X_{n,1}) \sim \mathcal{R}_n^\dagger,$$
$$Y_{n,1} | \boldsymbol{W}_{n,1}, U_{n,1}, X_{n,1} \ \sim \ \exp(Y_{n,1}((\boldsymbol{\eta}_n^\dagger)^T \boldsymbol{Z}_{n,1,c_n^\dagger}) - \phi((\boldsymbol{\eta}_n^\dagger)^T \boldsymbol{Z}_{n,1,c_n^\dagger})).$$

Condition (A.III) only concerns $P$ and is verified in Lemma B.1. Further, again by arguing along sub-sequences, we may assume almost surely, $\|\boldsymbol{\eta}_n^\dagger - \boldsymbol{\eta}_0\| \to 0$, which is condition (A.IV) with $\tau = 1$. Finally, by construction, $\sqrt{n}\gamma_n^\dagger \to \mathbb{H}_\gamma^\dagger$ and $c_n^\dagger \to \mathbb{C}^\dagger$, that is, condition (A.VIII) holds.

With all conditions verified, we apply Theorem A.5 and conclude that for almost surely $\omega \in \Omega$,

$$d_{\mathsf{Prok}}\left(\mathcal{L}_n(c_n^\dagger(\omega), \boldsymbol{\eta}_n^\dagger(\omega), \mathcal{R}_n^\dagger(\omega)),\ \mathcal{L}_\infty(\mathbb{C}^\dagger(\omega), \mathbb{H}_\gamma^\dagger(\omega))\right) \ \to \ 0, \text{ as } n \to \infty.$$

Note that the above statement is true if we argue along sub-sequences, which completes the proof for the in-probability convergence. □

Now we prove Theorem 3.2.

*Proof.* Recall that $\mathbb{S}$ in (2.7) is the limiting distribution of the test statistics $S_n$, and its law is a *fixed* element in $\mathcal{M}(\mathbb{R})$. Further, recall the coupling $\{(c_n^\dagger, \boldsymbol{\eta}_n^\dagger, \mathcal{R}_n^\dagger) : n \geq 1\}$ and $(\mathbb{C}^\dagger, \mathbb{H}^\dagger)$ in Theorem C.1, and in particular $\mathcal{L}_\infty(\mathbb{C}^\dagger, \mathbb{H}_\gamma^\dagger)$ is a random element in $\mathcal{M}(\mathbb{R})$. Thus for some $\epsilon > 0$, we have

$$\mathrm{pr}\left(d_{\mathsf{Prok}}\left(\mathcal{L}_\infty(\mathbb{C}^\dagger, \mathbb{H}_\gamma^\dagger),\ \mathbb{S}\right) \geq 2\epsilon\right) > 0,$$

where the second argument in $d_{\mathsf{Prok}}(\cdot,\cdot)$ refers to the law of $\mathbb{S}$. Then due to Theorem C.1, we have

$$\liminf_{n\to\infty}\ \mathrm{pr}\left(d_{\mathsf{Prok}}\left(\mathcal{L}_n(c_n^\dagger, \boldsymbol{\eta}_n^\dagger, \mathcal{R}_n^\dagger),\ \mathbb{S}\right) \geq \epsilon\right) > 0,$$

which completes the proof since $(c_n^\dagger, \boldsymbol{\eta}_n^\dagger, \mathcal{R}_n^\dagger)$ and $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n, \hat{\mathcal{R}}_n)$ have the same distribution for each $n \geq 1$. □

