A Unified Framework for Specification Tests of Continuous Treatment Effect Models

Wei Huang*
School of Mathematics and Statistics and Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS), University of Melbourne, Australia
and
Oliver Linton†
Faculty of Economics, University of Cambridge
and
Zheng Zhang‡
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 unconfoundedness 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 $O(N^{-1/2})$. 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.

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

1 Introduction

Causal inference is a central topic in economics, statistics, and machine learning. Although a randomized trial is the gold standard for identifying causal effects, such trials are often unavailable or even unethical in practice. Observational data, which are collected when the participation of an intervention is only observed rather than manipulated by scientists, are predominantly the type of data that are available. A major challenge for inferring causality in observational studies is confoundedness, whereby 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 binary treatment, whereby an individual either does or does not receive the treatment (e.g., Hahn, 1998; Hirano et al., 2003; Donald et al., 2014; Imai and Ratkovic, 2014; Abrevaya et al., 2015; Chan et al., 2016; Athey et al., 2018; Hsu et al., 2020; Chen et al., 2020; Fan et al., 2020; Sant'Anna et al., 2020; Ai et al., 2022). Some studies focus on multivalued treatment (see, e.g., Cattaneo, 2010; Lee, 2018; Ai et al., 2020; Ao et al., 2021). However, in many applications, the treatment variable is continuously valued, and its causal effect is of great interest to decision makers. For example, when 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, when evaluating how advertising affects the campaign contributions for political analysis, the causal effect may depend not only on whether any advertisements are released but also on how many of them are distributed.

Estimation of continuous treatment effects has received considerable attention from researchers (see Hirano and Imbens, 2004; Galvao and Wang, 2015; Kennedy et al., 2017; Fong et al., 2018; Dong et al., 2019; Huber et al., 2020; Colangelo and Lee, 2020; Ai et al., 2021, among others). Hirano and Imbens (2004), Galvao and Wang (2015), and Fong et al. (2018) applied fully parametric methods by modeling 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 modeling and testing the relationship between 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 et al. (2017) and Huber et al. (2020) estimated the continuous treatment effects using the nonparametric kernel method. Although nonparametric approaches are much more flexible than parametric ones, the former 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 difficult to interpret nonparametric results.

In a recent article, Ai et al. (2021) studied 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, average and quantile treatment effects. They proposed a generalized weighting estimator for the causal effect with the weights modeled 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 et al. (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 et al. (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 et al. (2021) as the null model while testing our hypothesis. 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 on the specification tests for conditional models (e.g., Ait-Sahalia et al. (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 et al. (2003); Crump et al. (2008)). Most authors have considered the problems of testing a parametric/semiparametric null model using an integrated type test statistic. Ait-Sahalia et al. (2001) and Chen and Fan (1999) considered testing nonparametric/semiparametric null models using nonparametric kernel methods. Li et al. (2003) considered testing the nonparametric/semiparametric using series methods. Crump et al. (2008) derived a nonparametric Wald test statistic for testing the conditional average treatment effects under the unconfoundedness condition. For the binary treatment effect model, Shaikh et al. (2009) parametrically modeled the propensity score as a conditional expectation of the treatment given the confounders and proposed an associated specification test. This differs from our problem in the sense that we consider the specification test for the function of the potential outcome with a continuous treatment variable.

Specifically, we estimate our semiparametric weighted null model using the framework developed in Ai et al. (2021) and construct a Cramér–von Mises test statistic and a Kolmogorov–Smirnov test one to test the null model. 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 $O(N^{-1/2})$.

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 adopt a special case of the exchangeable bootstrap 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. All proofs are detailed in the supplementary file.

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 the *marginal structural model* (Robins et al., 2000) and the *average structural function* in nonseparable models (Blundell and Powell, 2003; Imbens and Newey, 2009). 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, which is also called the *quantile structural model* (Imbens and Newey, 2009). This can recover the quantile treatment effect (QTE), which is given by $\mathrm{QTE}(t_1, t_0) = F^{-1}_{Y^*(t_1)}(\tau) - F^{-1}_{Y^*(t_0)}(\tau)$. 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 an observational study in which 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 $\boldsymbol{\theta}^*$:

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

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 et al., 2003; Imai and van Dyk, 2004; Fong et al., 2018; Ai et al., 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 et al. (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 et al. (2000).

