UNIVERSITY OF CAMBRIDGE
Faculty of Economics

Institute for New Economic Thinking

CAMBRIDGE WORKING PAPERS IN ECONOMICS
CAMBRIDGE-INET WORKING PAPERS

A Unified Framework for Specification Tests of Continuous Treatment Effect Models

Wei Huang	Oliver Linton	Zheng Zhang
University of Melbourn	University of Cambridge	Renmin University of China

Abstract

We propose a general framework for the specification testing of continuous treatment effect models. We assume a general residual function, which includes the average and quantile treatment effect models as special cases. The null models are identified under the confoundedness condition and contain a nonparametric weighting function. We propose a test statistic for the null model in which the weighting function is estimated by solving an expanding set of moment equations. We establish the asymptotic distributions of our test statistic under the null hypothesis and under fixed and local alternatives. The proposed test statistic is shown to be more efficient than that constructed from the true weighting function and can detect local alternatives deviated from the null models at the rate of $O_P(N^{-1/2})$. A simulation method is provided to approximate the null distribution of the test statistic. Monte-Carlo simulations show that our test exhibits a satisfactory finite-sample performance, and an application shows its practical value.

Reference Details

2113	Cambridge Working Papers in Economics
2021/07	Cambridge-INET Working Paper Series
Published	16 February 2021
Key Words	Consistent tests, Continuous treatment effect, Series estimation, Bootstrap
JEL Codes	C10, C11, C12
Websites	www.econ.cam.ac.uk/cwpe www.inet.econ.cam.ac.uk/working-papers

A Unified Framework for Specification Tests of Continuous Treatment Effect Models

Wei Huang [*,⋆], Oliver Linton [†,†], and Zheng Zhang [‡,‡]

[⋆] *School of Mathematics and Statistics and Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS), University of Melbourne, Australia*
[†] *Faculty of Economics, University of Cambridge*
[‡] *Center for Applied Statistics, Institute of Statistics & Big Data, Renmin University of China*

Abstract

We propose a general framework for the specification testing of continuous treatment effect models. We assume a general residual function, which includes the average and quantile treatment effect models as special cases. The null models are identified under the confoundedness condition and contain a nonparametric weighting function. We propose a test statistic for the null model in which the weighting function is estimated by solving an expanding set of moment equations. We establish the asymptotic distributions of our test statistic under the null hypothesis and under fixed and local alternatives. The proposed test statistic is shown to be more efficient than that constructed from the true weighting function and can detect local alternatives deviated from the null models at the rate of $O_P(N^{-1/2})$. A simulation method is provided to approximate the null distribution of the test statistic. Monte-Carlo simulations show that our test exhibits a satisfactory finite-sample performance, and an application shows its practical value.

Keywords: Consistent tests; Continuous treatment effect; Series estimation; Bootstrap.

1 Introduction

Causal inference is a central topic in economics, statistics, and machine learning. Although a randomized trial is the gold standard for identifying the causal effect, it is often unavail-

[*] E-mail: wei.huang@unimelb.edu.au
[†] E-mail: obl20@cam.ac.uk
[‡] E-mail: zhengzhang@ruc.edu.cn

able or even unethical in practice. Observational data, where the participation of an intervention is only observed rather than manipulated by scientists, are predominantly what is available. A major challenge for inferring the causality in observational study is the confoundedness, whereby the individual characteristics are correlated with both the treatment variable and the outcome of interest. To identify causality, the *unconfounded treatment assignment* condition is frequently imposed in the literature, see Rosenbaum and Rubin (1983, 1984). For a comprehensive review of causal inference and its applications, see Imbens and Wooldridge (2009) and Abadie and Cattaneo (2018).

Treatment effect models are used extensively in economics and statistics to evaluate the causal effect of a treatment or policy. Most of the existing literature focuses on the binary treatment where an individual either receives the treatment or does not (see e.g., Hahn, 1998, Hirano, Imbens, and Ridder, 2003, Donald, Hsu, and Lieli, 2014, Abrevaya, Hsu, and Lieli, 2015, Chan, Yam, and Zhang, 2016, Athey, Imbens, and Wager, 2018, Hsu, Lai, and Lieli, 2020, Chen, Hsu, and Wang, 2020, Fan, Hsu, Lieli, and Zhang, 2020 among others). Some literature focus on the multivalued treatment (see e.g., Cattaneo, 2010, Lee, 2018, and Ao, Calonico, and Lee, 2021). In many applications, however, the treatment variable is continuously valued, and its causal effect is of great interest to decision makers. For example, in evaluating how non-labor income affects the labor supply, the causal effect may depend on not only the introduction of the non-labor income but also the total non-labor income. Similarly, in evaluating how advertising affects the campaign contributions for political analysis, the causal effect may depend not only on whether any advertisements are imposed but also on how many of them are distributed.

Estimation of the continuous treatment effects has drawn great attention from researchers (see Hirano, Imbens, and Ridder (2003), Galvao and Wang (2015), Kennedy, Ma, McHugh, and Small (2017), Fong, Hazlett, and Imai (2018), Dong, Lee, and Gou (2019), Huber, Hsu, Lee, and Lettry (2020), Colangelo and Lee (2020) among others). Hirano, Imbens, and Ridder (2003), Galvao and Wang (2015), and Fong, Hazlett, and Imai (2018) applied fully parametric methods by modelling either the conditional distribution of the treatment given the confounders or that of the observed outcome given the treatment and the confounders. The shortcoming of these parametric methods is that modelling and testing the relations of the treatment and the observed outcome regarding the confounders are difficult, especially when multiple confounding variables are involved. If the model is mis-specified, the conclusion can be biased and completely misleading. Kennedy, Ma, McHugh, and Small (2017) and Huber, Hsu, Lee, and Lettry (2020) estimated the continuous treatment effects by using the nonparametric kernel method. Although nonparametric approaches are much more flexible than parametric ones, they require smoothing of the data rather than estimating finite dimensional parameters, which leads to less precise fits and slower convergence rates (slower than $N^{-1/2}$). Furthermore, it is usually hard to interpret nonparametric results.

In a recent article, Ai, Linton, Motegi, and Zhang (2021) studied the continuous treatment effects by imposing a *univariate generalized parametric* model for the functionals of the potential outcome over the treatment variable. The general framework includes many important causal parameters as special cases, for example, the average and quantile treatment effects. They proposed a generalized weighting estimator for the causal effect with the weights modelled nonparametrically and estimated by solving an expanding set of equations. They further derived the semiparametric efficiency bound for the causal effect of treatment under the unconfounded treatment assignment condition and showed that their estimator is $\sqrt{N}$-asymptotically normal and attains the semiparametric efficiency bound. Although Ai, Linton, Motegi, and Zhang (2021)'s estimator enjoys superior asymptotic properties and satisfactory finite sample performance, they did not detail the specifications of the parametric models for the functionals of the potential outcomes. If the parametric model is mis-specified, the results developed in Ai, Linton, Motegi, and Zhang (2021) do not hold.

We study the question of model specification. In particular, we propose a consistent specification test for the most generalized continuous treatment effect model. That is, we consider the generalized parametric model in Ai, Linton, Motegi, and Zhang (2021) as the null model in our hypothesis test. The potential outcome variable in the model is not observable. However, under the unconfounded treatment assignment condition, the model can be identified by a semiparametric weighted conditional model. There is abundant literature that studies the specification tests for conditional models (see e.g. Ait-Sahalia, Bickel, and Stoker (2001), Bierens (1982, 1990), Fan and Li (1996), Zheng (1996), Bierens and Ploberger (1997), Stute (1997), Li (1999), Chen and Fan (1999), Fan and Li (2000), Li, Hsiao, and Zinn (2003), Crump, Hotz, Imbens, and Mitnik (2008) among others). Most authors have considered the problems of testing a parametric/semiparametric null model using the integrated type test statistic.Ait-Sahalia, Bickel, and Stoker (2001) and Chen and Fan (1999) considered testing the nonparametric/semiparametric null models using nonparametric kernel methods. Li, Hsiao, and Zinn (2003) considered testing the nonparametric/semiparametric using series methods. Crump, Hotz, Imbens, and Mitnik (2008) derived a nonparametric Wald test statistic for testing the conditional average treatment effects under the unconfoundedness condition. We estimate our semiparametric weighted null model by using the approach developed in Ai, Linton, Motegi, and Zhang (2021) and construct an integrated-type test statistic. Although the weights in our null model are estimated nonparametrically, we show that our proposed test statistic is more efficient than that constructed from the true weights. Moreover, our proposed test statistic can detect local alternatives that deviate from the null model at the rate of $O_P(N^{-1/2})$.

Under the null hypothesis our test statistic is shown to converge in distribution to a weighted sum of independent chi-squared random variables. It is known that obtaining the exact critical values of such a distribution is extremely difficult in practice. Most of the literature suggests using a residual wild bootstrap procedure to approximate the critical values. This is not applicable in our case because our null model does not imply any explicit form of relationship among the observed outcome, the treatment, and the confounders for residual sampling. To resolve this problem, we propose a simulation method to approximate the null limiting distribution. Monte-Carlo simulations and real data analysis were conducted to demonstrate the numerical properties of our test method and limiting distribution approximation.

The remainder of the paper is organized as follows. We introduce the problem formulation and notations in Section 2. Section 3 constructs the test statistic, followed by the study of the asymptotic properties under null hypothesis, the fixed and the local alternatives in Section 4. In Section 5, we discuss how to approximate the limiting distribution under the null hypothesis. Finally, Section 6 discusses the choice of the tuning parameters in the estimation and investigates the finite sample performance through simulations and U.S. campaign advertisement data.

2 Basic framework

Let T denote a continuous treatment variable with support $\mathcal{T} \subset \mathbb{R}$, where $\mathcal{T}$ is a continuum subset, and T has a marginal density function $f_T(t)$. Let $Y^*(t)$ denote the potential response when treatment $T = t$ is assigned. We are interested in testing the null hypothesis:

$$H_0 : \exists \text{ some } \boldsymbol{\theta}^* \in \Theta, \text{ s.t. } \mathbb{E}[m\{Y^*(t); g(t; \boldsymbol{\theta}^*)\}] = 0 \text{ for all } t \in \mathcal{T}, \tag{2.1}$$

against the alternative hypothesis

$$H_1 : \nexists \text{ any } \boldsymbol{\theta} \in \Theta, \text{ s.t. } \mathbb{E}[m\{Y^*(t); g(t; \boldsymbol{\theta})\}] = 0 \text{ for all } t \in \mathcal{T},$$

where Θ is a compact set in $\mathbb{R}^p$ for some integer $p \geq 1$, $m(\cdot)$ is some generalized residual function which could possibly be *non-differentiable*, and $g(t; \boldsymbol{\theta})$ is a parametric working model which is differentiable with respect to $\boldsymbol{\theta}$. If H_0 holds, for each t, the dose-response function (DRF) is defined as the value $g(t; \boldsymbol{\theta}^*)$ that solves the moment condition in (2.1). The following examples show that the average dose-response function (ADRF) and the quantile dose-response function (QDRF) are special cases of $g(t;\boldsymbol{\theta}^*)$, which result from choosing specific forms of $m(\cdot)$.

- (Average) Setting $m\{Y^*(t); g(t;\boldsymbol{\theta}^*)\} = Y^*(t) - g(t;\boldsymbol{\theta}^*)$ and letting its first moment equal zero for each t, we obtain $g(t;\boldsymbol{\theta}^*) = \mathbb{E}\{Y^*(t)\}$, the unconditional ADRF, which is also called a *marginal structural model* in Robins, Hernán, and Brumback (2000). This can recover the average treatment effect (ATE), which is given by $\text{ATE}(t_1, t_0) = \mathbb{E}\{Y^*(t_1)\} - \mathbb{E}\{Y^*(t_0)\}$. Examples include the linear marginal structure model $\mathbb{E}\{Y^*(t)\} = \beta_0 + \beta_1 \cdot t$, and the nonlinear marginal structure model $\mathbb{E}\{Y^*(t)\} = \beta_0 \cdot t + 1/(t + \beta_1)^2$ studied in Hirano and Imbens (2004)).

- (Quantile) Let $\tau \in (0,1)$ and $F_{Y^*(t)}(\cdot)$ be the cumulative distribution function of $Y^*(t)$. Setting $m\{Y^*(t); g(t;\boldsymbol{\theta}^*)\} = \tau - \mathbb{1}\{Y^*(t) < g(t;\boldsymbol{\theta}^*)\}$ and letting its first moment equal zero for each t, we obtain $g(t;\boldsymbol{\theta}^*) = F^{-1}_{Y^*(t)}(\tau) := \inf\{q : \mathbb{P}(Y^*(t) \geq q) \leq \tau\}$, the unconditional QDRF. This can recover the quantile treatment effect (QTE), which is given by $\text{QTE}(t_1, t_0) = F^{-1}_{Y^*(t_1)}(\tau) - F^{-1}_{Y^*(t_0)}(\tau)$. See Firpo (2007) for detailed discussion on QTE. Examples include the linear model $g(t;\boldsymbol{\theta}) = \theta_0 + \theta_1 \cdot t$ and the Box-Cox transformation model $g(t;\boldsymbol{\theta}) = h_\lambda(\theta_0 + \theta_1 \cdot t)$ studied in Buchinsky (1995), where $h_\lambda(z) = \lambda z + 1)^{-1/\lambda}$.

We consider the observational study where the potential outcome $Y^*(t)$ is not observed for all t. Let $Y := Y^*(T)$ denote the observed response. Under the null hypothesis, one may attempt to solve the following equation to find $\boldsymbol{\theta}^*$:

$$\mathbb{E}[m\{Y; g(T;\boldsymbol{\theta})\nabla_\theta g(T;\boldsymbol{\theta}\}] = 0\,.$$

However, if there is a selection into treatment, even under the null hypothesis, the true value $\boldsymbol{\theta}^*$ does not solve the above equation. Indeed, in this case, the observed response and the treatment assignment data alone cannot identify $\boldsymbol{\theta}^*$. To address this identification issue, most studies in the literature impose a selection on the observable condition (e.g., Hirano, Imbens, and Ridder, 2003, Imai and van Dyk, 2004, Fong, Hazlett, and Imai, 2018, Ai, Linton, Motegi, and Zhang, 2021). Specifically, let $\boldsymbol{X} \in \mathbb{R}^r$, for some integer $r \geq 1$, denote a vector of observable covariates. The following condition shall be maintained throughout the paper.

Assumption 1 (*Unconfounded Treatment Assignment*)**.** *For all $t \in \mathcal{T}$, given $\boldsymbol{X}$, T is independent of $Y^*(t)$, that is, $Y^*(t) \perp T|\boldsymbol{X}$, for all $t \in \mathcal{T}$.*

Let $\{T_i, \boldsymbol{X}_i, Y_i\}_{i=1}^N$ be an independent and identically distributed ($i.i.d.$) sample drawn from the joint distribution of $(T, \boldsymbol{X}, Y)$. Let $f_{T|X}$ denote the conditional density of T given the observed covariates $\boldsymbol{X}$. Under Assumption 1, Ai, Linton, Motegi, and Zhang (2021) showed that $\mathbb{E}[m\{Y^*(t); g(t;\boldsymbol{\theta})\}]$ can be identified as follows:

$$\mathbb{E}[m\{Y^*(t); g(t;\boldsymbol{\theta})\}] = \mathbb{E}[\pi_0(T,\boldsymbol{X})m\{Y; g(T;\boldsymbol{\theta})\}|T=t],\ \forall t \in \mathcal{T}$$

where

$$\pi_0(T,\boldsymbol{X}) := \frac{f_T(T)}{f_{T|X}(T|\ \boldsymbol{X})}\,.$$

The function $\pi_0(T,\boldsymbol{X})$ is called the *stabilized weights* in Robins, Hernán, and Brumback (2000). Then under H_0, the true value $\boldsymbol{\theta}^*$ solves the following equation:

$$\mathbb{E}\left[\pi_0(T,\boldsymbol{X})m\{Y; g(T;\boldsymbol{\theta})\}\nabla_\theta g(T;\boldsymbol{\theta})\right] = 0\,, \tag{2.2}$$

where the "$\nabla_{\boldsymbol{\theta}}$" denotes the derivative with respect to $\boldsymbol{\theta}$.

The null and alternative hypothesis in (2.1) can then be re-written as

$$H_0 : \mathbb{P}\left(\mathbb{E}[\pi_0(T,\boldsymbol{X})m\{Y; g(T;\boldsymbol{\theta}^*)\}|T] = 0\right) = 1 \text{ for some } \boldsymbol{\theta}^* \in \Theta\,, \tag{2.3}$$

against the alternative hypothesis

$$H_1 : \mathbb{P}\left(\mathbb{E}[\pi_0(T,\boldsymbol{X})m\{Y; g(T;\boldsymbol{\theta})\}|T] \neq 0\right) > 0 \text{ for all } \boldsymbol{\theta} \in \Theta\,.$$

This converts the test for (2.1) to a goodness-of-fit test for a univariate regression model, if both $\pi_0(T,\boldsymbol{X})$ and $\boldsymbol{\theta}^*$ were given. Specially, letting

$$U_i := \pi_0(T_i,\boldsymbol{X}_i)m\{Y_i; g(T_i;\boldsymbol{\theta}^*)\}\,, \tag{2.4}$$

the null hypothesis H_0 is equivalent to $\mathbb{P}\{\mathbb{E}(U_i|T_i) = 0\} = 1$. A popular technique for testing such a conditional moment model is to convert it to an unconditional one.