## Appendix D: Proofs regarding power analysis

In this subsection, we consider the rejection probabilities under the local alternatives $H_{1,n} : \lambda_n = B_0/\sqrt{n}$ defined in (3.3), where recall that the constant $B_0 \neq 0$ does not depend on $n$. Further, recall that the other parameters $\boldsymbol{\eta}_0$ and $c_0$, as well as the distribution of $(\boldsymbol{W}, U, X)$, do not depend on $n$. That is, conditional on $(\boldsymbol{W}, U, X) = (\boldsymbol{w}, u, x)$, the $\nu$-density of $Y = y$ is

$$\exp\left(y\left(\boldsymbol{\eta}_0^T \boldsymbol{z}_{c_0} + \lambda_n u x_{c_0}\right) - \phi(\boldsymbol{\eta}_0^T \boldsymbol{z}_{c_0} + \lambda_n u x_{c_0})\right),$$

where $\boldsymbol{\eta}_0^T \boldsymbol{z}_{c_0} = \boldsymbol{\alpha}_0^T \boldsymbol{w} + \beta_0 u + \gamma_0 x_{c_0}$. Denote by $P_n$ the joint distribution of $(Y, \boldsymbol{W}, U, X)$, and by $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \delta_{\boldsymbol{\mathcal{D}}_i}$ the empirical measure on $S$ induced by $\boldsymbol{\mathcal{D}}_i, i \in [n]$, where $\delta_{\boldsymbol{\mathcal{D}}_i}$ is the Dirac measure at $\boldsymbol{\mathcal{D}}_i$, and define $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - P_n)$.

Next, we consider the identifiable and non-identifiable case separately.

### *D.1. Identifiable case - power analysis*

Recall the definitions of $\boldsymbol{V}^{(1)}_{c,\boldsymbol{\eta}}, \boldsymbol{V}^{(2)}_{c,\boldsymbol{\eta}}, V^{(3)}_{c,\boldsymbol{\eta}}, \sigma^2_{c,\boldsymbol{\eta}}$ in (2.5), and $\boldsymbol{g}^{(1)}_{c,\boldsymbol{\eta}}$ and $g^{(2)}_{c,\boldsymbol{\eta}}$ in (A.11). In the identifiable case, $\gamma_0 \neq 0$.

*Proof of Theorem 3.3.* We start with the first statement regarding the score test statistics $S_n$ based on the original data. The proof is similar to that for Theorem A.3, and we only show key calculations.

By similar arguments as for Theorem A.1, A.2 and 2.1, under the local alternatives $H_{1,n}$ in (3.3), we can show that the maximal likelihood estimators (MLE) $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n)$ have a $\sqrt{n}$-convergence rate in $d_0$-metric, i.e.,

$$\sqrt{n}\left(\sqrt{|\hat{c}_n - c_0|} + \|\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|\right) = O_{\text{pr}}(1).$$

Since $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n)$ is the MLE, we have

$$0 = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \boldsymbol{Z}_{i,\hat{c}_n}(Y_i - \phi'(\boldsymbol{Z}^T_{i,\hat{c}_n}\hat{\boldsymbol{\eta}}_n)) = \mathbb{G}_n \boldsymbol{g}^{(1)}_{\hat{c}_n,\hat{\boldsymbol{\eta}}_n} + \sqrt{n} P_n \boldsymbol{g}^{(1)}_{\hat{c}_n,\hat{\boldsymbol{\eta}}_n}.$$

By a similar argument as for Lemma A.6, $\mathbb{G}_n \boldsymbol{g}^{(1)}_{\hat{c}_n,\hat{\boldsymbol{\eta}}_n} = \mathbb{G}_n \boldsymbol{g}^{(1)}_{c_0,\boldsymbol{\eta}_0} + o_{\text{pr}}(1)$. Further,

$$\begin{aligned}
\sqrt{n} P_n \boldsymbol{g}^{(1)}_{\hat{c}_n,\hat{\boldsymbol{\eta}}_n} &= \sqrt{n} P_n \boldsymbol{g}^{(1)}_{c_0,\hat{\boldsymbol{\eta}}_n} + \sqrt{n} P_n [\boldsymbol{g}^{(1)}_{\hat{c}_n,\hat{\boldsymbol{\eta}}_n} - \boldsymbol{g}^{(1)}_{c_0,\hat{\boldsymbol{\eta}}_n}] \\
&=_{(1)} \sqrt{n} P_n \boldsymbol{g}^{(1)}_{c_0,\hat{\boldsymbol{\eta}}_n} + o_{\text{pr}}(1), \\
&=_{(2)} -\boldsymbol{V}^{(1)}_{c_0,\boldsymbol{\eta}_0} \sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) + B_0 (\boldsymbol{V}^{(2)}_{c_0,\boldsymbol{\eta}_0})^T + o_{\text{pr}}(1),
\end{aligned}$$