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

$$H_0 : \mathbb{P}\left(\mathbb{E}[\pi_0(T, \boldsymbol{X})m\{Y; g(T; \boldsymbol{\theta}^*)\}|T] = 0\right) = 1 \text{ for some } \boldsymbol{\theta}^* \in \Theta\,, \tag{2.2}$$

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 specification test for a univariate regression model, if both $\pi_0(T, \boldsymbol{X})$ and $\boldsymbol{\theta}^*$ were given. Specially, letting

$$U_i := \pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\,, \tag{2.3}$$

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

$$\mathbb{E}\left\{U_i\mathscr{H}(T_i, t)\right\} = 0 \text{ for all } t \in \mathcal{T}\,. \tag{2.4}$$

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 $\mathscr{H}(T_i, t) = \mathbb{1}(T_i \leq t)$ (see Stinchcombe and White, 1998 and Stute, 1997 for more detailed discussion). Now, letting

$$J_N^0(t) = \frac{1}{\sqrt{N}}\sum_{i=1}^{N} U_i\mathscr{H}(T_i, t)\,, \tag{2.5}$$

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

$$CM_N^0 = \int\{J_N^0(t)\}^2\widehat{F}_T(dt) = \frac{1}{N}\sum_{i=1}^{N}\{J_N^0(T_i)\}^2, \tag{2.6}$$

or the Kolmogorov-Smirnov (KS)-type statistic

$$KS_N^0 = \sup_{t \in \mathcal{T}} |J_N^0(t)|, \tag{2.7}$$

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 must replace the U_i's with some estimates, which is studied in the following section.

Remark 1. *Note that in our model, the stabilized weights $\pi_0(T, \boldsymbol{X})$ are nonparametric. The conditional distribution of the treatment, given the confounders $f_{T|X}(T|\boldsymbol{X})$, which is known as the* generalized propensity score *(Hirano and Imbens, 2004), is also nonparametric. Under the unconfoundedness assumption, an alternative identification of the null model is through the conditional distribution of the outcome given the treatment and the confounders and the marginal distribution of the confounders, that is,* $\mathbb{E}[m\{Y^*(t); g(t; \boldsymbol{\theta})\}] = \mathbb{E}\Big(\mathbb{E}[m\{Y; g(T; \boldsymbol{\theta})\}|\boldsymbol{X}, T = t]\Big)$ *holds under Assumption 1.*

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, Ai et al. (2021) directly estimated $\pi_0(T, \boldsymbol{X})$ as a whole using the generalized empirical likelihood (GEL). We adopt their estimator and elaborate its construction as follows. 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 et al. (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 et al. (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 et al. (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 immediately clear. 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 et al. (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 $\{\widehat{\pi}_K(T_i, \boldsymbol{X}_i)\}_{i=1}^N$ satisfies 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 propose an extremum estimator for $\boldsymbol{\theta}^*$ (e.g., Pakes and Pollard, 1989, Chen et al., 2003, de Castro et al., 2019). Note that under H_0, the true value $\boldsymbol{\theta}^*$ solves the following equation:

$$\mathbb{E}\left[\pi_0(T, \boldsymbol{X}) m\{Y; g(T; \boldsymbol{\theta})\} w(T; \boldsymbol{\theta})\right] = 0, \tag{3.6}$$

where $w(T; \boldsymbol{\theta})$ (which may possibly not involve $\boldsymbol{\theta}$) is a prespecified q-dimensional vector with $q \geq p$ such that, under H_0, $\boldsymbol{\theta}^*$ is identified or over-identified. Examples of such vectors include $w(T; \boldsymbol{\theta}) = (1, T, ..., T^{q-1})^\top$ or $w(T; \boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} g(T; \boldsymbol{\theta})$, where "$\nabla_{\boldsymbol{\theta}}$" denotes the derivative with respect to $\boldsymbol{\theta}$. We then estimate $\boldsymbol{\theta}^*$ by

$$\widehat{\boldsymbol{\theta}} := \arg\min_{\boldsymbol{\theta} \in \Theta} \|M_N(\boldsymbol{\theta}, \widehat{\pi}_K)\|, \tag{3.7}$$

where $\|\cdot\|$ is the Euclidean norm, and

$$M_N(\boldsymbol{\theta}, \pi) := \frac{1}{N} \sum_{i=1}^N \pi(T_i, \boldsymbol{X}_i) m\{Y_i; g(T_i; \boldsymbol{\theta})\} w(T_i; \boldsymbol{\theta}).$$

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 (2.5) by the $\widehat{U}_i$'s, we have a feasible test statistic for H_0 based on

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

the corresponding estimators of the Cramér–von Mises (CM)-type statistic in (2.6) and the Kolmogorov-Smirnov (KS)-type statistic in (2.7) are, respectively,

$$\widehat{CM}_N = \frac{1}{N} \sum_{i=1}^N \{\widehat{J}_N(T_i)\}^2 \quad \text{and} \quad \widehat{KS}_N = \sup_{t \in \mathcal{T}} |\widehat{J}_N(t)|, \tag{3.8}$$

where the supremum is calculated as the maximum value over a discretization of $\mathcal{T}$ in practice.