Note that $\mathbb{P}\{\mathbb{E}(U_i|T_i) = 0\} = 1$ if and only if $\mathbb{E}\{U_i M(T_i)\} = 0$ for all bounded and measurable functions $M(\cdot)$. Following Bierens and Ploberger (1997), Stinchcombe and White (1998), Stute (1997), and Li, Hsiao, and Zinn (2003), by choosing a proper weight function $\mathscr{H}(\cdot,\cdot)$, $\mathbb{E}(U_i|T_i) = 0$ is a.s. equivalent to

$$\mathbb{E}\left\{U_i \mathscr{H}(T_i,t)\right\} = 0 \text{ for all } t \in \mathcal{T}\,. \tag{2.5}$$

Popular choices of such a weight function are the logistic function $\mathscr{H}(T_i,t) = 1/\{1 + \exp(c - t\cdot T_i)\}$ with $c \neq 0$, cosine-sine function $\mathscr{H}(T_i,t) = \cos(t\cdot T_i) + \sin(t\cdot T_i)$ and the indicator function $\mathscr{H}(T_i,t) = \mathbb{1}(T_i \le t)$ (see Stinchcombe and White, 1998 and Stute, 1997 for more detailed discussion). Now, letting

$$J_N^0(t) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_i \mathscr{H}(T_i,t)\,, \tag{2.6}$$

the sample analogue of $\mathbb{E}\left\{U_i \mathscr{H}(T_i, t)\right\}$ multiplied by $\sqrt{N}$, one can test H_0 by using the Cramer-von Mises (CM)-type statistic

$$CM_N^0 = \int \{J_N^0(t)\}^2 \widehat{F}_T(dt) = \frac{1}{N}\sum_{i=1}^{N}\{J_N^0(T_i)\}^2, \tag{2.7}$$

where $\widehat{F}_T(\cdot)$ is the empirical distribution of $T_1, ..., T_N$. However, both $\pi_0(T, \boldsymbol{X})$ and $\boldsymbol{\theta}^*$ are unknown in practice so that the U_i's are unavailable. We have to replace the U_i's with some estimates, which is studied in the following section.

3 Test statistic

One obvious approach for estimating the U_i's is to estimate $f_T(T_i)$ and $f_{T|X}(T_i|\boldsymbol{X}_i)$, then construct the estimators of $\pi_0(T_i, \boldsymbol{X}_i)$ and $\boldsymbol{\theta}^*$. However, it is well-known that this ratio estimator of $\pi_0(T, \boldsymbol{X})$ is very sensitive to small values of $f_{T|X}(T|\boldsymbol{X})$ because small estimation errors in estimating $f_{T|X}(T|\boldsymbol{X})$ result in large estimation errors of the estimator of $\pi_0(T, \boldsymbol{X})$. To avoid or mitigate this problem, we follow Ai, Linton, Motegi, and Zhang (2021)'s idea of estimating the weighting function $\pi_0(T, \boldsymbol{X})$ by generalized empirical likelihood. Note that the weighting function satisfies

$$\mathbb{E}\left\{\pi_0(T, \boldsymbol{X}) u(T) v(\boldsymbol{X})\right\} = \mathbb{E}\{u(T)\} \cdot \mathbb{E}\{v(\boldsymbol{X})\} \tag{3.1}$$

for any suitable functions $u(t)$ and $v(\boldsymbol{x})$. Ai, Linton, Motegi, and Zhang (2021, Theorem 2) showed that the restriction (3.1) identifies the weighting function $\pi_0(T, \boldsymbol{X})$. This result suggests that one may estimate the $\pi_0(T_i, \boldsymbol{X}_i)$'s by solving the sample analogue of (3.1). The challenge is that (3.1) implies an infinite number of equations, which is impossible to solve with a finite sample of observations. To overcome this difficulty, Ai, Linton, Motegi, and Zhang (2021) suggested approximating the infinite-dimensional function space by a sequence of finite-dimensional sieve spaces. Specifically, let $u_{K_1}(T) = (u_{K_1,1}(T), \ldots, u_{K_1,K_1}(T))^\top$ and $v_{K_2}(\boldsymbol{X}) = \left(v_{K_2,1}(\boldsymbol{X}), \ldots, v_{K_2,K_2}(\boldsymbol{X})\right)^\top$ denote some known basis functions with dimensions $K_1 \in \mathbb{N}$ and $K_2 \in \mathbb{N}$ respectively, and let $K := K_1 \cdot K_2$. The functions $u_{K_1}(t)$ and $v_{K_2}(\boldsymbol{x})$ are called the *approximation sieves*, such as B-splines or power series (see Newey, 1997, Chen, 2007, for more discussion on sieve approximation). Because the sieve approximating space is a subspace of the original function space, $\pi_0(T, \boldsymbol{X})$ also satisfies

$$\mathbb{E}\left\{\pi_0(T, \boldsymbol{X}) u_{K_1}(T) v_{K_2}(\boldsymbol{X})^\top\right\} = \mathbb{E}\{u_{K_1}(T)\} \cdot \mathbb{E}\{v_{K_2}(\boldsymbol{X})\}^\top . \tag{3.2}$$

Following Ai, Linton, Motegi, and Zhang (2021), we estimate the $\pi_0(T_i, \boldsymbol{X}_i)$'s consistently by the $\widehat{\pi}_i$'s that maximize the generalized empirical likelihood (GEL) function, subject to the sample analog of (3.2):

$$\begin{cases} \{\widehat{\pi}_i\}_{i=1}^N = \arg\max\left(-N^{-1}\sum_{i=1}^N \pi_i \log \pi_i\right) \\ \text{subject to } \frac{1}{N}\sum_{i=1}^N \pi_i u_{K_1}(T_i) v_{K_2}(\boldsymbol{X}_i)^\top = \left\{\frac{1}{N}\sum_{i=1}^N u_{K_1}(T_i)\right\}\left\{\frac{1}{N}\sum_{j=1}^N v_{K_2}(\boldsymbol{X}_j)^\top\right\}. \end{cases} \tag{3.3}$$

Two observations are immediate. First, by including a constant of one in the sieve base functions, (3.3) guarantees that $N^{-1}\sum_{i=1}^N \widehat{\pi}_i = 1$. Second, we notice that

$$\max\left(-N^{-1}\sum_{i=1}^N \pi_i \log \pi_i\right) = -\min\left\{\sum_{i=1}^N (N^{-1}\pi_i)\cdot \log\left(\frac{N^{-1}\pi_i}{N^{-1}}\right)\right\}.$$

The entropy maximization problem minimizes the Kullback-Leibler divergence between the weights $\{N^{-1}\pi_i\}_{i=1}^N$ and the empirical frequencies $\{N^{-1}\}$, subject to the sample analogue of (3.2). Further, Ai, Linton, Motegi, and Zhang (2021) showed that the dual solution of the primal problem (3.3) is

$$\widehat{\pi}_K(T_i, \boldsymbol{X}_i) := \rho'\left\{u_{K_1}(T_i)^\top \widehat{\Lambda}_{K_1\times K_2} v_{K_2}(\boldsymbol{X}_i)\right\}, \tag{3.4}$$

where ρ' is the first derivative of ρ with $\rho(u) = -\exp(-u-1)$, and $\widehat{\Lambda}_{K_1\times K_2}$ is the maximizer of the strictly concave function $\widehat{G}_{K_1\times K_2}$ defined by

$$\begin{aligned} &\widehat{G}_{K_1\times K_2}(\Lambda) \\ &:= \frac{1}{N}\sum_{i=1}^N \rho\left\{u_{K_1}(T_i)^\top \Lambda v_{K_2}(\boldsymbol{X}_i)\right\} - \left\{\frac{1}{N}\sum_{i=1}^N u_{K_1}(T_i)\right\}^\top \Lambda \left\{\frac{1}{N}\sum_{j=1}^N v_{K_2}(\boldsymbol{X}_j)\right\}. \end{aligned} \tag{3.5}$$

The first order condition of (3.5) implies that $\{\widehat{\pi}_K(T_i, \boldsymbol{X}_i)\}_{i=1}^N$ satisfy the sample analog of (3.2), such restrictions reduce the chance of obtaining extreme weights. The concavity of (3.5) enables us to obtain the solution quickly via the Gauss-Newton algorithm. To ensure a consistent estimate of $\pi_0(T, \boldsymbol{X})$, the dimensions of the bases, K_1 and K_2, shall increase as the sample size increases. The choice of K_1 and K_2 in practice will be discussed in Section 6.1.

Having estimated the weights, we now estimate $\boldsymbol{\theta}^*$, denoted by $\widehat{\boldsymbol{\theta}}$, by solving the sample analogue of (2.2) with respect to $\boldsymbol{\theta}$, i.e.

$$\frac{1}{N}\sum_{i=1}^N \widehat{\pi}_K(T_i, \boldsymbol{X}_i) m\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\} \nabla_\theta g(T; \widehat{\boldsymbol{\theta}}) = o_P(N^{-1/2}). \tag{3.6}$$

With the estimators $\{\widehat{\pi}_K(T_i, \boldsymbol{X}_i)\}_{i=1}^N$ of $\{\pi_0(T_i, \boldsymbol{X}_i)\}_{i=1}^N$ and $\widehat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, we estimate U_i by $\widehat{U}_i = \widehat{\pi}_K(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\}$, for $i = 1, \ldots, N$. Replacing the U_i's in (2.6) by the $\widehat{U}_i$'s, we have the feasible test statistic for H_0 based on

$$\widehat{J}_N(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \widehat{U}_i \mathscr{H}(T_i, t) \, ,$$

and the corresponding estimator of the Cramer-von Mises (CM)-type statistic in (2.7) is

$$\widehat{CM}_N = \frac{1}{N} \sum_{i=1}^{N} \{\widehat{J}_N(T_i)\}^2 \, .$$

Remark 1. *Crump, Hotz, Imbens, and Mitnik (2008) considered the null hypothesis concerning the conditional average treatment effect (CATE) with a binary treatment, that is,* $H_0 : CATE(\boldsymbol{x}) := \mathbb{E}[Y^*(1) - Y^*(0)|\boldsymbol{X} = \boldsymbol{x}] = 0$ *for all* $\boldsymbol{x}$*. This null hypothesis indicates that there is no heterogeneity in average treatment effects by covariates. Under the unconfoundedness condition, the null hypothesis is identical to* $H_0 : CATE(\boldsymbol{x}) = \mathbb{E}[Y|T = 1, \boldsymbol{X} = \boldsymbol{x}] - \mathbb{E}[Y|T = 0, \boldsymbol{X} = \boldsymbol{x}] = 0$ *for all* $\boldsymbol{x}$*. Further, they proposed series estimators for the regression functions and formed the Wald test statistic. Their test method is applicable to a particular scenario included in our general formulation (2.1) that there is no continuous average treatment effect, that is,* $H_0' : \mathbb{E}[Y^*(t)] = \mathbb{E}[\pi_0(T, \boldsymbol{X})Y|T = t] = 0$ *for all* t*, given that* $\pi_0(T, \boldsymbol{X})$ *was known. However, in practice,* $\pi_0(T, \boldsymbol{X})$ *is usually unknown and needs to be estimated.*

Remark 2. *An alternative estimator of* $\boldsymbol{\theta}^*$ *can be constructed under* H_0*. Suppose that under* H_0*,* $\boldsymbol{\theta}^*$ *is identified by the unique solution of the following optimization problem:*

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta} \in \Theta} CM(\boldsymbol{\theta}) := N \times \int_{\mathcal{T}} \{\mathbb{E}\left[U_i(\boldsymbol{\theta})\mathcal{H}(T_i, t)\right]\}^2 f_T(t)dt,$$

where $U_i(\boldsymbol{\theta}) := \pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}$*. Let* $\widehat{U}_i(\boldsymbol{\theta}) := \widehat{\pi}_K(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}$ *and* $\widehat{J}_N(t; \boldsymbol{\theta}) := N^{-1/2} \sum_{i=1}^N \widehat{U}_i(\boldsymbol{\theta})\mathscr{H}(T_i, t)$*. Under* H_0*, the estimator of* $\boldsymbol{\theta}^*$ *can be defined by*

$$\widehat{\boldsymbol{\theta}}_{opt} := \arg\min_{\boldsymbol{\theta} \in \Theta} \widehat{CM}_N(\boldsymbol{\theta}) := \arg\min_{\boldsymbol{\theta} \in \Theta} \frac{1}{N} \sum_{i=1}^{N} \{\widehat{J}_N(T_i; \boldsymbol{\theta})\}^2. \tag{3.7}$$

As a result, the alternative test statistic is $\widehat{CM}_N(\widehat{\boldsymbol{\theta}}_{opt})$*. However, seeking the global minimizer of* $\widehat{CM}_N(\boldsymbol{\theta})$ *is difficult as* $\widehat{CM}_N(\boldsymbol{\theta})$ *may not be differentiable, convex, even may not be continuous. For example, taking* $m\{Y_i; g(T_i; \boldsymbol{\theta})\} = \tau - \mathbb{1}\{Y_i \leq g(T_i; \boldsymbol{\theta})\}$ *for QDRF, there does not exist a unique solution to the problem (3.7). Under a stronger condition that* $m(y; g)$ *is differentiable in* g*, we establish the asymptotic results for both* $\widehat{J}_N(t; \widehat{\boldsymbol{\theta}}_{opt})$ *and* $\widehat{CM}_N(\widehat{\boldsymbol{\theta}}_{opt})$ *in Appendix E.*

4 Large sample properties

This section studies the asymptotic properties of $\widehat{J}_N(\cdot)$ and the test statistic $\widehat{CM}_N$.

4.1 Asymptotic properties under null hypothesis

To establish the asymptotic properties of $\widehat{J}_N(\cdot)$ and $\widehat{CM}_N$, the following additional assumptions are imposed.

Assumption 2. *Suppose that* $N^{-1}\sum_{i=1}^{N}\widehat{\pi}_K(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\}\nabla_{\boldsymbol{\theta}}g(T_i;\widehat{\boldsymbol{\theta}})=o_p(N^{-1/2})$ *holds.*

Assumption 3. $Var(Y|T)$ *is bounded a.s. on the support of* T.

Assumption 4.

(i) $g(t;\boldsymbol{\beta})$ *is twice continuously differentiable in* $\boldsymbol{\theta}\in\Theta$*;*

(ii) $\mathbb{E}[m\{Y;g(T;\boldsymbol{\theta}^*)\}|T=t,\boldsymbol{X}=\boldsymbol{x}]$ *is continuously differentiable in* $(t,\boldsymbol{x})$*;*

(iii) $\mathbb{E}\left[\pi_0(T,\boldsymbol{X})m\{Y;g(T;\boldsymbol{\theta})\}\nabla_{\boldsymbol{\theta}}g(T;\boldsymbol{\theta})\right]$ *is differentiable w.r.t.* $\boldsymbol{\theta}$ *and* $\nabla_{\boldsymbol{\theta}}\mathbb{E}[\pi_0(T,\boldsymbol{X})$ $m\{Y;g(T;\boldsymbol{\theta})\}\nabla_{\boldsymbol{\theta}}g(T;\boldsymbol{\theta})]\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^*}$ *is nonsingular.*

Assumption 5. *(i)* $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\Theta}|m\{Y;g(T;\boldsymbol{\theta})\}|^{2+\delta}\right]<\infty$ *for some* $\delta>0$*; (ii) The function class* $\left\{m\{Y;g(T;\boldsymbol{\theta})\}:\boldsymbol{\theta}\in\Theta\right\}$ *satisfies:*

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}_1:\|\boldsymbol{\theta}_1-\boldsymbol{\theta}\|<\delta}|m\{Y;g(T;\boldsymbol{\theta}_1)\}-m\{Y;g(T;\boldsymbol{\theta})\}|^2\right]^{1/2}\leq a\cdot\delta^b$$

for any $\boldsymbol{\theta}\in\Theta$ *and any small* $\delta>0$ *and for some finite positive constants* a *and* $b\geq 1$.

Assumption 2 is essentially saying that the estimating equation is a.s. approximately satisfied, see Pakes and Pollard (1989). Assumption 3 is needed to bound the asymptotic variance of the test statistic. Assumption 4 (i) and (ii) impose sufficient regularity conditions on both the link function g and residual function m. Assumption 4 (iii) ensures that the variance of the test statistic is finite. Assumption 5 is a stochastic equicontinuity condition, which is needed for establishing the weak convergence of our test statistic, see Andrews (1994). Again, this is satisfied by the widely used loss functions such as $m\{y,g(t;\boldsymbol{\theta})\}=y-g(t;\boldsymbol{\theta})$ and $m\{y,g(t;\boldsymbol{\theta})\}=\tau-\mathbb{1}\{y<g(t;\boldsymbol{\theta})\}$ discussed in Section 2.

To aid presentation of the asymptotic properties of the test statistic, define the following quantities:

$$\begin{aligned}\phi(T_i, \boldsymbol{X}_i; t) :=& \pi_0(T_i, \boldsymbol{X}_i) \cdot \mathscr{H}(T_i, t) \cdot \mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X_i}] \\ &- \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} \cdot \mathscr{H}(T_i, t)|\boldsymbol{X}_i],\end{aligned}$$

and

$$\begin{aligned}\psi(T_i, \boldsymbol{X}_i, Y_i; t) :=& \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i] \cdot \nabla_{\boldsymbol{\theta}} g(T_i; \boldsymbol{\theta}^*)^\top \mathscr{H}(T_i, t)\right] \\ &\times \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i] \cdot \nabla_{\boldsymbol{\theta}} g(T_i; \boldsymbol{\theta}^*) \nabla_\theta^\top g(T_i; \boldsymbol{\theta}^*)\right]^{-1} \\ &\times \Big\{\pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} \nabla_\theta g(T_i; \boldsymbol{\theta}^*) \\ &\quad - \pi_0(T_i, \boldsymbol{X}_i) \nabla_\theta g(T_i; \boldsymbol{\theta}^*) \cdot \mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i] \\ &\quad + \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) \nabla_\theta g(T_i; \boldsymbol{\theta}^*) m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|\boldsymbol{X}_i]\Big\},\end{aligned}$$

and

$$\eta(T_i, \boldsymbol{X}_i, Y_i; t) := U_i \mathscr{H}(T_i, t) - \phi(T_i, \boldsymbol{X}_i; t) - \psi(T_i, \boldsymbol{X}_i, Y_i; t).$$

The next theorem establishes the weak convergence of $\widehat{J}_N(\cdot)$ and $\widehat{CM}_N$ under H_0.