where (1) can be verified by similar arguments as for Lemma A.7, and (2) is due to the Taylor Theorem and that $\|\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\|^2 = o_{\text{pr}}(n^{-1/2})$ and $\sqrt{n}\lambda_n = B_0$ under the $H_{1,n}$. From condition (C.I), $\boldsymbol{V}^{(1)}_{c_0,\boldsymbol{\eta}_0}$ is invertible, and thus we have

$$\sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) = (\boldsymbol{V}^{(1)}_{c_0,\boldsymbol{\eta}_0})^{-1} \mathbb{G}_n \boldsymbol{g}^{(1)}_{c_0,\boldsymbol{\eta}_0} + B_0 (\boldsymbol{V}^{(1)}_{c_0,\boldsymbol{\eta}_0})^{-1} (\boldsymbol{V}^{(2)}_{c_0,\boldsymbol{\eta}_0})^T + o_{\text{pr}}(1). \quad \text{(D.1)}$$

By similar arguments as above and for Lemma A.6 and A.7, we have

$$\begin{aligned}
S_n &= \mathbb{G}_n g^{(2)}_{\hat{c}_n,\hat{\boldsymbol{\eta}}_n} + \sqrt{n} P_n g^{(2)}_{\hat{c}_n,\hat{\boldsymbol{\eta}}_n} \\
&= \mathbb{G}_n g^{(2)}_{c_0,\boldsymbol{\eta}_0} + \sqrt{n} P_n g^{(2)}_{c_0,\hat{\boldsymbol{\eta}}_n} + \sqrt{n} P_n [g^{(2)}_{\hat{c}_n,\hat{\boldsymbol{\eta}}_n} - g^{(2)}_{c_0,\hat{\boldsymbol{\eta}}_n}] + o_{\text{pr}}(1) \\
&= \mathbb{G}_n g^{(2)}_{c_0,\boldsymbol{\eta}_0} - \boldsymbol{V}^{(2)}_{c_0,\boldsymbol{\eta}_0} \sqrt{n}(\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0) + B_0 V^{(3)}_{c_0,\boldsymbol{\eta}_0} + o_{\text{pr}}(1) \\
&= \mathbb{G}_n g^{(2)}_{c_0,\boldsymbol{\eta}_0} - \boldsymbol{V}^{(2)}_{c_0,\boldsymbol{\eta}_0} (\boldsymbol{V}^{(1)}_{c_0,\boldsymbol{\eta}_0})^{-1} \mathbb{G}_n \boldsymbol{g}^{(1)}_{c_0,\boldsymbol{\eta}_0} + B_0 \sigma^2_{c_0,\boldsymbol{\eta}_0} + o_{\text{pr}}(1).
\end{aligned}$$

Since $\text{cov}(\mathbb{G}_n(g^{(2)}_{c_0,\boldsymbol{\eta}_0} - \boldsymbol{V}^{(2)}_{c_0,\boldsymbol{\eta}_0}(\boldsymbol{V}^{(1)}_{c_0,\boldsymbol{\eta}_0})^{-1}\boldsymbol{g}^{(1)}_{c_0,\boldsymbol{\eta}_0}))$ converges to $\sigma^2_{c_0,\boldsymbol{\eta}_0}$, by the Lindeberg-Feller central limit theorem, $S_n \rightsquigarrow N(B_0\sigma^2_{c_0,\boldsymbol{\eta}_0}, \sigma^2_{c_0,\boldsymbol{\eta}_0})$ under $H_{1,n}$.