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 to 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

$$\widehat{\boldsymbol{\theta}}_{opt} := \arg\min_{\boldsymbol{\theta}\in\Theta} \widehat{CM}_N(\boldsymbol{\theta}) := \arg\min_{\boldsymbol{\theta}\in\Theta} \frac{1}{N}\sum_{i=1}^N \{\widehat{J}_N(T_i; \boldsymbol{\theta})\}^2. \tag{3.9}$$

Therefore, 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, and even 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, a unique solution to the problem does not exist. 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 section E in the supplementary file.

Remark 3. *In order to estimate $\pi_0(T, \boldsymbol{X})$, Fong et al. (2018) noted the moment conditions*

$$\mathbb{E}\left[\pi_0(T, \boldsymbol{X})T\boldsymbol{X}\right] = \mathbb{E}(T)\mathbb{E}(\boldsymbol{X}), \quad \mathbb{E}\left[\pi_0(T, \boldsymbol{X})T\right] = \mathbb{E}(T), \quad \mathbb{E}\left[\pi_0(T, \boldsymbol{X})\boldsymbol{X}\right] = \mathbb{E}(\boldsymbol{X}), \tag{3.10}$$

which are special cases of our moment condition (3.2). They then proposed estimating $\pi_0(T, \boldsymbol{X})$ by maximizing the empirical likelihood of T and $\boldsymbol{X}$ under the constraints of the sample analogue of (3.10) and estimating $\mathbb{E}\{Y^*(t)\}$ by a simple linear model. This can be considered as fixing $u_{K_1}(T) = (1, T)^\top$ and $v_{K_2}(\boldsymbol{X}) = (1, \boldsymbol{X}^\top)^\top$, taking $m\{Y^*(t), g(t, \boldsymbol{\theta}^*)\} = Y^*(t) - g(t, \boldsymbol{\theta}^*)$ and g as a simple linear model in the estimation method of Ai et al. (2021). However, the equation (3.10) is of finite dimension and cannot nonparametrically identify $\pi_0(T, \boldsymbol{X})$. Hence, Fong et al. (2018) imposed a parametric model for the stabilized weights to achieve consistent estimation. We adopt the estimator proposed by Ai et al. (2021) that does not impose any parametric structure on the stabilized weights.

Remark 4. *Once our specification test rejects the null model and no better parametric* *model can be proposed, several solutions are available. For example, one may consider the double robustness estimator. Colangelo and Lee (2020) estimate the average dose-response function* $\mathbb{E}[Y^*(t)]$ *based on the following double robustness representation:*

$$\mathbb{E}[Y^*(t)] = \mathbb{E}\left\{\mathbb{E}\left[Y|T=t,\boldsymbol{X}\right] + \lim_{h\to 0}\mathbb{E}\left[\frac{K_h\left(T-t\right)}{f_{T|X}(t|\boldsymbol{X})}\left\{Y-\mathbb{E}\left[Y|T=t,\boldsymbol{X}\right]\right\}\right]\right\},$$

where $K_h\left(T-t\right)$ *is a kernel weighting observation* T *with treatment value of approximately* t *in a distance of* h*. They estimate both the general propensity score* $f_{T|X}(t|\boldsymbol{X})$ *and the outcome regression function* $\mathbb{E}\left[Y|T=t,\boldsymbol{X}\right]$ *using nonparametric techniques with cross-fitting. An alternative solution to the misspecification of the dose-response function* $g(t;\boldsymbol{\theta})$ *is to consider a fully nonparametric specification* $g(t)$*, that is, to estimate* $g(t)$ *from the moment* $\mathbb{E}[m(Y^*(t);g(t))]=0$ *for all* $t\in\mathcal{T}$*. Under Assumption 1,* $g(t)$ *can be identified through the conditional moment* $\mathbb{E}[\pi_0(T,\boldsymbol{X})m(Y;g(T))|T]=0$*. We can define the sieve minimum distance (SMD) estimator (Ai and Chen, 2003) of* $g(T)$ *by*

$$\widehat{g}(\cdot) := \arg\min_{h(\cdot)\in\mathcal{H}_{K_3}}\frac{1}{N}\sum_{i=1}^{N}\left\{\widehat{E}[\widehat{\pi}_K(T,\boldsymbol{X})m(Y;h(T))|T=T_i]\right\}^2\Sigma^{-1}(T_i),$$