Theorem 1. *Suppose that Assumptions 1-5 and Assumptions 6-9 listed in Appendix A hold, then under H_0,*

$$\begin{aligned}&(i) \quad \widehat{J}_N(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \eta(T_i, \boldsymbol{X}_i, Y_i; t) + o_P(1), \\ &(ii) \quad \widehat{J}_N(\cdot) \text{ converges weakly to } J_\infty(\cdot) \text{ in } L_2\{\mathcal{T}, dF_T(t)\},\end{aligned}$$

where J_∞ is a Gaussian process with zero mean and covariance function given by

$$\Sigma(t, t') = \mathbb{E}\left\{\eta(T_i, \boldsymbol{X}_i, Y_i; t)\eta(T_i, \boldsymbol{X}_i, Y_i; t')\right\}.$$

Furthermore,

$$(iii) \quad \widehat{CM}_N \text{ converges to } \int \{J_\infty(t)\}^2 dF_T(t) \text{ in distribution.}$$

The proof of Theorem 1 is relegated to Appendix B. Similar to Bierens and Ploberger (1997), Chen and Fan (1999), it can be shown that $\int\{J_\infty(t)\}^2 dF_T(t)$ can be written as an infinite sum of weighted (independent) χ_1^2 random variables with weights depending on the unknown distribution of $(T_i, \boldsymbol{X}_i, Y_i)$. Hence, it is difficult to obtain the exact critical values. We suggest a simulation method to approximate the critical values for the null limiting distribution of $\widehat{CM}_N$, see Section 5.

The next theorem shows that the proposed test statistic is more efficient than the infeasible test statistic constructed by using the true $\pi_0(T, \boldsymbol{X})$. Suppose that $\pi_0(T, \boldsymbol{X})$ was known, let $\widehat{\boldsymbol{\theta}}_0$ be the estimator of $\boldsymbol{\theta}^*$ constructed by using the true ratio function $\pi_0(T, \boldsymbol{X})$, which is defined to be the solution of the following equation:

$$\frac{1}{N}\sum_{i=1}^{N} \pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i; \boldsymbol{\theta})\right\} \nabla_\theta g(T_i; \boldsymbol{\theta}) = o_P\left(\frac{1}{\sqrt{N}}\right),$$

w.r.t. $\boldsymbol{\theta}$. The infeasible test statistic for H_0 is then based on

$$\widehat{J}_0(t) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \widehat{U}_{0i} \mathscr{H}(T_i, t), \text{ where } \widehat{U}_{0i} = \pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}}_0)\right\}.$$

Let

$$\begin{aligned}
\psi_0(T_i, \boldsymbol{X}_i, Y_i; t) :=& \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i)\cdot \frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i]\cdot \nabla_{\boldsymbol{\theta}} g(T_i; \boldsymbol{\theta}^*)^\top \mathscr{H}(T_i, t)\right] \\
&\times \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i)\cdot \frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i]\cdot \nabla_{\boldsymbol{\theta}} g(T_i; \boldsymbol{\theta}^*)\nabla_\theta^\top g(T_i; \boldsymbol{\theta}^*)\right]^{-1} \\
&\times \pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}\nabla_\theta g(T_i; \boldsymbol{\theta}^*),
\end{aligned}$$

and

$$\eta_0(T_i, \boldsymbol{X}_i, Y_i; t) := U_i \mathscr{H}(T_i, t) - \psi_0(T_i, \boldsymbol{X}_i, Y_i; t).$$

The following theorem establishes the weak convergence of $\widehat{J}_0(\cdot)$ under H_0 and shows that the asymptotic variance of the proposed test statistic $\widehat{J}_N(t)$ is smaller than that of $\widehat{J}_0(t)$ for any $t \in \mathcal{T}$.

Theorem 2. *Suppose that Assumptions 3-5 hold, then under H_0,*

$$(i) \quad \widehat{J}_0(t) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \eta_0(T_i, \boldsymbol{X}_i, Y_i; t) + o_P(1),$$

$$(ii) \quad \widehat{J}_0(\cdot) \text{ converges weakly to } J_{0,\infty}(\cdot) \text{ in } L_2\{\mathcal{T}, dF_T(t)\},$$

where $J_{0,\infty}$ is a Gaussian process with zero mean and covariance function given by

$$\Sigma_0(t, t') = \mathbb{E}\left\{\eta_0(T_i, \boldsymbol{X}_i, Y_i; t)\eta_0(T_i, \boldsymbol{X}_i, Y_i; t')\right\}.$$

Furthermore, $\Sigma_0(t, t) > \Sigma(t, t)$ for any $t \in \mathcal{T}$.

The proof of Theorem 2 is presented in Appendix C.

4.2 Special cases

This section discusses two important special continuous treatment effect models, the average and quantile continuous treatment models. In the case of testing for the average dose-response model, that is,

$$H_0 : \exists \text{ some } \boldsymbol{\theta}^* \in \Theta \subset \mathbb{R}^p, \text{ s.t. } \mathbb{E}\{Y^*(t)\} = g(t;\boldsymbol{\theta}^*) \text{ for all } t \in \mathcal{T}, \tag{4.1}$$

against the alternative hypothesis

$$H_1 : \nexists \text{ any } \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p, \text{ s.t. } \mathbb{E}\{Y^*(t)\} = g(t;\boldsymbol{\theta}) = 0 \text{ for all } t \in \mathcal{T},$$

$m\{Y^*(t); g(t;\boldsymbol{\theta}^*)\} = Y^*(t) - g(t;\boldsymbol{\theta}^*)$, $U_i^{ADRF} = \pi_0(T_i, \boldsymbol{X}_i)\{Y_i - g(T_i;\boldsymbol{\theta}^*)\}$ and the test statistic for H_0 is

$$\widehat{CM}_N^{ADRF} = \frac{1}{N}\sum_{i=1}^{N}\{\widehat{J}_N^{ADRF}(T_i)\}^2,$$

where

$$\widehat{J}_N^{ADRF}(t) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{U}_i^{ADRF}\mathscr{H}(T_i,t),\ \widehat{U}_i^{ADRF} = \widehat{\pi}_K(T_i,\boldsymbol{X}_i)\left\{Y_i - g(T_i;\widehat{\boldsymbol{\theta}})\right\}.$$

In this special case, the notations $\phi(T_i, \boldsymbol{X}_i; t)$, $\psi(T_i, \boldsymbol{X}_i, Y_i; t)$, and $\eta(T_i, \boldsymbol{X}_i, Y_i; t)$ in Theorem 1 become

$$\begin{aligned}\phi^{ADRF}(T_i, \boldsymbol{X}_i; t) :=& \pi_0(T_i,\boldsymbol{X}_i)\cdot\mathscr{H}(T_i,t)\cdot\mathbb{E}\{Y_i - g(T_i;\boldsymbol{\theta}^*)|T_i,\boldsymbol{X_i}\}\\ &- \mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)\{Y_i - g(T_i;\boldsymbol{\theta}^*)\}\cdot\mathscr{H}(T_i,t)|\boldsymbol{X}_i],\end{aligned}$$

and

$$\begin{aligned}\psi^{ADRF}(T_i, \boldsymbol{X}_i, Y_i; t) :=& \mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*)^\top\mathscr{H}(T_i,t)\right]\\ &\times\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*)\nabla_\theta^\top g(T_i;\boldsymbol{\theta}^*)\right]^{-1}\\ &\times\Big\{\pi_0(T_i,\boldsymbol{X}_i)\nabla_\theta g(T_i;\boldsymbol{\theta}^*)Y_i - \pi_0(T_i,\boldsymbol{X}_i)\nabla_\theta g(T_i;\boldsymbol{\theta}^*)\cdot\mathbb{E}(Y_i|T_i,\boldsymbol{X}_i)\\ &\quad + \mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)\nabla_\theta g(T_i;\boldsymbol{\theta}^*)\{Y_i - g(T_i;\boldsymbol{\theta}^*)\}|\boldsymbol{X}_i]\Big\},\end{aligned}$$

and

$$\eta^{ADRF}(T_i,\boldsymbol{X}_i,Y_i;t) := U_i^{ADRF}\mathscr{H}(T_i,t) - \phi^{ADRF}(T_i,\boldsymbol{X}_i;t) - \psi^{ADRF}(T_i,\boldsymbol{X}_i,Y_i;t).$$

Then Theorem 1 implies the following result.

Corollary 3. *Suppose that Assumptions 1-3 and Assumptions 6-9 listed in Appendix A hold, then under H_0,*

$$(i) \quad \widehat{J}_N^{ADRF}(\cdot) \text{ converges weakly to } J_\infty^{ADRF}(\cdot) \text{ in } L_2\{\mathcal{T}, dF_T(t)\},$$

where J_∞^{ADRF} is a Gaussian process with zero mean and covariance function given by

$$\Sigma^{ADRF}(t,t') = \mathbb{E}\left\{\eta^{ADRF}(T_i, \boldsymbol{X}_i, Y_i; t)\eta^{ADRF}(T_i, \boldsymbol{X}_i, Y_i; t')\right\}.$$

Furthermore,

$$(ii) \quad \widehat{CM}_N^{ADRF} \text{ converges to } \int \{J_\infty^{ADRF}(t)\}^2 dF_T(t) \text{ in distribution.}$$

In the case of testing for the quantile dose-response model, that is,

$$H_0 : \exists \text{ some } \boldsymbol{\theta}^* \in \Theta \subset \mathbb{R}^p, \text{ s.t. } F_{Y^*(t)}^{-1}(\tau) = g(t;\boldsymbol{\theta}^*) \text{ for all } t \in \mathcal{T}, \tag{4.2}$$

against the alternative hypothesis

$$H_1 : \nexists \text{ any } \boldsymbol{\theta} \in \Theta \subset \mathbb{R}^p, \text{ s.t. } F_{Y^*(t)}^{-1}(\tau) = g(t;\boldsymbol{\theta}) \text{ for all } t \in \mathcal{T},$$

$m\{Y^*(t); g(t;\boldsymbol{\theta}^*)\} = \tau - \mathbb{1}\{Y^*(t) < g(t;\boldsymbol{\theta}^*)\}$, $U_i^{QDRF} = \pi_0(T_i, \boldsymbol{X}_i)\big[\tau - \mathbb{1}\{Y_i < g(T_i;\boldsymbol{\theta}^*)\}\big]$, and the test statistic for H_0 is

$$\widehat{CM}_N^{QDRF} = \frac{1}{N}\sum_{i=1}^N [\widehat{J}_N^{QDRF}(T_i)]^2,$$

where

$$\widehat{J}_N^{QDRF}(t) = \frac{1}{\sqrt{N}}\sum_{i=1}^N \widehat{U}_i^{QDRF}\mathscr{H}(T_i,t), \ \widehat{U}_i^{QDRF} = \widehat{\pi}_K(T_i,\boldsymbol{X}_i)\left[\tau - \mathbb{1}\{Y_i < g(t;\widehat{\boldsymbol{\theta}})\}\right].$$

Again, in this special case, the notations $\phi(T_i, \boldsymbol{X}_i; t)$, $\psi(T_i, \boldsymbol{X}_i, Y_i; t)$, and $\eta(T_i, \boldsymbol{X}_i, Y_i; t)$ in Theorem 1 become

$$\begin{aligned}\phi^{QDRF}(T_i, \boldsymbol{X}_i; t) :=& \pi_0(T_i, \boldsymbol{X}_i) \cdot \mathbb{E}\Big([\tau - \mathbb{1}\{Y_i < g(T_i;\boldsymbol{\theta}^*)\}] \cdot \mathscr{H}(T_i,t) | T_i, \boldsymbol{X_i}\Big) \\ &- \mathbb{E}\Big\{\pi_0(T_i, \boldsymbol{X}_i)[\tau - \mathbb{1}\{Y_i < g(T_i;\boldsymbol{\theta}^*)\}] \cdot \mathscr{H}(T_i,t) | \boldsymbol{X}_i\Big\},\end{aligned}$$

and

$$\psi^{QDRF}(T_i, \boldsymbol{X}_i, Y_i; t) := \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot f_{Y|T,X}\{g(T_i;\boldsymbol{\theta}^*)|T_i, \boldsymbol{X}_i\} \cdot \nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*)^\top \mathscr{H}(T_i,t)\right]$$

$$\times \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot f_{Y|T,X}\{g(T_i;\boldsymbol{\theta}^*)|T_i, \boldsymbol{X}_i\} \cdot \nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*)\nabla_{\theta}^{\top} g(T_i;\boldsymbol{\theta}^*)\right]^{-1}$$
$$\times \Bigg\{ - \pi_0(T_i, \boldsymbol{X}_i)\nabla_{\theta} g(T_i;\boldsymbol{\theta}^*)\mathbb{1}\{Y_i < g(T_i;\boldsymbol{\theta}^*)\}$$
$$+ \pi_0(T_i, \boldsymbol{X}_i)\nabla_{\theta} g(T_i;\boldsymbol{\theta}^*) \cdot \mathbb{E}[\mathbb{1}\{Y_i < g(T_i;\boldsymbol{\theta}^*)\}|T_i, \boldsymbol{X}_i]$$
$$+ \mathbb{E}\Big[\pi_0(T_i, \boldsymbol{X}_i)\nabla_{\theta} g(T_i;\boldsymbol{\theta}^*)\left[\tau - \mathbb{1}\{Y_i < g(T_i;\boldsymbol{\theta}^*)\}\right]|\boldsymbol{X}_i\Big]\Bigg\},$$

and

$$\eta^{QDRF}(T_i, \boldsymbol{X}_i, Y_i; t) := U_i^{QDRF}\mathscr{H}(T_i, t) - \phi^{QDRF}(T_i, \boldsymbol{X}_i; t) - \psi^{QDRF}(T_i, \boldsymbol{X}_i, Y_i; t).$$

Then Theorem 1 implies the following result.

Corollary 4. *Suppose that Assumptions 1-3 and Assumptions 6-9 listed in Appendix A hold, then under H_0,*

$$(i) \quad \widehat{J}_N^{QDRF}(\cdot) \text{ converges weakly to } J_\infty^{QDRF}(\cdot) \text{ in } L_2\{\mathcal{T}, dF_T(t)\},$$

where J_∞^{QDRF} is a Gaussian process with zero mean and covariance function given by

$$\Sigma^{QDRF}(t,t') = \mathbb{E}\left\{\eta^{QDRF}(T_i, \boldsymbol{X}_i, Y_i; t)\eta^{QDRF}(T_i, \boldsymbol{X}_i, Y_i; t')\right\}.$$

Furthermore,

$$(ii) \quad \widehat{CM}_N^{QDRF} \text{ converges to } \int \{J_\infty^{QDRF}(t)\}^2 dF_T(t) \text{ in distribution.}$$

4.3 Asymptotic properties under the fixed and local alternative hypothesis

This section studies the asymptotic distribution of $\widehat{J}_N(\cdot)$ under the fixed and Pitman local alternatives. The Pitman local alternative is given by

$$H_L : \mathbb{E}\left[m\left\{Y^*(t); g(t;\boldsymbol{\theta}_N^*) + \frac{1}{\sqrt{N}} \cdot \delta(t)\right\}\right] = 0 \text{ for some } \boldsymbol{\theta}_N^* \in \Theta \text{ and all } t \in \mathcal{T},$$

where $\int\{\delta(t)\}^2 dF_T(t) < \infty$. With Assumption 1, H_L can be represented by

$$H_L : \mathbb{E}\left[\pi_0(T,\boldsymbol{X})m\left\{Y; g(T;\boldsymbol{\theta}_N^*) + \frac{1}{\sqrt{N}} \cdot \delta(T)\right\}\middle| T = t\right] = 0 \text{ for some } \boldsymbol{\theta}_N^* \in \Theta \text{ and all } t \in \mathcal{T},$$

which deviates from the null model at the rate of $O_p(N^{-1/2})$. Let $\boldsymbol{\theta}^*$ be the limit of $\boldsymbol{\theta}_N^*$ as $N \to \infty$, hence it solves the following equation:

$$\mathbb{E}\left[\pi_0(T,\boldsymbol{X})m\left\{Y; g(T;\boldsymbol{\theta}^*)\right\}\middle| T = t\right] = 0 \text{ for all } t \in \mathcal{T}.$$

Define

$$\begin{aligned}\mu(t) :=& \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i] \cdot \nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*)^\top \mathscr{H}(T_i, t)\right] \\ &\times \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i] \cdot \nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*)\nabla_{\theta}^\top g(T_i;\boldsymbol{\theta}^*)\right]^{-1} \\ &\times \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}\middle|T_i, \boldsymbol{X}_i\right] \cdot \delta(T_i) \cdot \nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*)\right].\end{aligned}$$

The following theorem gives the asymptotic distribution of $\widehat{J}_N(\cdot)$ under the local alternative H_L and the fixed alternative H_1.