Next, we study the bootstrap distribution of $S^*_n$ under $H_{1,n}$. Consider the triangle array setup in Appendix A with $Q_n$ given in (B.1). By similar arguments as for Theorem 3.1(i), for each sub-sequence, we may extract a further subsequence such that conditions (A.I)-(A.VII) with $\tau = 0$ hold almost surely.

Then by Theorem A.3, we have that the bootstrap distribution of $S_n^*$ converges weakly to $N(0, \sigma^2_{c_0,\boldsymbol{\eta}_0})$ in probability. As a result, the limit of the rejection probabilities under $H_{1,n}$ in (3.3) is

$$\mathbb{P}(|\mathbb{Z}_S + B_0\sigma^2_{c,\boldsymbol{\eta}}| \geq \sigma_{c_0,\boldsymbol{\eta}_0}|\Phi^{-1}(\alpha/2)|),$$

which completes the proof. □

### D.2. Non-identifiable case - power analysis

Recall the definition of $\boldsymbol{V}^{(1)}_{c,\boldsymbol{\eta}}, \boldsymbol{V}^{(2)}_{c,\boldsymbol{\eta}}, V^{(3)}_{c,\boldsymbol{\eta}}$ in (2.5) and $\boldsymbol{V}^{(4)}_c, V^{(5)}_c$ in (3.4). Recall the zero mean Gaussian process $\{((\boldsymbol{\Delta}^{(1)}_c)^T, \Delta^{(2)}_c)^T : c \in [\ell, u]\}$ in Section 2.1, that is tight in $(\ell^\infty([\ell, u]))^{d+3}$ and uniformly $\rho$-continuous, where $\rho(c_1, c_2) = |c_1 - c_2|$ for any $c_1, c_2 \in \mathbb{R}$. Further, recall that $\overline{\mathbb{C}}$ and $\overline{\mathbb{S}}$ are defined in Subsection 3.3, and define

$$\overline{\mathbb{H}} := (\boldsymbol{V}^{(1)}_{\overline{\mathbb{C}},\boldsymbol{\eta}_0})^{-1}(\boldsymbol{\Delta}^{(1)}_{\overline{\mathbb{C}}} + B_0\boldsymbol{V}^{(4)}_{\overline{\mathbb{C}}}).$$

For any $\delta > 0$ let $\tilde{K}_\delta = \{(c, \boldsymbol{h}) \in [\ell, u] \times \mathbb{R}^{d+2} : c \in [\ell, u], \|\boldsymbol{h}\| \leq \delta\}$. For any $(c, \boldsymbol{h}) \in [\ell, u] \times \mathbb{R}^{d+2}$, define the following functions on the observation space $S$:

$$\begin{aligned}
&\overline{\varphi}_{c,\boldsymbol{h},n} := \sqrt{n}(\varphi_{c,\boldsymbol{\eta}_0+\boldsymbol{h}/\sqrt{n}} - \varphi_{c_0,\boldsymbol{\eta}_0}),\\
&\overline{f}^{(2)}_{c,n} = (\mathcal{Y} - \phi'(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_0} + \lambda_n\mathcal{U}\mathcal{X}_{c_0}))\mathcal{U}\mathcal{X}_c,\\
&\overline{\boldsymbol{f}}^{(3)}_c = \phi''(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_0})\mathcal{U}\mathcal{X}_c\boldsymbol{\mathcal{Z}}_c, \quad \overline{f}^{(4)}_c = \phi''(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_0})\mathcal{X}_{c_0}\mathcal{U}\mathcal{X}_c.
\end{aligned}$$

*Proof of Theorem 3.4.* In the non-identifiable case, $\gamma_0 = 0$. We start with the first statement regarding the score test statistics $S_n$ based on the original data. The proof is similar to that for Theorem A.5, and we only show key calculations.