where $\Sigma(T_i)$ *is a user-specified weighting function, and*

$$\widehat{E}[\widehat{\pi}_K(T,\boldsymbol{X})m(Y;h(T))|T] := \left[\sum_{i=1}^{N}\widehat{\pi}_K(T_i,\boldsymbol{X})m(Y_i;h(T_i))u_{K_3}^{\top}(T_i)\right] \times\left[\sum_{i=1}^{N}u_{K_3}(T_i)u_{K_3}^{\top}(T_i)\right]^{-1}u_{K_3}(T),$$

and $\mathcal{H}_{K_3}:=\{\lambda^{\top}u_{K_3}(T):\lambda\in\mathbb{R}^{K_3}\}$ *is a linear sieve space. Ai et al. (2021, Theorem 6) established the large sample property for the nonparametric estimator of the average dose-response function* $\widehat{E}[\widehat{\pi}_K(T,\boldsymbol{X})Y|T=t]$*.*

The extension of these methods to the general dose-response function including the quantile dose-response and the development of the corresponding large sample property is beyond the scope of this paper.

Remark 5*. If the residual function* $m(y;g(t;\boldsymbol{\theta}))$ *is smooth in* (t,y)*, the sieve minimum distance (SMD) estimator of* $\boldsymbol{\theta}^*$ *developed by Ai and Chen (2003) is semiparametrically efficient with respect to the conditional model* $\mathbb{E}[\pi_0(T,\boldsymbol{X})m\{T;g(T;\boldsymbol{\theta}^*)\}|T]=0$*. This efficient estimation result can also be achieved based on our unconditional moment (3.6); indeed, by* *replacing $\pi_0(T,\boldsymbol{X})$ with its estimate $\widehat{\pi}_K(T,\boldsymbol{X})$ and setting $w(T;\boldsymbol{\theta})$ to be a q-dimensional sieve basis, for example, $w(T;\boldsymbol{\theta})=(1,T,...,T^{q-1})^\top$, with $q\to\infty$ at an appropriate rate, it can be shown that the generalized method of moments (GMM) (Hansen, 1982) estimator of $\boldsymbol{\theta}^*$ constructed from (3.6) is asymptotically equivalent to the SMD estimator (see Ai et al. (2020) for an analogous finding). However, if the residual function $m(y;g(t;\boldsymbol{\theta}))$ is non-smooth, it remains an open problem regarding whether the efficient estimation of $\boldsymbol{\theta}^*$ from the conditional model $\mathbb{E}[\pi_0(T,\boldsymbol{X})m\{T;g(T;\boldsymbol{\theta}^*)\}|T]=0$ can be established. For this reason, and to avoid introducing an extra tuning parameter q, we estimate $\boldsymbol{\theta}^*$ through (3.6) with $w(T;\boldsymbol{\theta})$ as a fixed vector.*

4 Large sample properties

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

4.1 Asymptotic properties under null hypothesis

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

Assumption 2. *Under H_0, (i) $\boldsymbol{\theta}^*$ is an interior point of Θ, where Θ is a compact set in $\mathbb{R}^p$; (ii) $\|M_N(\widehat{\boldsymbol{\theta}},\widehat{\pi}_K)\|=\inf_{\boldsymbol{\theta}\in\Theta_\delta}\|M_N(\boldsymbol{\theta},\widehat{\pi}_K)\|+o_P(N^{-1/2})$, where $\Theta_\delta:=\{\boldsymbol{\theta}\in\Theta:\|\boldsymbol{\theta}-\boldsymbol{\theta}^*\|\le\delta\}$.*

Assumption 3. *Let $\eta(T,\boldsymbol{X},Y;t)$ be defined in (4.2), $Var\{\eta(T,\boldsymbol{X},Y;t)\}<\infty$ for all $t\in\mathcal{T}$.*

Assumption 4.