Theorem 5. *Suppose that Assumptions 1-5 and Assumptions 6-9 listed in Appendix A hold. Under the local alternative hypothesis H_L,*

$$(i) \quad \widehat{J}_N(t) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\eta(T_i, \boldsymbol{X}_i, Y_i; t) + \mu(t) + o_P(1)\,, \tag{4.3}$$

$$(ii) \quad \widehat{J}_N(\cdot) \text{ converges weakly to } J_{\infty,\mu}(\cdot) \text{ in } L_2\{\mathcal{T}, dF_T(t)\}\,,$$

where $J_{\infty,\mu}$ is a Gaussian process with mean function $\mu(t)$ and covariance function given by

$$\Sigma(t, t') = \mathbb{E}\left\{\eta(T_i, \boldsymbol{X}_i, Y_i; t)\eta(T_i, \boldsymbol{X}_i, Y_i; t')\right\}.$$

Under the fixed H_1,

$$(iii) \quad \frac{1}{\sqrt{N}}\widehat{J}_N(\cdot) \text{ converges to } \mu_1(\cdot) \text{ in probability in } L^2(\mathcal{T}, dt),$$

where $\mu_1(t) := \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i)m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}\mathscr{H}(T_i, t)\right]$.

Comparing Theorem 5 (ii) to Theorem 1 (ii), we see that our test statistic is able to detect the local alternatives deviated from the null model at the rate of $O_p(N^{-1/2})$.

5 Approximation for the null limiting distribution

We know from Theorem 1 that $\widehat{CM}_N$ converges in distribution to $\int\{J_\infty(t)\}^2\,dF_T(t)$. Using techniques similar to those in Bierens and Ploberger (1997) and Chen and Fan (1999), one can show that $\int\{J_\infty(t)\}^2\,dF_T(t)$ is an infinite sum of weighted (independent) χ_1^2 random variables, where the weights depend on the unknown distribution of the $(\boldsymbol{X}_i, T_i, Y_i)$'s (see also Li, Hsiao, and Zinn, 2003). Obtaining the exact critical values is difficult and we here propose a simulation method to approximate the null limiting distribution. The method is a special case of the *exchangeable bootstrap* (Praestgaard and Wellner, 1993, Van Der Vaart and Wellner, 1996, Chernozhukov, Fernández-Val, and Melly, 2013, Donald and Hsu, 2014). Specifically, we first generate B sets of N independent standard normal random variables $w_{1,b}, \ldots, w_{N,b}$, for $b = 1, \ldots, B$ and B a large enough integer. Then we define

$$\widehat{J}^*_{N,b}(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} w_{i,b} \widehat{\eta}(T_i, \boldsymbol{X}_i, Y_i; t) \,, \tag{5.1}$$

where $\widehat{\eta}(T_i, \boldsymbol{X}_i, Y_i; t) = \widehat{U}_i \mathscr{H}(T_i, t) - \widehat{\phi}(T_i, \boldsymbol{X}_i; t) - \widehat{\psi}(T_i, \boldsymbol{X}_i, Y_i; t)$, with $\widehat{\phi}(T_i, \boldsymbol{X}_i; t)$ and $\widehat{\psi}(T_i, \boldsymbol{X}_i, Y_i; t)$ respectively some consistent nonparametric plug-in estimators of $\phi(T_i, \boldsymbol{X}_i; t)$ and $\psi(T_i, \boldsymbol{X}_i, Y_i; t)$ defined above in Theorem 1, for example the additive penalized spline estimator(see Ruppert, Wand, and Carroll, 2003 for example) or the series estimator used in Donald and Hsu (2014).

It is easy to see that $\mathbb{E}^*\{w_{i,b} \widehat{\eta}(T_i, \boldsymbol{X}_i, Y_i; t)\} = 0$ and $\mathbb{E}^*\{w_{i,b}^2 \widehat{\eta}(T_i, \boldsymbol{X}_i, Y_i; t) \widehat{\eta}(T_i, \boldsymbol{X}_i, Y_i; t')\} = \widehat{\eta}(T_i, \boldsymbol{X}_i, Y_i; t) \widehat{\eta}(T_i, \boldsymbol{X}_i, Y_i; t')$, for $i = 1, \ldots, N$, $b = 1, \ldots, B$ and all $t, t' \in \mathcal{T}$, where $\mathbb{E}^*\{\cdot\}$ is the conditional expectation given the data $(T_i, \boldsymbol{X}_i, Y_i)_{i=1}^N$. Because $\widehat{\eta}$ is a consistent estimator of η, we can see that $\widehat{J}^*_{N,b}(\cdot)$ has the same limiting process as $\widehat{J}_N(\cdot)$ for $b = 1, \ldots, B$. Then, we can approximate the limiting distribution of $\widehat{CM}_N$ under H_0 by

$$\widehat{CM}^*_{N,b} = \frac{1}{N} \sum_{i=1}^{N} \left\{ \widehat{J}^*_{N,b}(T_i) \right\}^2,$$

for $b = 1, \ldots, B$. That is, we can approximate the p-value by $B^{-1} \sum_{b=1}^{B} \mathbb{1}(\widehat{CM}^*_{N,b} \geq \widehat{CM}_N)$.

6 Numerical studies

6.1 Choosing K_1 and K_2

The large-sample properties of the proposed estimator hold for a range of values of K_1 and K_2. This presents a dilemma for applied researchers, who have only one finite sample. Too little smoothing yields a large variance and too much smoothing yields a large bias. Therefore, applied researchers would like to have some guidance on the choice of K_1 and K_2. In this section, we propose a cross-validation method for choosing the smoothing parameters K_1 and K_2. Specifically, we split the data set into F sets (say $F = 5$ or 10), and select K_1 and K_2 that minimize the following quantity

$$CV(K_1, K_2) = \sum_{j=1}^{F} \sum_{k \in S_j} \left[\widehat{\pi}_K^{(-j)}(T_k, \boldsymbol{X}_k) m \left\{ Y_k; g\left(T_k; \widehat{\boldsymbol{\theta}}^{(-j)} \right) \right\} \right]^2 , \tag{6.1}$$

where S_j denotes the jth set of data of T, $\boldsymbol{X}$ and Y, and for $j = 1, \ldots, F$,

$$\widehat{\boldsymbol{\theta}}^{(-j)} = \arg\min_{\boldsymbol{\theta}} \sum_{i \notin S_j} \widehat{\pi}_K^{(-j)}(T_i, \boldsymbol{X}_i) m \left\{ Y_i; g(T_i; \boldsymbol{\theta}) \right\} \nabla_{\boldsymbol{\theta}} g(T_i, \boldsymbol{\theta}) ,$$

with $\widehat{\pi}_K^{(-j)}(T_i, \boldsymbol{X}_i)$ obtained in the same way as that introduced in Section 2 via (3.4) and (3.5), but without using the individuals in S_j.

6.2 Simulation study

To assess the performance of our goodness-of-fit test method, we conducted Monte Carlo simulation studies on the following four data generating processes (DGPs):

DGP0-L $\quad T = 1 + 0.2X + \xi, \quad$ and $\quad Y = 1 + X + T + \epsilon,$

DGP0-NL $\quad T = 0.1X^2 + \xi, \quad$ and $\quad Y = X^2 + T + \epsilon,$

DGP1-L $\quad T = 1 + 0.2X + \xi, \quad$ and $\quad Y = 1 + X + T^3 + \epsilon,$

DGP1-NL $\quad T = 0.1X^2 + \xi, \quad$ and $\quad Y = X^2 + T^3 + \epsilon,$

where ξ and ϵ are independent standard normal random variables, and X is a uniform random variable supported on $[0, 1]$. For all the four scenarios, we considered the two-sided hypothesis testing in (2.1), where $m\{Y^*(t); g(t; \boldsymbol{\theta}^*)\} = Y^*(t) - g(t; \boldsymbol{\theta}^*)$ (average) and $m\{Y^*(t); g(t; \boldsymbol{\theta}^*)\} = 0.5 - \mathbb{1}\{Y^*(t) < g(t; \boldsymbol{\theta}^*)\}$ (median), and

$$g\{t; (\theta_0^*, \theta_1^*)\} = \theta_0^* + \theta_1^* t \,.$$

Clearly, H_0 is true for **DGP0-L** and **DGP0-NL**, but fails for **DGP1-L** and **DGP1-NL**. For each case, we generated 1000 samples of size 100, 200, and 500. The number of samples for the simulation-based approximation of the limiting process is $B = 500$ and the number of folds in the cross-validation (6.1) was taken to be $F = 10$. We compared the three commonly used weight functions $\mathscr{H}$ that are mentioned in Section 2, the logistic, the cosine-sine and the indicator ones. Specifically, for the logistic weight function, we took the constant $c = 5$.

Table 1: Estimated sizes

$m(\cdot)$	Model	N	Logistic 1%	Logistic 5%	Logistic 10%	Cosine-Sine 1%	Cosine-Sine 5%	Cosine-Sine 10%	Indicator 1%	Indicator 5%	Indicator 10%
Average	**DGP0-L**	100	0.017	0.078	0.133	0.017	0.076	0.132	0.011	0.062	0.113
		200	0.013	0.056	0.112	0.014	0.055	0.113	0.005	0.058	0.125
		500	0.013	0.051	0.101	0.010	0.047	0.099	0.013	0.053	0.106
	DGP0-NL	100	0.028	0.096	0.174	0.016	0.081	0.135	0.014	0.062	0.110
		200	0.019	0.084	0.140	0.007	0.062	0.127	0.010	0.059	0.120
		500	0.018	0.070	0.115	0.011	0.053	0.113	0.015	0.052	0.111
Median	**DGP0-L**	100	0.028	0.101	0.162	0.019	0.071	0.133	0.026	0.077	0.144
		200	0.018	0.074	0.142	0.016	0.063	0.117	0.012	0.066	0.120
		500	0.014	0.059	0.117	0.007	0.064	0.117	0.012	0.052	0.117
	DGP0-NL	100	0.063	0.148	0.233	0.016	0.082	0.136	0.018	0.074	0.148
		200	0.032	0.100	0.159	0.014	0.060	0.115	0.016	0.063	0.129
		500	0.018	0.069	0.128	0.009	0.056	0.111	0.013	0.057	0.107

Tables 1 and 2 summarize the empirical rejection probabilities computed at significance levels $1\%, 5\%$, and 10% for each case, which respectively show the estimated sizes (DGP0-L and DGP0NL) and the estimated powers (DGP1-L and DGP1-NL) of our test method.

We can see from Table 1 that the estimated sizes of our method with cosine-sine and indicator weight functions are quite close to the nominal sizes from $N = 100$ to 500 for all cases. The estimated sizes when using the logistic weight function are obviously over-sized when the sample size is small, especially for nonlinear $\boldsymbol{X}$ cases, but they also improve as the sample size increases and are close to the nominal sizes when $N = 500$.

From Table 2, we observed that all tests are quite powerful even when $N = 100$.

Overall, the simulation studies confirmed our asymptotic theorems and showed that in practice, the cosine-sine and indicator weight functions might perform better than the logistic one for nonlinear $\boldsymbol{X}$ cases.

6.3 Real data analysis

In this section, we applied our method to examine the model assumption made on the U.S. presidential campaign data in Ai, Linton, Motegi, and Zhang (2021). The data have been analyzed a lot in the treatment effect literature (Urban and Niebler, 2014, Fong, Hazlett, and Imai, 2018), where the interest was to explore the casual relationship between advertising and campaign contributions. The treatment of interest is the number of political advertisements aired in each zip code from non-competitive states, which ranges from 0 to 22379

Table 2: Estimated power

			Logistic			Cosine-Sine			Indicator		
$m(\cdot)$	Model	N	1%	5%	10%	1%	5%	10%	1%	5%	10%
Average	**DGP1-L**	100	0.999	0.999	0.999	0.999	0.999	1	0.998	1	1
		200	1	1	1	1	1	1	1	1	1
	DGP1-NL	100	0.995	1	1	0.998	0.998	1	0.998	0.999	1
		200	1	1	1	1	1	1	1	1	1
Median	**DGP1-L**	100	1	1	1	1	1	1	1	1	1
		200	1	1	1	1	1	1	1	1	1
	DGP1-NL	100	0.961	0.989	0.995	0.953	0.985	0.996	0.964	0.983	0.995
		200	0.999	1	1	1	1	1	1	1	1

across $N = 16265$ zip codes.

The data was first analyzed by Urban and Niebler (2014), who used a binary model to compare the campaign contributions of the 5230 zip codes that received more than 1000 advertisements with those of the other 11035 zip codes that received less than 1000 advertisements. Their research suggested that advertising in non-competitive states had a significant casual effect on the level of campaign contributions.

By contrast, Ai, Linton, Motegi, and Zhang (2021) treated the treatment variable (number of political advertisements) as a continuous variable and assumed that

$$\mathbb{E}\{Y^*(t)\} = \beta_1 + \beta_2 t + \beta_3 t^2 ,$$

where the observed outcome $Y^*(T) = \log(\text{Contribution}+1)$ and $T = \log(\#\text{Advertisements}+1)$. The covariates $\boldsymbol{X}$ considered were

$$\boldsymbol{X} = \begin{bmatrix} \log(\text{Population}) \\ \%\text{Age over 65} \\ \log(\text{Median Income}) \\ \%\text{Hispanic} \\ \%\text{Black} \\ \log(\text{Population density} + 1) \\ \%\text{College graduates} \\ \mathbb{1}(\text{Can commute to a competitive state}) \end{bmatrix} .$$

The definition of each covariate is almost self-explanatory and one can refer to Fong, Hazlett, and Imai (2018) for more details. Ai, Linton, Motegi, and Zhang (2021) found

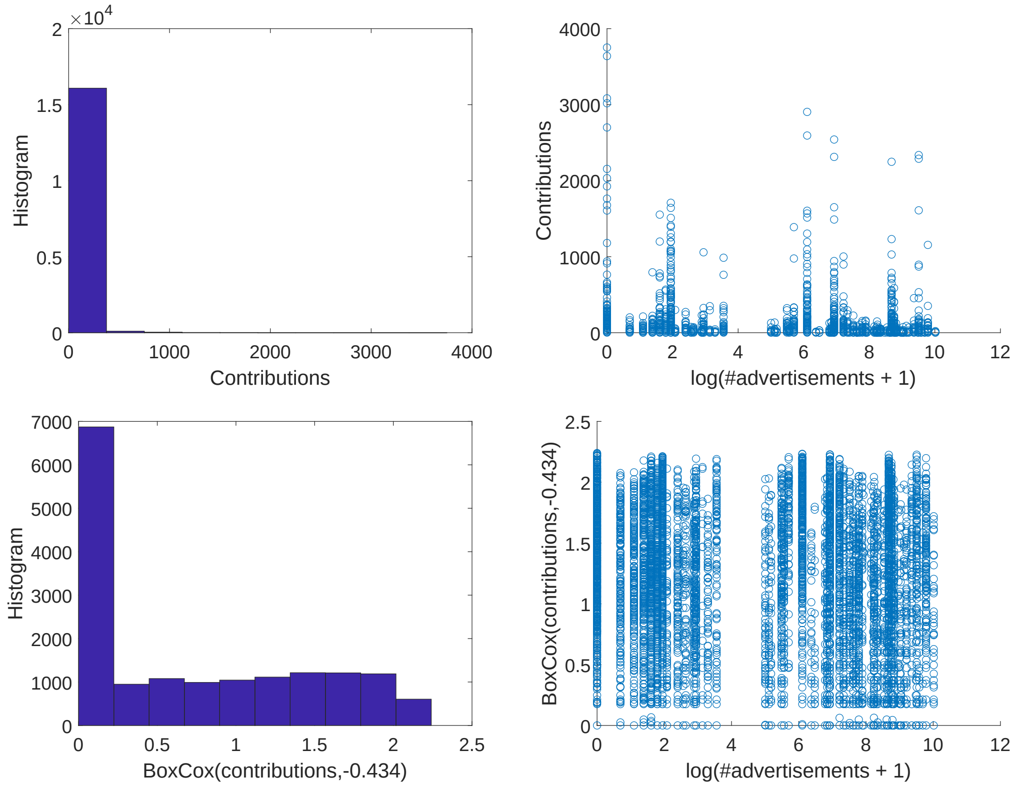

Figure 1: The histogram of the original campaign contribution data (top left) and the Box-Cox transformed contributions (bottom left), the scatter plot of the original campaign contribution data (top right) and the Box-Cox transformed ones (bottom right) versus the log-transformed number of political advertisements.

that the 95% confidence intervals for β_2 and β_3 were respectively $[-0.025, 0.232]$ and $[-0.025, 0.001]$, indicating that no significant causal link between advertising and campaign contributions was found from the linear model. Similar results were also reported by Fong, Hazlett, and Imai (2018). The authors then concluded that such opposing results from binary models and continuous linear models suggested a rather complex relationship between advertising and campaign contributions.

We reached the same conclusion in our data analysis. First, we examined the histogram of the original campaign contribution data and the scatter plot of the campaign contributions versus the log-transformed number of advertisements T. From the first row of Figure 1, we can see that the campaign contribution data are highly right-skewed both unconditionally and conditionally on T. That is, they are not likely to fit any linear models. We then conducted a log-transformation on the contribution data as in Ai, Linton, Motegi, and Zhang (2021). However, the results were similar.