By similar arguments as for Theorem A.1, A.4 and 2.2, under the local alternatives $H_{1,n}$ in (3.3), we can show that the maximal likelihood estimators (MLE) $(\hat{c}_n, \hat{\boldsymbol{\eta}}_n)$ have a $\sqrt{n}$-convergence rate in $d_1$-metric, i.e., $\hat{\boldsymbol{h}}_n := \sqrt{n}\|\hat{\boldsymbol{\eta}}_n - \boldsymbol{\eta}_0\| = O_{\text{pr}}(1)$. Note that by definition,

$$(\hat{c}_n, \hat{\boldsymbol{h}}_n) = \underset{(c,\boldsymbol{h})\in[\ell,u]\times\mathbb{R}^{d+2}}{\text{sargmax}}\ \mathbb{G}_n\overline{\varphi}_{c,\boldsymbol{h},n} + \sqrt{n}P_n\overline{\varphi}_{c,\boldsymbol{h},n}.$$

Further, since $\gamma_0 = 0$, $\lambda_n = B_0/\sqrt{n}$, and $\sqrt{n}\|\boldsymbol{\eta}_n - \boldsymbol{\eta}_0\| = O_{\text{pr}}(1)$, we have

$$\begin{aligned}
S_n &= \mathbb{G}_n[(\mathcal{Y} - \phi'(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_0} + \lambda_n\mathcal{U}\mathcal{X}_{c_0}))\mathcal{U}\mathcal{X}_{\hat{c}_n}]\\
&\quad + \sqrt{n}\mathbb{P}_n[(\phi'(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_0} + \lambda_n\mathcal{U}\mathcal{X}_{c_0}) - \phi'(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{c_0}))\mathcal{U}\mathcal{X}_{\hat{c}_n}]\\
&\quad - \sqrt{n}\mathbb{P}_n[(\phi'((\hat{\boldsymbol{\eta}}_n)^T\boldsymbol{\mathcal{Z}}_{\hat{c}_n}) - \phi'(\boldsymbol{\eta}_0^T\boldsymbol{\mathcal{Z}}_{\hat{c}_n}))\mathcal{U}\mathcal{X}_{\hat{c}_n}]\\
&= \mathbb{G}_n[\overline{f}^{(2)}_{\hat{c}_n,n}] + B_0\mathbb{P}_n[\overline{f}^{(4)}_{\hat{c}_n}] - \mathbb{P}_n[(\overline{\boldsymbol{f}}^{(3)}_{\hat{c}_n})^T]\hat{\boldsymbol{h}}_n + o_{\text{pr}}(1).
\end{aligned}$$

By similar arguments as for Lemma A.8-A.10, for any $\delta > 0$, in $(\ell^\infty(\tilde{K}_\delta))^{d+6}$,

$$\left\{ \begin{bmatrix} \sqrt{n}P_n\overline{\varphi}_{c,\boldsymbol{h},n} \\ \mathbb{P}_n[(\overline{\boldsymbol{f}}_c^{(3)})^T] \\ \mathbb{P}_n[(\overline{f}_c^{(4)})^T] \\ \mathbb{G}_n\overline{\varphi}_{c,\boldsymbol{h},n} \\ \mathbb{G}_n\overline{f}_{c,n}^{(2)} \end{bmatrix} : (c,\boldsymbol{h}) \in \tilde{K}_\delta \right\} \rightsquigarrow \left\{ \begin{bmatrix} -\frac{1}{2}\boldsymbol{h}^T\boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(1)}\boldsymbol{h} + B_0\boldsymbol{h}^T\boldsymbol{V}_c^{(4)} \\ \boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(2)} \\ V_c^{(5)} \\ \boldsymbol{h}^T\boldsymbol{\Delta}_c^{(1)} \\ \Delta_c^{(2)} \end{bmatrix} : (c,\boldsymbol{h}) \in \tilde{K}_\delta \right\}.$$