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

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:*

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

for any $\boldsymbol{\theta}\in\Theta$ *and any small* $\delta>0$ *and for some finite positive constant* C*.*

Assumption 2 is essentially stating that the estimating equation is a.s. approximately satisfied; see Pakes and Pollard (1989) and Chen et al. (2003). 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 to establish the weak convergence of our test statistic; see Andrews (1994). Again, this is satisfied by widely used residual 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

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

$$+ \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) w(T_i; \boldsymbol{\theta}^*) m \{Y_i; g(T_i; \boldsymbol{\theta}^*)\} | \boldsymbol{X}_i] \Big\}, \tag{4.1}$$

and

$$\eta(T_i, \boldsymbol{X}_i, Y_i; t) := U_i \mathscr{H}(T_i, t) - \phi(T_i, \boldsymbol{X}_i; t) - \psi(T_i, \boldsymbol{X}_i, Y_i; t). \tag{4.2}$$

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 A.1-A.4 listed in section A of the supplementary file hold; then, under H_0,*

$$(i) \quad \widehat{J}_N(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \eta(T_i, \boldsymbol{X}_i, Y_i; t) + o_P(1) \text{ holds uniformly over } t \in \mathcal{T},$$

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

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

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

Furthermore,

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

$$(iv) \quad \widehat{KS}_N \text{ converges to } \sup_{t \in \mathcal{T}} |J_\infty(t)| \text{ in distribution.}$$

The proof of Theorem 1 is relegated to section B in the supplementary file. 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 effect of the vector $w(T; \boldsymbol{\theta})$ on the asymptotic property of our test statistic is reflected in the term $\psi(T_i, \boldsymbol{X}_i, Y_i, t)$. It is unclear which choice of $w(T; \boldsymbol{\theta})$ would minimize the variance of $\Sigma(t, t)$. A common choice is $w(T; \boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} g(T; \boldsymbol{\theta})$. Then, the second and fourth terms of $\psi(T_i, \boldsymbol{X}_i, Y_i, t)$ in (4.1) are canceled out, which also simplifies the calculation approximating the null limiting distribution in practice.

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 by minimizing the following criterion function:

$$\widehat{\boldsymbol{\theta}}_0 = \arg\min_{\boldsymbol{\theta}\in\Theta} \|M_N(\boldsymbol{\theta}, \pi_0)\|.$$

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

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

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) $\widehat{J}_0(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \eta_0(T_i, \boldsymbol{X}_i, Y_i; t) + o_P(1)$ *holds uniformly over* $t \in \mathcal{T}$,

(ii) $\widehat{J}_0(\cdot)$ *converges weakly to* $J_{0,\infty}(\cdot)$ *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 section C in the supplementary file. In the estimation of the average treatment effects with binary and multiple treatments, it is a well-known paradox that using a nonparametric estimated propensity score is more efficient than using the true one; see Hirano et al. (2003), Chan et al. (2016), and Lee (2018) among others. Theorem 2 shows that this is also the case for continuous treatments.

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,

$$H_0 : \exists \text{ some } \boldsymbol{\theta}^* \in \Theta \subset \mathbb{R}^p, \text{ s.t. } \mathbb{E}\{Y^*(t)\} = g(t; \boldsymbol{\theta}^*) \text{ for all } t \in \mathcal{T}, \tag{4.3}$$

against the alternative hypothesis

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

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

$$\widehat{CM}_N^{ADRF} = \frac{1}{N}\sum_{i=1}^N \{\widehat{J}_N^{ADRF}(T_i)\}^2 \text{ and } \widehat{KS}_N^{ADRF} = \sup_{t \in \mathcal{T}} |\widehat{J}_N^{ADRF}(t)|,$$

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

$$\phi^{ADRF}(T_i, \boldsymbol{X}_i; t) := \pi_0(T_i, \boldsymbol{X}_i) \cdot \mathscr{H}(T_i, t) \cdot \mathbb{E}\{Y_i - g(T_i; \boldsymbol{\theta}^*)|T_i, \boldsymbol{X}_{\boldsymbol{i}}\}$$

$$- \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)\{Y_i - g(T_i;\boldsymbol{\theta}^*)\} \cdot \mathscr{H}(T_i,t)|\boldsymbol{X}_i],$$

and

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

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 A.1-A.4 listed in section A of the supplementary file hold; then, under H_0,*

$$(i) \quad \widehat{J}_N^{ADRF}(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^N \eta^{ADRF}(T_i, \boldsymbol{X}_i, Y_i; t) + o_P(1) \text{ holds uniformly over } t \in \mathcal{T},$$

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

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,

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

$$(iv) \quad \widehat{KS}_N^{ADRF} \text{ converges to } \sup_{t\in\mathcal{T}} \left|J_\infty^{ADRF}(t)\right| \text{ in distribution.}$$

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

$$H_0 : \exists \text{ some } \boldsymbol{\theta}^* \in \Theta \subset \mathbb{R}^p, \text{ s.t. } F_{Y^*(t)}^{-1}(\tau) = g(t;\boldsymbol{\theta}^*) \text{ for all } t \in \mathcal{T}, \tag{4.4}$$

against the alternative hypothesis

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

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

$$\widehat{CM}_N^{QDRF} = \frac{1}{N}\sum_{i=1}^N [\widehat{J}_N^{QDRF}(T_i)]^2 \text{ and } \widehat{KS}_N^{QDRF} = \sup_{t\in\mathcal{T}}\left|\widehat{J}_N^{QDRF}(t)\right|,$$