Table 3: Estimated power with $J = 100$ and $B = 500$ from subsamples of U.S. presidential campaign data

	Logistic			Cosine-Sine			Indicator		
N	1%	5%	10%	1%	5%	10%	1%	5%	10%
200	0.67	0.83	0.90	0.72	0.84	0.91	0.64	0.83	0.85
500	0.71	0.84	0.91	0.73	0.84	0.91	0.65	0.81	0.87
1000	0.93	0.96	0.98	0.91	0.97	0.98	0.90	0.96	0.98

To make the data more likely to fit a linear model, we searched across Box-Cox transformations of the form $\{(\text{Contribution} + 1)^{\lambda} - 1\}/\lambda$ w.r.t. λ to find a transformation of the contribution whose sample quantiles have the largest correlation with those of a standard normal distribution. This yielded $\lambda = -0.4336$. The histogram and the scatter plot of the Box-Cox-transformed contribution data are shown in the second row of Figure 1. We can see that the transformed data are still highly right-skewed unconditionally. However, now, given T, the scatter plot no longer shows as much skewness.

We then applied our method with logistic, cosine-sine, and indicator weight functions with a $B = 500$ simulation-based approximation to the Box-Cox-transformed data to verify if they fit a linear model. Following Ai, Linton, Motegi, and Zhang (2021), we took $g(t; \boldsymbol{\theta}) = \theta_0 + \theta_1 t + \theta_2 t^2$, $u_{K_1}(T) = (1, T, T^2)^\top$ and $v_{K_2}(\boldsymbol{X}) = (1, \boldsymbol{X}^\top)^\top$ for estimating π_0. Unsurprisingly, all tests rejected the null hypothesis of a simple linear model. This leads to the same conclusion as Ai, Linton, Motegi, and Zhang (2021) that the relationship between advertisements and campaign contributions is rather complex, or there are some other confounding variables not included in $\boldsymbol{X}$.

Finally, we treated the full sample $N = 16256$ as a population, knowing that the linear model is not true for this population. We then randomly took some subsamples to see the power performance of our test. There, when we tried sample sizes larger than 1500, nearly all tests rejected the null hypothesis 100% of the time. We report the estimated power of our tests computed from 100 random subsamples of sample sizes 200, 500 and 1000 in Table 3. We can see that all tests perform similarly and powerfully.

Acknowledgement

The first author, Wei Huang's research was supported by the Professor Maurice H. Belz Fund of the University of Melbourne. The second author, Oliver Linton, acknowledges Cambridge INET for financial support. The last author, Zheng Zhang, acknowledges financial support from the National Natural Science Foundation of China through project 12001535, and the fund for building world-class universities (disciplines) of the Renmin University of China.

References

ABADIE, A., AND M. D. CATTANEO (2018): "Econometric methods for program evaluation," *Annual Review of Economics*, 10, 465–503.

ABREVAYA, J., Y.-C. HSU, AND R. P. LIELI (2015): "Estimating conditional average treatment effects," *Journal of Business & Economic Statistics*, 33(4), 485–505.

AI, C., O. LINTON, K. MOTEGI, AND Z. ZHANG (2021): "A Unified Framework for Efficient Estimation of General Treatment Models," *Quantitative Economics*, forthcoming.

AIT-SAHALIA, Y., P. J. BICKEL, AND T. M. STOKER (2001): "Goodness-of-fit tests for kernel regression with an application to option implied volatilities," *Journal of Econometrics*, 105(2), 363–412.

ANDREWS, D. W. K. (1994): "Empirical Process Methods in Econometrics," in *Handbook of Econometrics*, ed. by R. F. Engle, and D. L. McFadden, vol. 4, chap. 37, pp. 2247–2294. Citeseer.

AO, W., S. CALONICO, AND Y.-Y. LEE (2021): "Multivalued treatments and decomposition analysis: An application to the WIA program," *Journal of Business & Economic Statistics*, 39(1), 358–371.

ATHEY, S., G. W. IMBENS, AND S. WAGER (2018): "Approximate residual balancing: debiased inference of average treatment effects in high dimensions," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 80(4), 597–623.

BIERENS, H. J. (1982): "Consistent model specification tests," *Journal of Econometrics*, 20(1), 105–134.

——— (1990): "A consistent conditional moment test of functional form," *Econometrica: Journal of the Econometric Society*, pp. 1443–1458.

BIERENS, H. J., AND W. PLOBERGER (1997): "Asymptotic theory of integrated conditional moment tests," *Econometrica: Journal of the Econometric Society*, pp. 1129–1151.

BUCHINSKY, M. (1995): "Quantile regression, Box-Cox transformation model, and the U.S. wage structure, 1963–1987," *Journal of Econometrics*, 65(1), 109–154.

CATTANEO, M. D. (2010): "Efficient semiparametric estimation of multi-valued treatment effects under ignorability," *Journal of Econometrics*, 155(2), 138–154.

CHAN, K. C. G., S. C. P. YAM, AND Z. ZHANG (2016): "Globally efficient nonparametric inference of average treatment effects by empirical balancing calibration weighting," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 78(3), 673–700.

CHEN, X. (2007): "Large sample sieve estimation of semi-nonparametric models," *Handbook of Econometrics*, 6(B), 5549–5632.

CHEN, X., AND Y. FAN (1999): "Consistent hypothesis testing in semiparametric and nonparametric models for econometric time series.," *Journal of Econometrics*, 91, 373–401.

CHEN, Y.-T., Y.-C. HSU, AND H.-J. WANG (2020): "A stochastic frontier model with endogenous treatment status and mediator," *Journal of Business & Economic Statistics*, 38(2), 243–256.

CHERNOZHUKOV, V., I. FERNÁNDEZ-VAL, AND B. MELLY (2013): "Inference on counterfactual distributions," *Econometrica*, 81(6), 2205–2268.

COLANGELO, K., AND Y.-Y. LEE (2020): "Double debiased machine learning nonparametric inference with continuous treatments," *arXiv preprint arXiv:2004.03036*.

CRUMP, R. K., V. J. HOTZ, G. W. IMBENS, AND O. A. MITNIK (2008): "Nonparametric tests for treatment effect heterogeneity," *The Review of Economics and Statistics*, 90(3), 389–405.

DONALD, S. G., AND Y.-C. HSU (2014): "Estimation and inference for distribution functions and quantile functions in treatment effect models," *Journal of Econometrics*, 178(3), 383–397.

DONALD, S. G., Y.-C. HSU, AND R. P. LIELI (2014): "Testing the unconfoundedness assumption via inverse probability weighted estimators of (L) ATT," *Journal of Business & Economic Statistics*, 32(3), 395–415.

DONG, Y., Y.-Y. LEE, AND M. GOU (2019): "Regression discontinuity designs with a continuous treatment," *Available at SSRN 3167541*.

FAN, Q., Y.-C. HSU, R. P. LIELI, AND Y. ZHANG (2020): "Estimation of conditional average treatment effects with high-dimensional data," *Journal of Business & Economic Statistics*, pp. 1–15.

FAN, Y., AND Q. LI (1996): "Consistent model specification tests: omitted variables and semiparametric functional forms," *Econometrica: Journal of the econometric society*, pp. 865–890.

——— (2000): "Consistent model specification tests: Kernel-based tests versus Bierens' ICM tests," *Econometric Theory*, pp. 1016–1041.

FIRPO, S. (2007): "Efficient Semiparametric Estimation of Quantile Treatment Effects," *Econometrica*, 75(1), 259–276.

FONG, C., C. HAZLETT, AND K. IMAI (2018): "Covariate Balancing Propensity Score for a Continuous Treatment: Application to the Efficacy of Political Advertisements," *Annals of Applied Statistics*, 12(1), 156–177.

GALVAO, A. F., AND L. WANG (2015): "Uniformly semiparametric efficient estimation of treatment effects with a continuous treatment," *Journal of the American Statistical Association*, 110(512), 1528–1542.

HAHN, J. (1998): "On the role of the propensity score in efficient semiparametric estimation of average treatment effects," *Econometrica*, 66(2), 315–331.

HIRANO, K., AND G. W. IMBENS (2004): "The propensity score with continuous treatments," in *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, ed. by A. Gelman, and X.-L. Meng, chap. 7, pp. 73–84. John Wiley & Sons Ltd.

HIRANO, K., G. W. IMBENS, AND G. RIDDER (2003): "Efficient Estimation of Average Treatment Effects Using the Estimated Propensity Score," *Econometrica*, 71(4), 1161–1189.

HSU, Y.-C., T.-C. LAI, AND R. P. LIELI (2020): “Counterfactual Treatment Effects: Estimation and Inference,” *Journal of Business & Economic Statistics*, pp. 1–16.

HUBER, M., Y.-C. HSU, Y.-Y. LEE, AND L. LETTRY (2020): “Direct and indirect effects of continuous treatments based on generalized propensity score weighting,” *Journal of Applied Econometrics*, 35(7), 814–840.

IMAI, K., AND D. A. VAN DYK (2004): “Causal inference with general treatment regimes: Generalizing the propensity score,” *Journal of the American Statistical Association*, 99(467), 854–866.

IMBENS, G. W., AND J. M. WOOLDRIDGE (2009): “Recent developments in the econometrics of program evaluation,” *Journal of Economic Literature*, 47(1), 5–86.

KENNEDY, E. H., Z. MA, M. D. MCHUGH, AND D. S. SMALL (2017): “Non-parametric methods for doubly robust estimation of continuous treatment effects,” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(4), 1229–1245.

LEE, Y.-Y. (2018): “Efficient propensity score regression estimators of multivalued treatment effects for the treated,” *Journal of Econometrics*, 204(2), 207–222.

LI, Q. (1999): “Consistent model specification tests for time series econometric models,” *Journal of Econometrics*, 92(1), 101–147.

LI, Q., C. HSIAO, AND J. ZINN (2003): “Consistent specification tests for semiparametric/nonparametric models based on series estimation methods,” *Journal of Econometrics*, 112(2), 295–325.

NEWEY, W. K. (1997): “Convergence Rates and Asymptotic Normality for Series Estimators,” *Journal of Econometrics*, 79(1), 147–168.

PAKES, A., AND D. POLLARD (1989): “Simulation and the asymptotics of optimization estimators,” *Econometrica*, 57(5), 1027–1057.

PRAESTGAARD, J., AND J. A. WELLNER (1993): “Exchangeably weighted bootstraps of the general empirical process,” *The Annals of Probability*, 21(4), 2053–2086.

ROBINS, J. M., M. A. HERNÁN, AND B. BRUMBACK (2000): “Marginal structural models and causal inference in epidemiology,” *Epidemiology*, 11(5), 550–560.

ROSENBAUM, P. R., AND D. B. RUBIN (1983): "The central role of the propensity score in observational studies for causal effects," *Biometrika*, 70(1), 41–55.

ROSENBAUM, P. R., AND D. B. RUBIN (1984): "Reducing bias in observational studies using subclassification on the propensity score," *J. Am. Statist. Ass.*, 79(387), 516–524.

RUPPERT, D., M. P. WAND, AND R. J. CARROLL (2003): *Semiparametric Regression*. Cambridge University Press.

STINCHCOMBE, M. B., AND H. WHITE (1998): "Consistent specification testing with nuisance parameters present only under the alternative," *Econometric theory*, 14(3), 295–325.

STUTE, W. (1997): "Nonparametric model checks for regression," *The Annals of Statistics*, pp. 613–641.

URBAN, C., AND S. NIEBLER (2014): "Dollars on the Sidewalk: Should U.S. Presidential Candidates Advertise in Uncontested States?," *American Journal of Political Science*, 58(2), 322–336.

VAN DER VAART, A. W. (1998): *Asymptotic Statistics*. Cambridge University Press.

VAN DER VAART, A. W., AND J. A. WELLNER (1996): *Weak convergence and empirical processes with applications to statistics*. Springer.

ZHENG, J. X. (1996): "A consistent test of functional form via nonparametric estimation techniques," *Journal of Econometrics*, 75(2), 263–289.

Appendix

A Some preliminary results

We recall some preliminary results which have been established in Ai, Linton, Motegi, and Zhang (2021). The following conditions are inherited from Ai, Linton, Motegi, and Zhang (2021):

Assumption 6. *(i) The support $\mathcal{X}$ of $\boldsymbol{X}$ is a compact subset of $\mathbb{R}^r$. The support $\mathcal{T}$ of the treatment variable T is a compact subset of $\mathbb{R}$. (ii) There exist two positive constants η_1 and η_2 such that*

$$0 < \eta_1 \leq \pi_0(t, \boldsymbol{x}) \leq \eta_2 < \infty \ , \ \forall (t, \boldsymbol{x}) \in \mathcal{T} \times \mathcal{X} \ .$$

Assumption 7. *There exist* $\Lambda_{K_1\times K_2}\in\mathbb{R}^{K_1\times K_2}$ *and a positive constant* $\alpha>0$ *such that*

$$\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}\left|\rho'^{-1}\left\{\pi_0(t,\boldsymbol{x})\right\}-u_{K_1}(t)^\top\Lambda_{K_1\times K_2}v_{K_2}(\boldsymbol{x})\right|=O(K^{-\alpha}),$$

where $\rho(u)=-\exp(-u-1)$ *and* ρ'^{-1} *is the inverse function of* ρ'.

Assumption 8. *(i) For every* K_1 *and* K_2*, the smallest eigenvalues of* $\mathbb{E}\left[u_{K_1}(T)u_{K_1}(T)^\top\right]$ *and* $\mathbb{E}\left[v_{K_2}(\boldsymbol{X})v_{K_2}(\boldsymbol{X})^\top\right]$ *are bounded away from zero uniformly in* K_1 *and* K_2*. (ii) There are two sequences of constants* $\zeta_1(K_1)$ *and* $\zeta_2(K_2)$ *satisfying* $\sup_{t\in\mathcal{T}}\|u_{K_1}(t)\|\le\zeta_1(K_1)$ *and* $\sup_{\boldsymbol{x}\in\mathcal{X}}\|v_{K_2}(\boldsymbol{x})\|\le\zeta_2(K_2)$*,* $K=K_1(N)K_2(N)$ *and* $\zeta(K):=\zeta_1(K_1)\zeta_2(K_2)$*, such that* $\zeta(K)K^{-\alpha}\to0$ *and* $\zeta(K)\sqrt{K/N}\to0$ *as* $N\to\infty$.

Assumption 9. $\zeta(K)\sqrt{K^2/N}\to0$ *and* $\sqrt{N}K^{-\alpha}\to0$.

See Ai, Linton, Motegi, and Zhang (2021) for a detailed discussion on Assumptions 6 -9. Under these conditions, Ai, Linton, Motegi, and Zhang (2021, Theorem 3) established the following results:

Proposition 6. *Suppose that Assumptions 6-8 hold. Then, we obtain the following:*

$$\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}|\hat{\pi}_K(t,\boldsymbol{x})-\pi_0(t,\boldsymbol{x})|=O_p\left[\max\left\{\zeta(K)K^{-\alpha},\zeta(K)\sqrt{\frac{K}{N}}\right\}\right],$$
$$\int_{\mathcal{T}\times\mathcal{X}}|\hat{\pi}_K(t,\boldsymbol{x})-\pi_0(t,\boldsymbol{x})|^2dF_{T,X}(t,\boldsymbol{x})=O_p\left\{\max\left(K^{-2\alpha},\frac{K}{N}\right)\right\},$$
$$\frac{1}{N}\sum_{i=1}^N|\hat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)|^2=O_p\left\{\max\left(K^{-2\alpha},\frac{K}{N}\right)\right\}.$$

Furthermore, for any estimand with the form of $\mathbb{E}\{\pi_0(T,\boldsymbol{X})R(T,\boldsymbol{X},Y)\}$, where $R(T,\boldsymbol{X},Y)\in L^1(dF_{T,X,Y})$, Theorem 5 of Ai, Linton, Motegi, and Zhang (2021) provides an asymptotically equivalent representation for the plug-in estimator $N^{-1}\sum_{i=1}^N\widehat{\pi}_K(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i)$:

Proposition 7. *Suppose that Assumptions 6-9 hold. For any integrable function* $R(T,\boldsymbol{X},Y)$ *where* $\mathbb{E}\{R(T,\boldsymbol{X},Y)|T=t,\boldsymbol{X}=\boldsymbol{x}\}$ *is continuously differentiable. Then,*

$$\begin{aligned}&\frac{1}{\sqrt{N}}\sum_{i=1}^N\left[\widehat{\pi}_K(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i)-\mathbb{E}\{\pi_0(T,\boldsymbol{X})R(T,\boldsymbol{X},Y)\}\right]\\=&\frac{1}{\sqrt{N}}\sum_{i=1}^N\Big[\pi_0(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i)-\mathbb{E}\{\pi_0(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i)|T_i,\boldsymbol{X}_i\}\\&\qquad+\mathbb{E}\{\pi_0(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i)|T_i\}-\mathbb{E}\{\pi_0(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i)\}\\&\qquad+\mathbb{E}\{\pi_0(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i)|\boldsymbol{X}_i\}-\mathbb{E}\{\pi_0(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i)\}\Big]+o_p(1).\end{aligned}$$

B Proof of Theorem 1

Proof. Note that

$$\begin{aligned}\widehat{U}_i =&\widehat{\pi}_K(T_i, \boldsymbol{X}_i)m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\}\\ =&U_i + \{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\}m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\\ &+ \pi_0(T_i, \boldsymbol{X}_i)\left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\right]\\ &+ \{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\}\left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\right],\end{aligned}$$

where $U_i = \pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}$. Then, we have

$$\widehat{J}_N(t) =\frac{1}{\sqrt{N}}\sum_{i=1}^N \widehat{U}_i\mathscr{H}(T_i, t) = \frac{1}{\sqrt{N}}\sum_{i=1}^N U_i\mathscr{H}(T_i, t) \tag{B.1}$$

$$+ \frac{1}{\sqrt{N}}\sum_{i=1}^N \{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\}m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\mathscr{H}(T_i, t) \tag{B.2}$$

$$+ \frac{1}{\sqrt{N}}\sum_{i=1}^N \pi_0(T_i, \boldsymbol{X}_i)\left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\right]\mathscr{H}(T_i, t) \tag{B.3}$$

$$+ \frac{1}{\sqrt{N}}\sum_{i=1}^N \{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\}\left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\right]\mathscr{H}(T_i, t). \tag{B.4}$$