For each $c \in [\ell, u]$, the maximizer and the maximum value of the function $\mathbb{R}^{d+2} \ni \boldsymbol{h} \mapsto \boldsymbol{h}^T\boldsymbol{\Delta}_c^{(1)} + B_0\boldsymbol{h}^T\boldsymbol{V}_c^{(4)} - \frac{1}{2}\boldsymbol{h}^T\boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(1)}\boldsymbol{h} \in \mathbb{R}$ are respectively:

$$(\boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(1)})^{-1}(\boldsymbol{\Delta}_c^{(1)} + B_0\boldsymbol{V}_c^{(4)}), \quad \text{and}$$
$$\frac{1}{2}\left(\boldsymbol{\Delta}_c^{(1)} + B_0\boldsymbol{V}_c^{(4)}\right)^T (\boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(1)})^{-1}(\boldsymbol{\Delta}_c^{(1)} + B_0\boldsymbol{V}_c^{(4)}).$$

Further, note that $\{((\boldsymbol{\Delta}_c^{(1)})^T, \Delta_c^{(2)})^T : c \in [\ell, u]\}$ is uniformly $\rho$-continuous, and so are $\boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(1)}, \boldsymbol{V}_{c,\boldsymbol{\eta}_0}^{(2)}, V_{c,\boldsymbol{\eta}_0}^{(3)}$ and $\boldsymbol{V}_c^{(4)}, V_c^{(5)}$ due to (C.I), and that $\hat{\boldsymbol{h}}_n = O_{\text{pr}}(1)$. Then by the continuous mapping theorem, we have

$$\left(\hat{c}_n, \hat{\boldsymbol{h}}_n, \mathbb{G}_n\overline{f}_{\hat{c}_n,n}^{(2)}, \mathbb{P}_n[\overline{\boldsymbol{f}}_{\hat{c}_n}^{(3)}]^T, \mathbb{P}_n[\overline{f}_{\hat{c}_n}^{(4)}]\right) \rightsquigarrow (\overline{\mathbb{C}}, \overline{\mathbb{H}}, \Delta_{\overline{\mathbb{C}}}^{(2)}, \boldsymbol{V}_{\overline{\mathbb{C}},\boldsymbol{\eta}_0}^{(2)}, V_{\overline{\mathbb{C}},\boldsymbol{\eta}_0}^{(5)}),$$

and further $S_n$ converges in distribution to $\overline{\mathbb{S}}$.

Next, we study the bootstrap distribution of $S_n^*$ under $H_{1,n}$. Consider the triangle array setup in Appendix A with $Q_n$ given in (B.1). Since $m_n/n \to 0$, by similar arguments as for Theorem 3.1(ii), for each sub-sequence, we may extract a further sub-sequence such that conditions (A.i)-(A.iv) with $\tau = 1$, (A.viii) with $B_1 = 0$ and (A.ix) hold almost surely. Then by Theorem A.5, we have that the bootstrap distribution of $S_n^*$ converges weakly to $\mathbb{S}$ in probability, since $\tilde{\mathbb{S}}$ has the same distribution as $\mathbb{S}$ when $B_1 = 0$. The proof is complete. □

## Acknowledgments

We would like to thank the associated editor and two reviewers for their constructive comments.

## References

[1] Abrevaya, J. and Huang, J. (2005). On the bootstrap of the maximum score estimator. *Econometrica* **73** 1175–1204. MR2149245

[2] Andrews, D. W. K. (2001). Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica* **69** 683–734. MR1828540

[3] Andrews, D. W. and Cheng, X. (2012). Estimation and inference with weak, semi-strong, and strong identification. *Econometrica* **80** 2153–2211. MR3013721

[4] Andrews, D. W. and Guggenberger, P. (2010). Asymptotic size and a problem with subsampling and with the m out of n bootstrap. *Econometric Theory* **26** 426–468. MR2600570

[5] Ballman, K. V. (2015). Biomarker: predictive or prognostic? *J. Clin. Oncol.* **33** 3968–3971.