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

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

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 A.1-A.4 listed in section A of the supplementary file hold; then, under* H_0,

$$(i) \quad \widehat{J}_N^{QDRF}(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \eta^{QDRF}(T_i, \boldsymbol{X}_i, Y_i; t) + o_P(1) \text{ holds uniformly over } t \in \mathcal{T},$$

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

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,

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

$$(iv) \quad \widehat{KS}_N^{QDRF} \text{ converges to } \sup_{t \in \mathcal{T}} \left| J_\infty^{QDRF}(t) \right| \text{ in distribution.}$$

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 $O(N^{-1/2})$. Let $\boldsymbol{\theta}^*$ be the limit of $\boldsymbol{\theta}_N^*$ as $N \to \infty$, hence it solves the following equation:

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

Define

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

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

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

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

$$\times \, \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) \cdot \frac{\partial}{\partial g} \mathbb{E}\left[m\left\{ Y_i; g(T_i; \boldsymbol{\theta}_N^*)) \right\} \middle| T_i, \boldsymbol{X}_i \right] \cdot \delta(T_i) \cdot w(T_i; \boldsymbol{\theta}^*) \right].$$

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 A.1-A.4 listed in section A of the supplementary file hold. Under the local alternative hypothesis H_L,*

$$(i) \quad \widehat{J}_N(t) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} \eta(T_i, \boldsymbol{X}_i, Y_i; t) + \mu(t) + o_P(1) \text{ holds uniformly over } t \in \mathcal{T}, \tag{4.5}$$

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

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 $O(N^{-1/2})$.

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 et al., 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 et al., 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 et al., 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, then $\widehat{J}^*_{N,b}(\cdot)$ has the same asymptotic behavior as $\widehat{J}_N(\cdot)$ for $b = 1, \ldots, B$. Then, we can approximate the limiting distributions of $\widehat{CM}_N$ and $\widehat{KS}_N$ under H_0, respectively, by

$$\widehat{CM}^*_{N,b} = \frac{1}{N}\sum_{i=1}^{N}\left\{\widehat{J}^*_{N,b}(T_i)\right\}^2 \quad \text{and} \quad \widehat{KS}^*_{N,b} = \sup_{t\in\mathcal{T}}\left|\widehat{J}^*_{N,b}(t)\right|,$$

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

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 benefit from 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

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

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

$$\widehat{\boldsymbol{\theta}}^{(-j)} = \arg\min_{\boldsymbol{\theta} \in \Theta} \left\| M_N^{(-j)} \left(\boldsymbol{\theta}, \widehat{\pi}_K^{(-j)} \right) \right\| ,$$

where

$$M_N^{(-j)} \left(\boldsymbol{\theta}, \widehat{\pi}_K^{(-j)} \right) := \frac{1}{N} \sum_{i \notin S_j} \widehat{\pi}_K^{(-j)}(T_i, \boldsymbol{X}_i) m\{Y_i; g(T_i, \boldsymbol{\theta})\} w(T_i; \boldsymbol{\theta}) ,$$

with $\widehat{\pi}_K^{(-j)}(T_i, \boldsymbol{X}_i)$ obtained in a method identical to that introduced in Section 2 via (3.4) and (3.5), but excluding samples in S_j.

6.2 Simulation study

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

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

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

DGP1-L $T = 1 + 0.2X + \xi$, and $Y = 1 + X + 0.1T^3 + \epsilon$,

DGP1-NL $T = 0.1X^2 + \xi$, and $Y = X^2 + 0.2T^3 + \epsilon$, 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

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

We take the vector $w(T;\boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} g(T;\boldsymbol{\theta})$ and adopt the algorithm of de Castro et al. (2019) for estimating the quantile dose-response function to overcome the computational difficulties with the discontinuity of the indicator function.

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, namely logistic, cosine-sine, and indicator functions. Specifically, for the logistic weight function, we took the constant $c = 5$. We tested all models using both CM-type and KS-type statistics. The results of the two methods are similar; here, we present those of the CM-type statistic. Results of the KS-type one can be found in the supplementary materials of this paper.

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 observe that all tests become more and more powerful as N or significance level increases and reach a considerably high power level when $N = 200$.

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.