Using Proposition 7, under $H_0 : \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}|T_i] = 0$, we have

$$\begin{aligned}(\text{B.2}) =& -\frac{1}{\sqrt{N}}\sum_{i=1}^N \pi_0(T_i, \boldsymbol{X}_i)\cdot\mathbb{E}[m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\cdot\mathscr{H}(T_i, t)|T_i, \boldsymbol{X_i}]\\ &+ \frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\cdot\mathscr{H}(T_i, t)|T_i]\\ &+ \frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\cdot\mathscr{H}(T_i, t)|\boldsymbol{X}_i]\\ &- \frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\cdot\mathscr{H}(T_i, t)] + o_P(1)\\ =& -\frac{1}{\sqrt{N}}\sum_{i=1}^N \pi_0(T_i, \boldsymbol{X}_i)\cdot\mathbb{E}[m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\cdot\mathscr{H}(T_i, t)|T_i, \boldsymbol{X_i}]\\ &+ \frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\cdot\mathscr{H}(T_i, t)|\boldsymbol{X}_i] + o_P(1)\end{aligned}$$

$$= -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\phi(T_i, \boldsymbol{X}_i; t) + o_P(1). \tag{B.5}$$

By Ai, Linton, Motegi, and Zhang (2021, Theorems 4 and 5), under H_0, we have

$$\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| = O_P(N^{-1/2}). \tag{B.6}$$

We next find the expression for $\sqrt{N}\{\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\}$. Note that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\pi}_K(T_i, \boldsymbol{X}_i)m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\}\nabla_{\boldsymbol{\theta}}g(T_i; \widehat{\boldsymbol{\theta}}) = o_P(1).$$

Note that $m(\cdot)$ may not be differentiable, and we cannot simply apply the mean value theorem to obtain the expression for $\sqrt{N}\{\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\}$. We apply the empirical process theory in Andrews (1994) to solve this problem. Let

$$\nu_N(f) := \frac{1}{\sqrt{N}}\sum_{i=1}^{N}[f(T_i, \boldsymbol{X}_i, Y_i) - \mathbb{E}\{f(T_i, \boldsymbol{X}_i, Y_i)\}]$$

be the empirical process indexed by $f(\cdot)$. Note that

$$\begin{aligned}
o_P(1) =& \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\pi}_K(T_i, \boldsymbol{X}_i)m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\}\nabla_{\boldsymbol{\theta}}g(T_i; \widehat{\boldsymbol{\theta}}) \\
=& \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\pi_0(T_i, \boldsymbol{X}_i)m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\}\nabla_{\boldsymbol{\theta}}g(T_i; \widehat{\boldsymbol{\theta}}) && \text{(B.7)} \\
&+ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\}\, m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\nabla_{\boldsymbol{\theta}}g(T_i; \boldsymbol{\theta}^*) && \text{(B.8)} \\
&+ \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\} && \text{(B.9)} \\
&\qquad \times \left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\}\nabla_{\boldsymbol{\theta}}g(T_i; \widehat{\boldsymbol{\theta}}) - m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\nabla_{\boldsymbol{\theta}}g(T_i; \boldsymbol{\theta}^*)\right].
\end{aligned}$$

For (B.8), by Proposition 7, under H_0, we have

$$\begin{aligned}
\text{(B.8)} =& \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\}\, m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\nabla_{\boldsymbol{\theta}}g(T_i; \boldsymbol{\theta}^*) \\
=& -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\pi_0(T_i, \boldsymbol{X}_i)\cdot\mathbb{E}\left[m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\,|T_i, \boldsymbol{X}_i\right]\cdot\nabla_{\boldsymbol{\theta}}g(T_i; \boldsymbol{\theta}^*)
\end{aligned}$$

$$
\begin{aligned}
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i\right]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)|\boldsymbol{X}_i\right]\\
&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\right]+o_P(1)\\
=&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i\right]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)|\boldsymbol{X}_i\right]+o_P(1).
\end{aligned}
$$

For (B.9), we have

$$
\begin{aligned}
|(\text{B.9})|=&\left\|\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}\right.\\
&\left.\times\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}\nabla_{\boldsymbol{\theta}}g(T_i;\widehat{\boldsymbol{\theta}})-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\right]\right\|\\
\leq&\sqrt{N}\cdot\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}|\widehat{\pi}_K(t,\boldsymbol{x})-\pi_0(t,\boldsymbol{x})|\\
&\cdot\frac{1}{N}\sum_{i=1}^{N}\left[\left\|m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\right\|\right.\\
&\left.+\left\|m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}\nabla^2_{\boldsymbol{\theta}}g(T_i;\widetilde{\boldsymbol{\theta}})\right\|\cdot\left\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right\|\right]\\
=&\sqrt{N}\cdot O_P\left(\zeta(K)K^{-\alpha}+\zeta(K)\sqrt{\frac{K}{N}}\right)\\
&\cdot\left\{\mathbb{E}\left[\left|m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right|\cdot\|\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\|\right]+O_P\left(N^{-1/2}\right)\right\}\\
\leq&O_P\left(\zeta(K)K^{-\alpha}+\zeta(K)\sqrt{\frac{K}{N}}\right)\cdot\sqrt{N}\cdot\left\{O(1)\cdot\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|+O_P\left(N^{-1/2}\right)\right\}\\
=&o_P(1),
\end{aligned}
\tag{B.10}
$$

where the second equality holds by Proposition 6 and the law of large numbers; the second inequality holds by Assumption 5; and the last equality holds by (B.6) and Assumption 8.

For (B.7), we have

$$
\begin{aligned}
(\text{B.7}) =& \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \pi_0(T_i, \boldsymbol{X}_i) m \left\{ Y_i; g(T_i; \widehat{\boldsymbol{\theta}}) \right\} \nabla_{\boldsymbol{\theta}} g(T_i; \widehat{\boldsymbol{\theta}}) \\
=& \nu_N \left(\pi_0(T_i, \boldsymbol{X}_i) m \left\{ Y_i; g(T_i; \boldsymbol{\theta}^*) \right\} \nabla_{\boldsymbol{\theta}} g(T_i; \boldsymbol{\theta}^*) \right) \\
&+ \left\{ \nu_N \left(\pi_0(T_i, \boldsymbol{X}_i) m \left\{ Y_i; g(T_i; \widehat{\boldsymbol{\theta}}) \right\} \nabla_{\boldsymbol{\theta}} g(T_i; \widehat{\boldsymbol{\theta}}) \right) - \nu_N \left(\pi_0(T_i, \boldsymbol{X}_i) m \left(Y_i; g(T_i; \boldsymbol{\theta}^*) \right) \nabla_{\boldsymbol{\theta}} g(T_i; \boldsymbol{\theta}^*) \right) \right\} \\
&+ \sqrt{N} \cdot \mathbb{E} \left[\pi_0(T_i, \boldsymbol{X}_i) m \left\{ Y_i; g(T_i; \widehat{\boldsymbol{\theta}}) \right\} \nabla_{\boldsymbol{\theta}} g(T_i; \widehat{\boldsymbol{\theta}}) \right] \\
=& \nu_N \left[\pi_0(T_i, \boldsymbol{X}_i) m \left\{ Y_i; g(T_i; \boldsymbol{\theta}^*) \right\} \nabla_{\boldsymbol{\theta}} g(T_i; \boldsymbol{\theta}^*) \right] + o_P(1) \\
&+ \sqrt{N} \cdot \mathbb{E} \left[\pi_0(T_i, \boldsymbol{X}_i) m \left\{ Y_i; g(T_i; \widehat{\boldsymbol{\theta}}) \right\} \nabla_{\boldsymbol{\theta}} g(T_i; \widehat{\boldsymbol{\theta}}) \right],
\end{aligned}
$$

where the last equality holds because Assumption 5 and the compactness of Θ imply the empirical process

$$
\left\{ \nu_N \left[\pi_0(T_i, \boldsymbol{X}_i) m \left\{ Y_i; g(T_i; \boldsymbol{\theta}) \right\} \nabla_{\boldsymbol{\theta}} g(T_i; \boldsymbol{\theta}) \right] : \boldsymbol{\theta} \in \Theta \right\}
$$

is stochastically equicontinuous (Andrews (1994, Theorems 4 and 5)), and $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \xrightarrow{p} 0$, then

$$
\nu_N \left[\pi_0(T_i, \boldsymbol{X}_i) m \left\{ Y_i; g(T_i; \widehat{\boldsymbol{\theta}}) \right\} \nabla_{\boldsymbol{\theta}} g(T_i; \widehat{\boldsymbol{\theta}}) \right] - \nu_N \left[\pi_0(T_i, \boldsymbol{X}_i) m \left\{ Y_i; g(T_i; \boldsymbol{\theta}^*) \right\} \nabla_{\boldsymbol{\theta}} g(T_i; \boldsymbol{\theta}^*) \right] = o_P(1).
$$

Note that under H_0, $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}^*$:

$$
\begin{aligned}
&\sqrt{N} \cdot \mathbb{E} \left[\pi_0(T_i, \boldsymbol{X}_i) m \left\{ Y_i; g(T_i; \widehat{\boldsymbol{\theta}}) \right\} \nabla_{\boldsymbol{\theta}} g(T_i; \widehat{\boldsymbol{\theta}}) \right] \\
=& \sqrt{N} \cdot \mathbb{E} \left[\pi_0(T_i, \boldsymbol{X}_i) m \left\{ Y_i; g(T_i; \boldsymbol{\theta}^*) \right\} \nabla_{\boldsymbol{\theta}} g(T_i; \boldsymbol{\theta}^*) \right] \\
&+ \sqrt{N} \cdot \nabla_{\boldsymbol{\theta}} \mathbb{E} \left[\pi_0(T_i, \boldsymbol{X}_i) m \left\{ Y_i; g(T_i; \widetilde{\boldsymbol{\theta}}) \right\} \nabla_{\boldsymbol{\theta}} g(T_i; \widetilde{\boldsymbol{\theta}}) \right] \cdot \left\{ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \right\} \\
=& \nabla_{\boldsymbol{\theta}} \mathbb{E} \left[\pi_0(T_i, \boldsymbol{X}_i) m \left\{ Y_i; g(T_i; \widetilde{\boldsymbol{\theta}}) \right\} \nabla_{\boldsymbol{\theta}} g(T_i; \widetilde{\boldsymbol{\theta}}) \right] \cdot \sqrt{N} \left\{ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \right\} + o_P(1),
\end{aligned}
$$

where $\widetilde{\boldsymbol{\theta}}$ lies between $\boldsymbol{\theta}^*$ and $\widehat{\boldsymbol{\theta}}$. Then we get

$$
\begin{aligned}
(\text{B.7}) =& \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \pi_0(T_i, \boldsymbol{X}_i) m \left\{ Y_i; g(T_i; \boldsymbol{\theta}^*) \right\} \nabla_{\theta} g(T_i; \boldsymbol{\theta}^*) \\
&+ \nabla_{\boldsymbol{\theta}} \mathbb{E} \left[\pi_0(T_i, \boldsymbol{X}_i) m \left\{ Y_i; g(T_i; \widetilde{\boldsymbol{\theta}}) \right\} \nabla_{\boldsymbol{\theta}} g(T_i; \widetilde{\boldsymbol{\theta}}) \right] \cdot \sqrt{N} \left\{ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \right\} + o_P(1)
\end{aligned}
$$

Hence, combining the expressions of (B.7), (B.8), and (B.9), we get

$$
\sqrt{N} \left\{ \widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^* \right\} \tag{B.11}
$$

$$
\begin{aligned}
&= \left\{-\nabla_{\boldsymbol{\theta}}\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_\theta g(T_i;\boldsymbol{\theta}^*)\right]\right\}^{-1} \\
&\quad \times \frac{1}{\sqrt{N}}\sum_{i=1}^N \Big\{\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_\theta g(T_i;\boldsymbol{\theta}^*) \\
&\qquad\qquad - \pi_0(T_i,\boldsymbol{X}_i)\nabla_\theta g(T_i;\boldsymbol{\theta}^*)\cdot\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*\right\})|T_i,\boldsymbol{X}_i] \\
&\qquad\qquad + \mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)\nabla_\theta g(T_i;\boldsymbol{\theta}^*)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i] \\
&\qquad\qquad - \mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)\nabla_\theta g(T_i;\boldsymbol{\theta}^*)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}] \\
&\qquad\qquad + \mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)\nabla_\theta g(T_i;\boldsymbol{\theta}^*)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|\boldsymbol{X}_i] + o_P(1) \\
&= \left\{-\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\nabla_\theta^\top g(T_i;\boldsymbol{\theta}^*)\right]\right\}^{-1} \\
&\quad \times \frac{1}{\sqrt{N}}\sum_{i=1}^N \Big\{\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_\theta g(T_i;\boldsymbol{\theta}^*) \\
&\qquad\qquad - \pi_0(T_i,\boldsymbol{X}_i)\nabla_\theta g(T_i;\boldsymbol{\theta}^*)\cdot\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i] \\
&\qquad\qquad + \mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)\nabla_\theta g(T_i;\boldsymbol{\theta}^*)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|\boldsymbol{X}_i]\Big\} + o_P(1),
\end{aligned}
$$

where the second equality holds by noting $\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i\right] = 0$ under H_0.

Consider the term (B.3). Note that

$$
\begin{aligned}
\text{(B.3)} =& \nu_N\left\{\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\} - m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t)\right\} \\
&+ \sqrt{N}\cdot\mathbb{E}\left\{\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta})\right\} - m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t)\right\}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}.
\end{aligned}
$$

By Assumption 5, the compactness of Θ, and Andrews (1994, Theorems 4 and 5), then the empirical process

$$
\left\{\nu_N\left[\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta})\right\} - m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t)\right] : \boldsymbol{\theta}\in\Theta\right\}
$$

is stochastically equicontinuous. With $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \xrightarrow{p} 0$ under H_0, we have

$$
\nu_N\left\{\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\} - m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t)\right\} = o_P(1).
$$

Using the mean value theorem and $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \xrightarrow{p} 0$ under H_0, we have

$$
\begin{aligned}
&\sqrt{N}\cdot\mathbb{E}\left\{\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta})\right\} - m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t)\right\}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}} \\
&= \left\{\nabla_{\boldsymbol{\theta}}\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta})\right\}\mathscr{H}(T_i,t)\right]\Big|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}\right\}^\top\cdot\sqrt{N}\left\{\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right\}
\end{aligned}
$$

$$=\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}\{m\left(Y_i;g(T_i;\boldsymbol{\theta})\right\}|T_i,\boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta})^\top\mathscr{H}(T_i,t)\right]\Bigg|_{\boldsymbol{\theta}=\widetilde{\boldsymbol{\theta}}}\cdot\sqrt{N}\left\{\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right\}$$
$$=\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)^\top\mathscr{H}(T_i,t)\right]\cdot\sqrt{N}\left\{\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right\}+o_P(1).$$

By (B.11), we have

$$(\text{B.3})=-\frac{1}{\sqrt{N}}\sum_{i=1}^N\psi(T_i,\boldsymbol{X}_i,Y_i;t)+o_p(1),\tag{B.12}$$

where

$$\begin{aligned}\psi(T_i,\boldsymbol{X}_i,Y_i;t):=&\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)^\top\mathscr{H}(T_i,t)\right]\\&\times\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\nabla_\theta^\top g(T_i;\boldsymbol{\theta}^*)\right]^{-1}\\&\times\Big\{\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_\theta g(T_i;\boldsymbol{\theta}^*)\\&\quad-\pi_0(T_i,\boldsymbol{X}_i)\nabla_\theta g(T_i;\boldsymbol{\theta}^*)\cdot\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\\&\quad+\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)\nabla_\theta g(T_i;\boldsymbol{\theta}^*)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|\boldsymbol{X}_i]\Big\}.\end{aligned}$$

For the term (B.4), we have

$$\begin{aligned}|(\text{B.4})|=&\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t)\right|\\\le&\sqrt{N}\cdot\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}|\widehat{\pi}_K(t,\boldsymbol{x})-\pi_0(t,\boldsymbol{x})|\\&\cdot\frac{1}{N}\sum_{i=1}^N\left|m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\mathscr{H}(T_i,t)\right|\\=&\sqrt{N}\cdot O_P\left(\zeta(K)K^{-\alpha}+\zeta(K)\sqrt{\frac{K}{N}}\right)\\&\cdot\left\{\mathbb{E}\left[\left|m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right|\cdot|\mathscr{H}(T_i,t)|\right]+O_P\left(N^{-1/2}\right)\right\}\\\le&O_P\left(\zeta(K)K^{-\alpha}+\zeta(K)\sqrt{\frac{K}{N}}\right)\cdot\sqrt{N}\cdot\left\{O(1)\cdot\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|+O_P\left(N^{-1/2}\right)\right\}\\=&o_P(1),\end{aligned}\tag{B.13}$$

where the second equality holds by Proposition 6 and the law of large numbers; the second inequality holds by Assumption 5; and the last equality holds by (B.6) and Assumption 8.