[6] Bickel, P. J. and Freedman, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217. MR630103

[7] Bickel, P. J., Götze, F. and van Zwet, W. R. (1997). Resampling fewer than $n$ observations: gains, losses, and remedies for losses. *Statist. Sinica* **7** 1–31. MR1441142

[8] Billingsley, P. (1999). *Convergence of probability measures*, second ed. *Wiley Series in Probability and Statistics: Probability and Statistics*. John Wiley & Sons, Inc., New York. MR1700749

[9] Blok, E. J., Engels, C. C., Dekker-Ensink, G., Kranenbarg, E. M.-K., Putter, H., Smit, V. T., Liefers, G.-J., Morden, J. P., Bliss, J. M., Coombes, R. C., Bartlett, J. M., Kroep, J. R., van de Velde, C. J. and Kuppen, P. J. (2018). Exploration of tumour-infiltrating lymphocytes as a predictive biomarker for adjuvant endocrine therapy in early breast cancer. *Breast Cancer Res. Treat.* **171** 65–74.

[10] Bose, A. and Chatterjee, S. (2001). Generalised bootstrap in non-regular $M$-estimation problems. *Statist. Probab. Lett.* **55** 319–328. MR1867535

[11] Chernozhukov, V., Chetverikov, D. and Kato, K. (2014). Gaussian approximation of suprema of empirical processes. *Ann. Statist.* **42** 1564–1597. MR3262461

[12] Davies, R. B. (1977). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **64** 247–254. MR501523

[13] Davies, R. B. (1987). Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika* **74** 33–43. MR885917

[14] Dudley, R. M. (2018). *Real analysis and probability*. CRC Press. MR0982264

[15] Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7** 1–26. MR515681

[16] Einmahl, U. and Mason, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *Ann. Statist.* **33** 1380–1403. MR2195639

[17] Fan, A., Song, R. and Lu, W. (2017). Change-plane analysis for subgroup detection and sample size calculation. *J. Amer. Statist. Assoc.* **112** 769–778. MR3671769

[18] Jespersen, N. (1986). Dichotomizing a Continuous Covariate in the Cox Regression Model Technical Report, Statistical Research Unit, University of Copenhagen.

[19] Jonker, D. J., O'Callaghan, C. J., Karapetis, C. S., Zalcberg, J. R., Tu, D., Au, H.-J., Berry, S. R., Krahn, M., Price, T., Simes, R. J., Tebbutt, N. C., van Hazel, G., Wierzbicki, R.,

Langer, C. and Moore, M. J. (2007). Cetuximab for the treatment of colorectal cancer. *N. Engl. J. Med.* **357** 2040–2048.

[20] Jonker, D. J., Karapetis, C., Harbison, C., O'Callaghan, C. J., Tu, D., Simes, R. J., Malone, D. P., Langer, C., Tebbutt, N., Price, T. J. et al. (2014). Epiregulin gene expression as a biomarker of benefit from cetuximab in the treatment of advanced colorectal cancer. *British journal of cancer* **110** 648–655.

[21] Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. *Springer Series in Statistics*. Springer, New York. MR2724368

[22] Koul, H. L., Qian, L. and Surgailis, D. (2003). Asymptotics of $M$-estimators in two-phase linear regression models. *Stochastic Process. Appl.* **103** 123–154. MR1947962

[23] Lausen, B. and Schumacher, M. (1992). Maximally selected rank statistics. *Biometrics* **48** 73–85.