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	0.020	0.076	0.131	0.016	0.071	0.138	0.014	0.055	0.115
		200	0.015	0.050	0.105	0.016	0.056	0.118	0.011	0.057	0.114
		500	0.011	0.051	0.109	0.012	0.050	0.101	0.014	0.049	0.111
	DGP0-NL	100	0.033	0.094	0.172	0.016	0.080	0.134	0.018	0.061	0.124
		200	0.024	0.079	0.138	0.010	0.060	0.117	0.010	0.062	0.120
		500	0.024	0.067	0.129	0.011	0.056	0.101	0.011	0.052	0.118
Median	**DGP0-L**	100	0.036	0.117	0.171	0.028	0.090	0.141	0.035	0.100	0.178
		200	0.016	0.076	0.144	0.017	0.066	0.124	0.018	0.068	0.135
		500	0.010	0.047	0.104	0.009	0.063	0.117	0.012	0.063	0.120
	DGP0-NL	100	0.044	0.135	0.217	0.014	0.072	0.139	0.022	0.083	0.168
		200	0.026	0.101	0.160	0.015	0.060	0.113	0.016	0.066	0.141
		500	0.020	0.078	0.130	0.010	0.054	0.111	0.013	0.061	0.119

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 et al. (2021). The data have been analyzed several times in the treatment effect literature (Urban and Niebler, 2014; Fong et al., 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 across $N = 16265$ zip codes.

The data were 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 et al. (2021) considered the treatment variable (number of political advertisements) as continuous and assumed that

$$\mathbb{E}\{Y^*(t)\} = \beta_1 + \beta_2 t + \beta_3 t^2\,, \tag{6.2}$$

Table 2: Estimated power

$m(\cdot)$	Model	N	Logistic 1%	Logistic 5%	Logistic 10%	Cosine-Sine 1%	Cosine-Sine 5%	Cosine-Sine 10%	Indicator 1%	Indicator 5%	Indicator 10%
Average	**DGP1-L**	100	0.829	0.931	0.963	0.566	0.797	0.868	0.697	0.848	0.895
		200	0.987	0.998	0.999	0.918	0.980	0.994	0.969	0.996	0.998
	DGP1-NL	100	0.543	0.739	0.838	0.531	0.743	0.835	0.458	0.669	0.767
		200	0.854	0.945	0.972	0.865	0.950	0.971	0.835	0.925	0.953
Median	**DGP1-L**	100	0.529	0.744	0.832	0.259	0.517	0.656	0.391	0.633	0.743
		200	0.870	0.957	0.984	0.580	0.818	0.891	0.760	0.903	0.946
	DGP1-NL	100	0.250	0.476	0.619	0.207	0.438	0.590	0.196	0.407	0.542
		200	0.545	0.769	0.876	0.511	0.762	0.849	0.471	0.710	0.809

where the observed outcome $Y^*(T) = \log(\text{Contribution} + 1)$ and $T = \log(\#\text{ads} + 1)$, where #ads denotes the number of advertisements. 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 et al. (2018) for more details. Ai et al. (2021) found 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 et al. (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. Indeed, when we applied our method with logistic, cosine-sine, and indicator weight functions with a $B = 500$ simulation-based approximation to test the model in (6.2), all the methods rejected the model with