Hence, combining (B.1), (B.5), (B.12), and (B.13), we have

$$\begin{aligned}\widehat{J}_N(t) =& \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\eta(T_i, \boldsymbol{X}_i, Y_i; t) + o_P(1)\\ =& \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\{U_i\mathscr{H}(T_i, t) - \phi(T_i, \boldsymbol{X}_i; t) - \psi(T_i, \boldsymbol{X}_i, Y_i; t)\} + o_P(1),\end{aligned}$$

where $\mathbb{E}\{\phi(T_i, \boldsymbol{X}_i; t)\} = 0$ and $\mathbb{E}\{\psi(T_i, \boldsymbol{X}_i, Y_i; t)\} = 0$. Therefore, under the null hypothesis H_0, $\widehat{J}_N(\cdot)$ weakly converges to $J_\infty(\cdot)$ in $L_2(\mathcal{T}, dt)$, where $J_\infty(\cdot)$ is a Gaussian process with zero mean and covariance function given by

$$\Sigma(t, t') = \mathbb{E}\left\{\eta(T_i, \boldsymbol{X}_i, Y_i; t)\eta(T_i, \boldsymbol{X}_i, Y_i; t')\right\}.$$

(ii) Obviously, $h(J) := \int\{J(t)\}^2 dF_T(t)$ is a continuous function in $L_2(\mathcal{T}, dF_T)$. Given that $F_T(\cdot)$ is absolutely continuous with respect to the Lebesgue measure, $h(J)$ is also continuous in $L_2(\mathcal{T}, dt)$. Therefore, by Theorem 1 (i) and the continuous mapping theorem, we have that $\int\{\widehat{J}(t)\}^2 dF_T(t)$ converges to $\int\{J_\infty(t)\}^2 dF_T(t)$ in distribution. $\square$

C Proof of Theorem 2

Similar to Theorem 1, results (i) and (ii) can be established. We next prove $\Sigma_0(t, t) > \Sigma(t, t)$ for any fixed $t \in \mathcal{T}$. Let

$$\begin{aligned}A_t :=& \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}} g(T_i; \boldsymbol{\theta}^*)^\top\mathscr{H}(T_i, t)\right]\\ &\times\mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}} g(T_i; \boldsymbol{\theta}^*)\nabla_\theta^\top g(T_i; \boldsymbol{\theta}^*)\right]^{-1}.\end{aligned}$$

Then

$$\begin{aligned}\psi(T_i, \boldsymbol{X}_i, Y_i; t) :=& A_t\cdot\Big\{\pi_0(T_i, \boldsymbol{X}_i)m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}\nabla_\theta g(T_i; \boldsymbol{\theta}^*)\\ &- \pi_0(T_i, \boldsymbol{X}_i)\nabla_\theta g(T_i; \boldsymbol{\theta}^*)\cdot\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X}_i]\\ &+ \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)\nabla_\theta g(T_i; \boldsymbol{\theta}^*)m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|\boldsymbol{X}_i]\Big\},\end{aligned}$$

and

$$\psi_0(T_i, \boldsymbol{X}_i, Y_i; t) := A_t\cdot\pi_0(T_i, \boldsymbol{X}_i)m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}\nabla_\theta g(T_i; \boldsymbol{\theta}^*).$$

We have

$$
\begin{aligned}
&\Sigma(t,t) = \mathbb{E}\left[\{\eta(T_i,\boldsymbol{X}_i,Y_i;t)\}^2\right] \\
=&\mathbb{E}\left[\{U_i\mathscr{H}(T_i,t) - \phi(T_i,\boldsymbol{X}_i;t) - \psi(T_i,\boldsymbol{X}_i,Y_i;t)\}^2\right] \\
=&\mathbb{E}\Bigg[\Bigg\{U_i\mathscr{H}(T_i,t) - A_t\cdot\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_\theta g(T_i;\boldsymbol{\theta}^*) \\
&\qquad - \pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X_i}]\cdot(\mathscr{H}(T_i,t) - A_t\cdot\nabla_\theta g(T_i;\boldsymbol{\theta}^*)) \\
&\qquad + \mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\cdot(\mathscr{H}(T_i,t) - A_t\cdot\nabla_\theta g(T_i;\boldsymbol{\theta}^*))\,|\boldsymbol{X}_i]\Bigg\}^2\Bigg] \\
=&\mathbb{E}\left[\left\{U_i\mathscr{H}(T_i,t) - A_t\cdot\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_\theta g(T_i;\boldsymbol{\theta}^*)\right\}^2\right] \\
&+\mathbb{E}\left[\left\{\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X_i}]\cdot(\mathscr{H}(T_i,t) - A_t\cdot\nabla_\theta g(T_i;\boldsymbol{\theta}^*))\right\}^2\right] \\
&+\mathbb{E}\left[\left\{\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\cdot(\mathscr{H}(T_i,t) - A_t\cdot\nabla_\theta g(T_i;\boldsymbol{\theta}^*))\,|\boldsymbol{X}_i]\right\}^2\right] \\
&-2\cdot\mathbb{E}\Bigg[\left\{U_i\mathscr{H}(T_i,t) - A_t\cdot\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_\theta g(T_i;\boldsymbol{\theta}^*)\right\} \\
&\qquad\quad\times\left\{\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X_i}]\cdot(\mathscr{H}(T_i,t) - A_t\cdot\nabla_\theta g(T_i;\boldsymbol{\theta}^*))\right\}\Bigg] \\
&+2\cdot\mathbb{E}\Bigg[\left\{U_i\mathscr{H}(T_i,t) - A_t\cdot\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_\theta g(T_i;\boldsymbol{\theta}^*)\right\} \\
&\qquad\quad\times\left\{\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\cdot(\mathscr{H}(T_i,t) - A_t\cdot\nabla_\theta g(T_i;\boldsymbol{\theta}^*))\,|\boldsymbol{X}_i]\right\}\Bigg] \\
&-2\cdot\mathbb{E}\Bigg[\left\{\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X_i}]\cdot(\mathscr{H}(T_i,t) - A_t\cdot\nabla_\theta g(T_i;\boldsymbol{\theta}^*))\right\} \\
&\qquad\quad\times\left\{\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\cdot(\mathscr{H}(T_i,t) - A_t\cdot\nabla_\theta g(T_i;\boldsymbol{\theta}^*))\,|\boldsymbol{X}_i]\right\}\Bigg] \\
=&\mathbb{E}\left[\left\{U_i\mathscr{H}(T_i,t) - A_t\cdot\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_\theta g(T_i;\boldsymbol{\theta}^*)\right\}^2\right] \\
&+\mathbb{E}\left[\left\{\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X_i}]\cdot(\mathscr{H}(T_i,t) - A_t\cdot\nabla_\theta g(T_i;\boldsymbol{\theta}^*))\right\}^2\right] \\
&+\mathbb{E}\left[\left\{\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\cdot(\mathscr{H}(T_i,t) - A_t\cdot\nabla_\theta g(T_i;\boldsymbol{\theta}^*))\,|\boldsymbol{X}_i]\right\}^2\right]
\end{aligned}
$$

$$
\begin{aligned}
&- 2 \cdot \mathbb{E}\left[\left\{\pi_0(T_i, \boldsymbol{X}_i) \cdot \mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X_i}] \cdot (\mathscr{H}(T_i, t) - A_t \cdot \nabla_\theta g(T_i; \boldsymbol{\theta}^*))\right\}^2\right] \\
&+ 2 \cdot \mathbb{E}\left[\left\{\mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} \cdot (\mathscr{H}(T_i, t) - A_t \cdot \nabla_\theta g(T_i; \boldsymbol{\theta}^*))\,|\boldsymbol{X}_i]\right\}^2\right] \\
&- 2 \cdot \mathbb{E}\left[\left\{\mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} \cdot (\mathscr{H}(T_i, t) - A_t \cdot \nabla_\theta g(T_i; \boldsymbol{\theta}^*))\,|\boldsymbol{X}_i]\right\}^2\right] \\
=&\mathbb{E}\left[\left\{U_i \mathscr{H}(T_i, t) - A_t \cdot \pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} \nabla_\theta g(T_i; \boldsymbol{\theta}^*)\right\}^2\right] \\
&+ \mathbb{E}\left[\left\{\mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} \cdot (\mathscr{H}(T_i, t) - A_t \cdot \nabla_\theta g(T_i; \boldsymbol{\theta}^*))\,|\boldsymbol{X}_i]\right\}^2\right] \\
&- \mathbb{E}\left[\left\{\pi_0(T_i, \boldsymbol{X}_i) \cdot \mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}|T_i, \boldsymbol{X_i}] \cdot (\mathscr{H}(T_i, t) - A_t \cdot \nabla_\theta g(T_i; \boldsymbol{\theta}^*))\right\}^2\right] \\
<&\mathbb{E}\left[\left\{U_i \mathscr{H}(T_i, t) - A_t \cdot \pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} \nabla_\theta g(T_i; \boldsymbol{\theta}^*)\right\}^2\right] = \Sigma_0(t, t),
\end{aligned}
$$

where the fourth equality holds by using the tower property of the conditional expectation, the inequality holds by using Jensen's inequality.

D Proof of Theorem 5

Proof. We prove parts (i) and (ii). The proof is similar to that for Theorem 1. Let

$$
g_N(t, \boldsymbol{\theta}) := g(t; \boldsymbol{\theta}) + \frac{\delta(t)}{\sqrt{N}} \text{ and } U_{iN} = \pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g_N(T_i; \boldsymbol{\theta}_N^*)\right\}.
$$

Obviously, $g_N(t, \boldsymbol{\theta}) \to g(t, \boldsymbol{\theta})$ and $U_{iN} \xrightarrow{a.s.} U_i$. Then

$$
\begin{aligned}
\widehat{U}_i =& \widehat{\pi}_K(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} \\
=& U_{iN} + \{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\} m\left\{Y_i; g_N(T_i; \boldsymbol{\theta}_N^*)\right\} \\
&+ \pi_0(T_i, \boldsymbol{X}_i)\left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\left\{Y_i; g_N(T_i; \boldsymbol{\theta}_N^*)\right\}\right] \\
&+ \{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\}\left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\left\{Y_i; g_N(T_i; \boldsymbol{\theta}_N^*)\right\}\right].
\end{aligned}
$$

Then, we have

$$
\widehat{J}_N(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \widehat{U}_i \mathscr{H}(T_i, t) = \frac{1}{\sqrt{N}} \sum_{i=1}^N U_{iN} \mathscr{H}(T_i, t) \tag{D.1}
$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}m\left\{Y_i;g_N(T_i;\boldsymbol{\theta}_N^*)\right\}\mathscr{H}(T_i,t) \tag{D.2}$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g_N(T_i;\boldsymbol{\theta}_N^*)\right\}\right]\mathscr{H}(T_i,t) \tag{D.3}$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g_N(T_i;\boldsymbol{\theta}_N^*)\right\}\right]\mathscr{H}(T_i,t). \tag{D.4}$$

Obviously, by Chebyshev's inequality, we have

$$(\text{D.1})=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}U_i\mathscr{H}(T_i,t)+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}(U_{iN}-U_i)\mathscr{H}(T_i,t)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}U_i\mathscr{H}(T_i,t)+o_P(1).$$

Using Proposition 7, under $H_L:\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g_N(T_i;\boldsymbol{\theta}_N^*)\}|T_i]=0$, we have

$$\begin{aligned}
(\text{D.2})=&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[m\left\{Y_i;g_N(T_i;\boldsymbol{\theta}_N^*)\right\}\cdot\mathscr{H}(T_i,t)|T_i,\boldsymbol{X_i}]\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g_N(T_i;\boldsymbol{\theta}_N^*)\right\}\cdot\mathscr{H}(T_i,t)|T_i]\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g_N(T_i;\boldsymbol{\theta}_N^*))\right\}\cdot\mathscr{H}(T_i,t)|\boldsymbol{X}_i]\\
&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g_N(T_i;\boldsymbol{\theta}_N^*)\right\}\cdot\mathscr{H}(T_i,t)]+o_P(1)\\
=&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[m\left\{Y_i;g_N(T_i;\boldsymbol{\theta}_N^*)\right\}\cdot\mathscr{H}(T_i,t)|T_i,\boldsymbol{X_i}]\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g_N(T_i;\boldsymbol{\theta}_N^*)\right\}\cdot\mathscr{H}(T_i,t)|\boldsymbol{X}_i]+o_P(1)\\
=&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\cdot\mathscr{H}(T_i,t)|T_i,\boldsymbol{X_i}]\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\cdot\mathscr{H}(T_i,t)|\boldsymbol{X}_i]+o_P(1)\\
=&-\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\phi(T_i,\boldsymbol{X}_i;t)+o_P(1),
\end{aligned}$$

where the third equality holds by using Chebyshev's inequality.

We consider the term (D.3). We first find the expression for $\sqrt{N}\{\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_N^*\}$. Note from Assumption 2 that

$$\frac{1}{\sqrt{N}}\sum_{i=1}^N \widehat{\pi}_K(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}\nabla_{\boldsymbol{\theta}}g(T_i;\widehat{\boldsymbol{\theta}})=o_P(1).$$

Note that

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^N \widehat{\pi}_K(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}\nabla_{\boldsymbol{\theta}}g(T_i;\widehat{\boldsymbol{\theta}})\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^N \pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}\nabla_{\boldsymbol{\theta}}g(T_i;\widehat{\boldsymbol{\theta}}) \qquad \text{(D.5)}\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^N \{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}\,m\left\{Y_i;g(T_i;\boldsymbol{\theta}_N^*)\right\}\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}_N^*) \qquad \text{(D.6)}\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^N \{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\} \qquad \text{(D.7)}\\
&\qquad\times\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}\nabla_{\boldsymbol{\theta}}g(T_i;\widehat{\boldsymbol{\theta}})-m\left\{Y_i;g(T_i;\boldsymbol{\theta}_N^*)\right\}\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}_N^*)\right].
\end{aligned}$$

For (D.6), Proposition 7, under H_L, we have

$$\begin{aligned}
&\frac{1}{\sqrt{N}}\sum_{i=1}^N \{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}\,m\left\{Y_i;g(T_i;\boldsymbol{\theta}_N^*\}\right)\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}_N^*)\\
=&-\frac{1}{\sqrt{N}}\sum_{i=1}^N \pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}_N^*)\right\}|T_i,\boldsymbol{X}_i\right]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}_N^*)\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot m\left\{Y_i;g(T_i;\boldsymbol{\theta}_N^*)\right\}|T_i\right]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}_N^*)\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot m\left\{Y_i;g(T_i;\boldsymbol{\theta}_N^*)\right\}\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}_N^*)|\boldsymbol{X}_i\right]\\
&-\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot m\left\{Y_i;g(T_i;\boldsymbol{\theta}_N^*)\right\}\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}_N^*)\right]+o_P(1)\\
=&-\frac{1}{\sqrt{N}}\sum_{i=1}^N \pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i\right]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\\
&+\frac{1}{\sqrt{N}}\sum_{i=1}^N \mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)|\boldsymbol{X}_i\right]+o_P(1),
\end{aligned}$$

where the last equality holds by using Chebyshev's inequality and the fact that $\lim_{N\to\infty} \boldsymbol{\theta}_N^* = \boldsymbol{\theta}^*$.

Recall the definition of $\boldsymbol{\theta}_N^*$ in section 4.3, we can see that $\|\boldsymbol{\theta}_N^* - \boldsymbol{\theta}^*\| = O_p(N^{-1/2})$. Then for (D.7), by using a similar argument used in establishing (B.13), we can obtain (D.7) $= o_P(1)$.