[24] Lee, S. M. S. and Pun, M. C. (2006). On $m$ out of $n$ bootstrapping for nonstandard M-estimation with nuisance parameters. *J. Amer. Statist. Assoc.* **101** 1185–1197. MR2328306

[25] Li, J. and Jin, B. (2018). Multi-threshold accelerated failure time model. *Ann. Statist.* **46** 2657–2682. MR3851751

[26] Li, J., Tai, B. C. and Nott, D. J. (2009). Confidence interval for the bootstrap P-value and sample size calculation of the bootstrap test. *Journal of Nonparametric Statistics* **21** 649–661. MR2543578

[27] Li, N., Song, Y., Lin, D. and Tu, D. (2021a). Bootstrap Adjustment to Minimum $p$-Value Method for Predictive Classification. *Statistica Sinica, In press.*

[28] Li, J., Li, Y., Jin, B. and Kosorok, M. R. (2021b). Multithreshold change plane model: Estimation theory and applications in subgroup identification. *Statistics in Medicine* **40** 3440–3459. MR4269063

[29] Mallik, A., Sen, B., Banerjee, M. and Michailidis, G. (2011). Threshold estimation based on a $p$-value framework in dose-response and regression settings. *Biometrika* **98** 887–900. MR2860331

[30] Mazumdar, M. and Glassman, J. R. (2000). Categorizing a prognostic variable: review of methods, code for easy implementation and applications to decision-making about cancer treatments. *Statist. Med.* **19** 113–132.

[31] McCullagh, P. and Nelder, J. A. (1989). *Generalized linear models*. *Monographs on Statistics and Applied Probability*. Chapman & Hall, London. MR3223057

[32] Miller, R. and Siegmund, D. (1982). Maximally selected chi square statistics. *Biometrics* **38** 1011–1016. MR695368

[33] Mukherjee, D., Banerjee, M. and Ritov, Y. (2020). Asymptotic normality of a linear threshold estimator in fixed dimension with near-optimal rate. *arXiv* 2001.06955.

[34] Schilsky, R. L. (2010). Personalized medicine in oncology: the future is now. *Nat. Rev. Drug Discov.* **9** 363–366.

[35] Schwartz, L. H., Litière, S., De Vries, E., Ford, R., Gwyther, S.,

Mandrekar, S., Shankar, L., Bogaerts, J., Chen, A., Dancey, J. et al. (2016). RECIST 1.1—Update and clarification: From the RECIST committee. *Eur. J. Cancer* **62** 132–137.

[36] Seijo, E. and Sen, B. (2011). Change-point in stochastic design regression and the bootstrap. *Ann. Statist.* **39** 1580–1607. MR2850213

[37] Sen, B., Banerjee, M. and Woodroofe, M. (2010). Inconsistency of bootstrap: the Grenander estimator. *Ann. Statist.* **38** 1953–1977. MR2676880

[38] Shao, J. and Tu, D. S. (1995). *The jackknife and bootstrap*. *Springer Series in Statistics*. Springer-Verlag, New York. MR1351010

[39] Therasse, P., Arbuck, S. G., Eisenhauer, E. A., Wanders, J., Kaplan, R. S., Rubinstein, L., Verweij, J., Van Glabbeke, M., van Oosterom, A. T., Christian, M. C. et al. (2000). New guidelines to evaluate the response to treatment in solid tumors. *J. Natl. Cancer Inst.* **92** 205–216.

[40] Tunes-da Silva, G. and Klein, J. P. (2011). Cutpoint selection for discretizing a continuous covariate for generalized estimating equations. *Comput. Statist. Data Anal.* **55** 226–235. MR2736550

[41] van der Vaart, A. W. (1998). *Asymptotic statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge. MR1652247

[42] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes: with applications to statistics*. *Springer Series in Statistics*. Springer-Verlag, New York. MR1385671

[43] Xu, G., Sen, B. and Ying, Z. (2014). Bootstrapping a change-point Cox model for survival data. *Electron. J. Stat.* **8** 1345–1379. MR3263125

[44] Yu, P. (2014). The bootstrap in threshold regression. *Econometric Theory* **30** 676–714. MR3205610

[45] Ziegler, A., Koch, A., Krockenberger, K. and Grosshennig, A. (2012). Personalized medicine using DNA biomarkers: a review. *Hum. genet.* **131** 1627–1638.