For (D.5), we have

$$
\begin{aligned}
&(\text{D.5}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} \pi_0(T_i,\boldsymbol{X}_i) m\left\{Y_i; g(T_i;\widehat{\boldsymbol{\theta}})\right\}\nabla_{\boldsymbol{\theta}} g(T_i;\widehat{\boldsymbol{\theta}})\\
=&\nu_N\left[\pi_0(T_i,\boldsymbol{X}_i) m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*)\right]\\
&+\left\{\nu_N\left(\pi_0(T_i,\boldsymbol{X}_i) m\left\{Y_i; g(T_i;\widehat{\boldsymbol{\theta}})\right\}\nabla_{\boldsymbol{\theta}} g(T_i;\widehat{\boldsymbol{\theta}})\right) - \nu_N\left(\pi_0(T_i,\boldsymbol{X}_i) m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*)\right)\right\}\\
&+\sqrt{N}\cdot\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i) m\left\{Y_i; g(T_i;\widehat{\boldsymbol{\theta}})\right\}\nabla_{\boldsymbol{\theta}} g(T_i;\widehat{\boldsymbol{\theta}})\right]\\
=&\nu_N\left[\pi_0(T_i,\boldsymbol{X}_i) m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*)\right] + o_P(1)\\
&+\sqrt{N}\cdot\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i) m\left\{Y_i; g(T_i;\widehat{\boldsymbol{\theta}})\right\}\nabla_{\boldsymbol{\theta}} g(T_i;\widehat{\boldsymbol{\theta}})\right],
\end{aligned}
$$

where the last equality holds because the empirical process

$$
\left\{\nu_N\left(\pi_0(T_i,\boldsymbol{X}_i) m\left\{Y_i; g(T_i;\boldsymbol{\theta})\right\}\nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta})\right) : \boldsymbol{\theta}\in\Theta\right\}
$$

is stochastically equicontinuous, and $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| \xrightarrow{p} 0$, then

$$
\nu_N\left(\pi_0(T_i,\boldsymbol{X}_i) m\left\{Y_i; g(T_i;\widehat{\boldsymbol{\theta}})\right\}\nabla_{\boldsymbol{\theta}} g(T_i;\widehat{\boldsymbol{\theta}})\right) - \nu_N\left(\pi_0(T_i,\boldsymbol{X}_i) m\left\{Y_i; g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}^*)\right) = o_P(1).
$$

Note that under H_L, $\boldsymbol{\theta}_N^* \to \boldsymbol{\theta}^*$, and $\widehat{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}^*$:

$$
\begin{aligned}
&\sqrt{N}\cdot\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i) m\left\{Y_i; g(T_i;\widehat{\boldsymbol{\theta}})\right\}\nabla_{\boldsymbol{\theta}} g(T_i;\widehat{\boldsymbol{\theta}})\right]\\
=&\sqrt{N}\cdot\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i) m\left\{Y_i; g(T_i;\boldsymbol{\theta}_N^*)\right\}\nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}_N^*)\right]\\
&+\sqrt{N}\cdot\nabla_{\boldsymbol{\theta}}\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i) m\left\{Y_i; g(T_i;\widetilde{\boldsymbol{\theta}}_N)\right\}\nabla_{\boldsymbol{\theta}} g(T_i;\widetilde{\boldsymbol{\theta}}_N)\right]\cdot\left\{\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N^*\right\}\\
=&\sqrt{N}\cdot\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}\left[m\left\{Y_i; g(T_i;\boldsymbol{\theta}_N^*)\right\}|T_i,\boldsymbol{X}_i\right]\cdot\nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}_N^*)\right]\\
&+\sqrt{N}\cdot\nabla_{\boldsymbol{\theta}}\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i) m\left\{Y_i; g(T_i;\widetilde{\boldsymbol{\theta}}_N)\right\}\nabla_{\boldsymbol{\theta}} g(T_i;\widetilde{\boldsymbol{\theta}}_N)\right]\cdot\left\{\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N^*\right\}\\
=&\sqrt{N}\cdot\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}\left[m\left\{Y_i; g_N(T_i;\boldsymbol{\theta}_N^*)\right\}|T_i,\boldsymbol{X}_i\right]\cdot\nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}_N^*)\right]\\
&-\sqrt{N}\cdot\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i;\widetilde{g}_N(T_i;\boldsymbol{\theta}_N^*)\right\}\middle|T_i,\boldsymbol{X}_i\right]\cdot\{g_N(T_i;\boldsymbol{\theta}_N^*) - g(T_i;\boldsymbol{\theta}_N^*)\}\nabla_{\boldsymbol{\theta}} g(T_i;\boldsymbol{\theta}_N^*)\right]\\
&+\nabla_{\boldsymbol{\theta}}\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i) m\left\{Y_i; g(T_i;\widetilde{\boldsymbol{\theta}}_N)\right\}\nabla_{\boldsymbol{\theta}} g(T_i;\widetilde{\boldsymbol{\theta}}_N)\right]\cdot\sqrt{N}\left\{\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N^*\right\}
\end{aligned}
$$

$$
\begin{aligned}
=& -\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\big|T_i,\boldsymbol{X}_i\right]\cdot\delta(T_i)\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\right]\\
&+\left\{\nabla_{\boldsymbol{\theta}}\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\right]+o_P(1)\right\}\cdot\sqrt{N}\left\{\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_N^*\right\}\\
=& -\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\big|T_i,\boldsymbol{X}_i\right]\cdot\delta(T_i)\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\right]\\
&+\Bigg\{\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i\right]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\nabla_{\boldsymbol{\theta}}^\top g(T_i;\boldsymbol{\theta}^*)\right]\\
&\qquad+\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_{\boldsymbol{\theta}}^2 g(T_i;\boldsymbol{\theta}^*)\right]+o_P(1)\Bigg\}\cdot\sqrt{N}\left\{\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_N^*\right\}\\
=& -\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\big|T_i,\boldsymbol{X}_i\right]\cdot\delta(T_i)\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\right]\\
&+\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i\right]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\nabla_{\boldsymbol{\theta}}^\top g(T_i;\boldsymbol{\theta}^*)\right]\cdot\sqrt{N}\left\{\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_N^*\right\}\\
&+o_P(1),
\end{aligned}
$$

where $\widetilde{\boldsymbol{\theta}}_N$ lies between $\boldsymbol{\theta}_N^*$ and $\widehat{\boldsymbol{\theta}}$, and $\widetilde{g}_N(T_i;\boldsymbol{\theta}_N^*)$ lies between $g_N(T_i;\boldsymbol{\theta}_N^*)$ and $g(T_i;\boldsymbol{\theta}_N^*)$. Hence, we get

$$
\begin{aligned}
&\sqrt{N}\left\{\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_N^*\right\}\\
=&\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i\right]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\nabla_{\boldsymbol{\theta}}^\top g(T_i;\boldsymbol{\theta}^*)\right]^{-1}\\
&\quad\times\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}_N^*))\right\}\big|T_i,\boldsymbol{X}_i\right]\cdot\delta(T_i)\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}_N^*)\right]\\
&\quad-\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i\right]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\nabla_{\boldsymbol{\theta}}^\top g(T_i;\boldsymbol{\theta}^*)\right]^{-1}\\
&\quad\times\frac{1}{\sqrt{N}}\sum_{i=1}^N\Bigg\{\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\\
&\qquad\qquad-\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i\right]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\\
&\qquad\qquad+\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)|\boldsymbol{X}_i\right]\Bigg\}+o_P(1).
\end{aligned}
$$

Then similar to (B.12), we have

$$
\text{(D.3)}=-\frac{1}{\sqrt{N}}\sum_{i=1}^N\psi(T_i,\boldsymbol{X}_i,Y_i;t)+\mu(t)+o_p(1),
$$

where

$$
\mu(t)=\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)^\top\mathscr{H}(T_i,t)\right]
$$

$$\times \mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)\nabla_\theta^\top g(T_i;\boldsymbol{\theta}^*)\right]^{-1}$$
$$\times \mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{d}{dg}\mathbb{E}\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta}_N^*)\right\}\Big|T_i,\boldsymbol{X}_i\right]\cdot\delta(T_i)\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}_N^*)\right].$$

Hence, we have

$$\begin{aligned}\widehat{J}_N(t)=&\frac{1}{\sqrt{N}}\sum_{i=1}^N\eta(T_i,\boldsymbol{X}_i,Y_i;t)+\mu(t)+o_P(1)\\=&\frac{1}{\sqrt{N}}\sum_{i=1}^N\left\{U_i\mathscr{H}(T_i,t)-\phi(T_i,\boldsymbol{X}_i;t)-\psi(T_i,\boldsymbol{X}_i,Y_i;t)\right\}+\mu(t)+o_P(1),\end{aligned}$$

where $\mathbb{E}\{\phi(T_i,\boldsymbol{X}_i;t)\}=0$ and $\mathbb{E}\{\psi(T_i,\boldsymbol{X}_i,Y_i;t)\}=0$. Therefore, under the null hypothesis H_0, $\widehat{J}_N(\cdot)$ weakly converges to $J_{\infty,\mu}(\cdot)$ in $L_2(\mathcal{T},dt)$, where $J_{\infty,\mu}(\cdot)$ is a Gaussian process with mean function $\mu(t)$ and covariance function given by

$$\Sigma(t,t')=\mathbb{E}\left\{\eta(T_i,\boldsymbol{X}_i,Y_i;t)\eta(T_i,\boldsymbol{X}_i,Y_i;t')\right\}.$$

We prove part (iii). Because

$$\frac{1}{\sqrt{N}}\widehat{J}_N(t)=\frac{1}{N}\sum_{i=1}^N\widehat{U}_i\mathscr{H}(T_i,t)$$

$$=\frac{1}{N}\sum_{i=1}^N U_i\mathscr{H}(T_i,t) \tag{D.8}$$

$$+\frac{1}{N}\sum_{i=1}^N\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\mathscr{H}(T_i,t) \tag{D.9}$$

$$+\frac{1}{N}\sum_{i=1}^N\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t) \tag{D.10}$$

$$+\frac{1}{N}\sum_{i=1}^N\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t). \tag{D.11}$$

By applying a similar argument for (B.2)-(B.4), we have that (D.9)-(D.11) are of $o_P(1)$. Under H_1, the law of large nubers implies (D.8) $=\mu_1(t)+o_P(1)$. Hence, we conclude the proof.

□

E Asymptotic properties of $\widehat{J}_N(t;\widehat{\theta}_{opt})$ and $\widehat{CM}_N(\widehat{\theta}_{opt})$

Theorem 8. *Suppose that $m(y;g)$ is differentiable with respect to g, Assumptions 1-5 and Assumptions 6-9 listed in Appendix A hold, then under H_0,*

$$(i)\quad \widehat{J}_N(t;\widehat{\boldsymbol{\theta}}_{opt}) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\eta_{opt}(T_i,\boldsymbol{X}_i,Y_i;t)+o_P(1),$$

$$(ii)\quad \widehat{J}_N(\cdot;\widehat{\boldsymbol{\theta}}_{opt}) \text{ converges weakly to } J_{\infty,opt}(\cdot) \text{ in } L_2\{\mathcal{T},dF_T(t)\},$$

where $J_{\infty,opt}$ is a Gaussian process with zero mean and covariance function given by

$$\Sigma_{opt}(t,t') = \mathbb{E}\left\{\eta_{opt}(T_i,\boldsymbol{X}_i,Y_i;t)\eta_{opt}(T_i,\boldsymbol{X}_i,Y_i;t')\right\}.$$

Furthermore,

$$(iii)\quad \widehat{CM}_N(\widehat{\boldsymbol{\theta}}_{opt}) \text{ converges to } \int\{J_{\infty,opt}(t)\}^2 dF_T(t) \text{ in distribution.}$$

Proof. We first claim $\|\widehat{\boldsymbol{\theta}}_{opt}-\boldsymbol{\theta}^*\| \xrightarrow{P} 0$ under H_0. Since

- Θ is compact;
- by Proposition 6, $|N^{-1}\cdot\widehat{CM}_N(\boldsymbol{\theta})-CM(\boldsymbol{\theta})| \xrightarrow{P} 0$ for every $\boldsymbol{\theta}\in\Theta$;
- $CM(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$;
- $|\widehat{U}_i(\boldsymbol{\theta})| = |\widehat{\pi}_K(T_i,\boldsymbol{X}_i)m(Y_i;g(T_i;\boldsymbol{\theta}))| \le O_p(1)\times\sup_{\boldsymbol{\theta}\in\Theta}|m(Y_i;g(T_i;\boldsymbol{\theta}))|$ and $\mathbb{E}[\sup_{\boldsymbol{\theta}\in\Theta} |m(Y_i;g(T_i;\boldsymbol{\theta}))|]<\infty$;

then it follows from van der Vaart (1998, Theorem 5.7) that $\|\widehat{\boldsymbol{\theta}}_{opt}-\boldsymbol{\theta}^*\| \xrightarrow{P} 0$.

We then find the asymptotic expression for $\sqrt{N}\{\widehat{\boldsymbol{\theta}}_{opt}-\boldsymbol{\theta}^*\}$. By the first order condition, we get

$$\frac{1}{N}\sum_{i=1}^{N}\widehat{J}_N(T_i;\widehat{\boldsymbol{\theta}}_{opt})\cdot\nabla_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\widehat{\boldsymbol{\theta}}_{opt})=0$$

Using the mean value theorem, we get

$$0=\frac{1}{N}\sum_{i=1}^{N}\widehat{J}_N(T_i;\boldsymbol{\theta}^*)\cdot\frac{\nabla_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\boldsymbol{\theta}^*)}{\sqrt{N}}$$
$$+\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\nabla_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\widetilde{\boldsymbol{\theta}}_{opt})}{\sqrt{N}}\cdot\frac{\nabla_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\widetilde{\boldsymbol{\theta}}_{opt})^\top}{\sqrt{N}}+\frac{\widehat{J}_N(T_i;\widetilde{\boldsymbol{\theta}}_{opt})}{\sqrt{N}}\cdot\frac{\nabla^2_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\widetilde{\boldsymbol{\theta}}_{opt})}{\sqrt{N}}\right\}\cdot\sqrt{N}\left\{\widehat{\boldsymbol{\theta}}_{opt}-\boldsymbol{\theta}^*\right\},$$

where $\widetilde{\boldsymbol{\theta}}_{opt}$ lies on the joining from $\widehat{\boldsymbol{\theta}}_{opt}$ to $\boldsymbol{\theta}^*$. Using the fact that $\|\widehat{\boldsymbol{\theta}}_{opt} - \boldsymbol{\theta}^*\| \xrightarrow{p} 0$ and Proposition 6, under H_0, it is easy to obtain

$$
\begin{aligned}
&\frac{1}{N}\sum_{i=1}^{N}\left\{\frac{\nabla_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\widetilde{\boldsymbol{\theta}}_{opt})}{\sqrt{N}}\cdot\frac{\nabla_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\widetilde{\boldsymbol{\theta}}_{opt})^\top}{\sqrt{N}}+\frac{\widehat{J}_N(T_i;\widetilde{\boldsymbol{\theta}}_{opt})}{\sqrt{N}}\cdot\frac{\nabla^2_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\widetilde{\boldsymbol{\theta}}_{opt})}{\sqrt{N}}\right\}\\
=&\int_{\mathcal{T}}\mathbb{E}\left[\pi_0(T,\boldsymbol{X})\cdot\frac{\partial}{g}m\{Y;g(T;\boldsymbol{\theta}^*)\}\nabla_{\boldsymbol{\theta}}g(T;\boldsymbol{\theta}^*)\mathcal{H}(T;t)\right]\\
&\quad\times\mathbb{E}\left[\pi_0(T,\boldsymbol{X})\cdot\frac{\partial}{g}m\{Y;g(T;\boldsymbol{\theta}^*)\}\nabla_{\boldsymbol{\theta}}g(T;\boldsymbol{\theta}^*)^\top\mathcal{H}(T;t)\right]f_T(t)dt+o_P(1)\\
=&\int_{\mathcal{T}}B_tB_t^\top f_T(t)dt+o_P(1),
\end{aligned}
$$

where

$$
B_t := \mathbb{E}\left[\pi_0(T,\boldsymbol{X})\cdot\frac{\partial}{g}m\{Y;g(T;\boldsymbol{\theta}^*)\}\nabla_{\boldsymbol{\theta}}g(T;\boldsymbol{\theta}^*)\mathcal{H}(T;t)\right].
$$

For $\widehat{J}_N(t;\boldsymbol{\theta}^*)$, under $H_0 : \mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}|T=t]=0$, by using Proposition 7, we get

$$
\begin{aligned}
&\widehat{J}_N(t;\boldsymbol{\theta}^*)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\widehat{\pi}_K(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}\mathcal{H}(T_i;t)\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\Big\{\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}\mathcal{H}(T_i;t)-\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}|T_i,\boldsymbol{X}_i]\cdot\mathcal{H}(T_i;t)\\
&\qquad+\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}\mathcal{H}(T_i;t)|\boldsymbol{X}_i]\Big\}+o_P(1)\\
=&\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\varphi_{opt}(T_i,\boldsymbol{X}_i,Y_i;t)+o_P(1),
\end{aligned}
$$

where

$$
\begin{aligned}
\varphi_{opt}(T_i,\boldsymbol{X}_i,Y_i;t):=&\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}\mathcal{H}(T_i;t)\\
&-\pi_0(T_i,\boldsymbol{X}_i)\cdot\mathbb{E}[m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}|T_i,\boldsymbol{X}_i]\cdot\mathcal{H}(T_i;t)\\
&+\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta}^*)\}\mathcal{H}(T_i;t)|\boldsymbol{X}_i]
\end{aligned}
$$

Now, we have

$$
\sqrt{N}\left\{\widehat{\boldsymbol{\theta}}_{opt}-\boldsymbol{\theta}^*\right\}
$$

$$
\begin{aligned}
&= -\left\{\int_{\mathcal{T}} B_t B_t^\top f_T(t)dt\right\}^{-1}\left\{\frac{1}{N}\sum_{i=1}^N \widehat{J}_N(T_i;\boldsymbol{\theta}^*)\cdot\frac{\nabla_{\boldsymbol{\theta}}\widehat{J}_N(T_i;\boldsymbol{\theta}^*)}{\sqrt{N}}\right\} \\
&= -\frac{1}{\sqrt{N}}\sum_{i=1}^N\left\{\int_{\mathcal{T}} B_t B_t^\top f_T(t)dt\right\}^{-1}\cdot\int_{\mathcal{T}}\varphi_{opt}(T_i,\boldsymbol{X}_i,Y_i;t)\cdot B_t\cdot f_T(t)dt.
\end{aligned}
$$

Let

$$
\psi_{opt}(T_i,\boldsymbol{X}_i,Y_i;t) = \left\{\int_{\mathcal{T}} B_t^\top f_T(t)dt\right\}\left\{\int_{\mathcal{T}} B_t B_t^\top f_T(t)dt\right\}^{-1}\int_{\mathcal{T}}\varphi_{opt}(T_i,\boldsymbol{X}_i,Y_i;t)B_t f_T(t)dt.
$$

Following a similar argument of establishing Theorem 1, we get

$$
\begin{aligned}
\widehat{J}_N(t;\widehat{\boldsymbol{\theta}}_{opt}) =& \frac{1}{\sqrt{N}}\sum_{i=1}^N \eta_{opt}(T_i,\boldsymbol{X}_i,Y_i;t) + o_P(1) \\
=& \frac{1}{\sqrt{N}}\sum_{i=1}^N \{U_i\mathscr{H}(T_i,t) - \phi(T_i,\boldsymbol{X}_i;t) - \psi_{opt}(T_i,\boldsymbol{X}_i,Y_i;t)\} + o_P(1).
\end{aligned}
$$

The remaining results follow by using a similar argument of estalishing Theorem 1. □