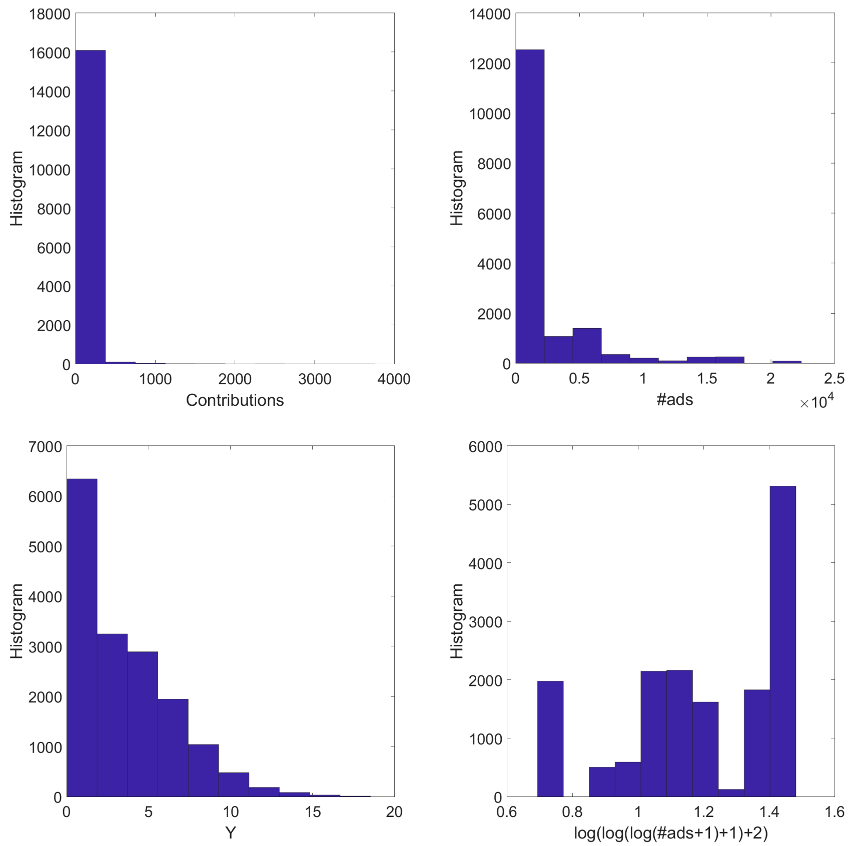

Figure 1: The histogram of the original campaign contribution data (top left) and the Box-Cox transformed contributions defined at (6.3) (bottom left), the histogram of the original counts of advertisements data (top right), and that of the transformed ones (bottom right).

the p-values equal to 0.

We examined the histogram of the original campaign contribution data and the number of advertisements T. From the first row of Figure 1, we can see that both histograms are highly right-skewed. That is, they are not likely to fit any linear models. We then conducted a log-transformation, as in Ai et al. (2021). However, the results were similar. To make the data more likely to fit a linear model, we searched across Box-Cox transformations of the response data of the form $\text{BoxCox}(\text{Contribution}, \lambda_1, \lambda_2) := \{(\text{Contribution} + \lambda_2)^{\lambda_1} - 1\}/\lambda_1$ w.r.t. λ_1, λ_2 to find a transformation of the contribution whose sample quantiles have the largest correlation with those of a standard normal distribution. This yielded $(\tilde{\lambda}_1, \tilde{\lambda}_2) =$

(0.1397, 0.0176). We then take

$$Y = \text{BoxCox}(\text{Contribution}, \tilde{\lambda}_1, \tilde{\lambda}_2) - \min\left\{\text{BoxCox}(\text{Contribution}, \tilde{\lambda}_1, \tilde{\lambda}_2)\right\}, \qquad (6.3)$$

so that the minimum response data is 0. The histogram of Y is shown in the bottom left of Figure 1. We can see that the transformed data remain highly right-skewed. However, now, it appears to be a truncated normal distribution.

Table 3: Estimated p-values from the U.S. presidential campaign data

Statistic	Logistic	Cosine-Sine	Indicator
CM	0.610	0.430	0.740
KS	0.718	0.667	0.718

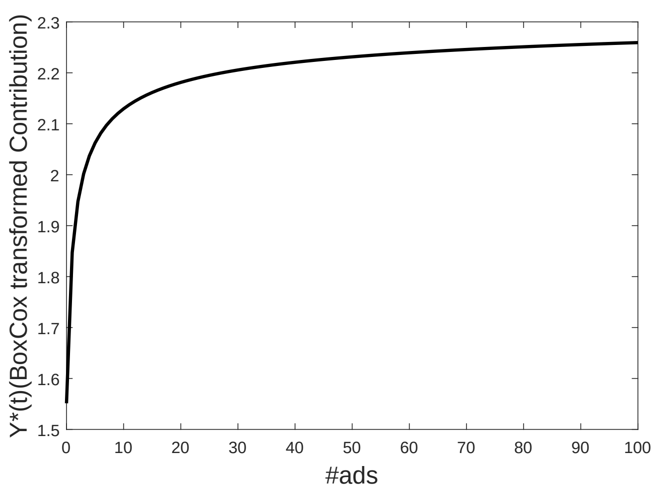

Figure 2: The plot of the estimated Tobit Model of the BoxCox transformed Campaign contribution versus the number of advertisements distributed.

Now, it seems more reasonable to assume a Tobit model for the data. Specifically, let

$$Y(t) = \boldsymbol{\beta}^\top \boldsymbol{t} + \epsilon\,,$$

for some unknown parameter $\boldsymbol{\beta}$ in a compact set in $\mathbb{R}^p$, where $\boldsymbol{t} = (1, t, t^2, \ldots, t^{p-1})^\top$ and ϵ is a normal random variable with mean 0 and unknown variance σ^2. We test a Tobit linear model on the potential outcome:

$$Y^*(t) = \begin{cases} Y(t) & \text{if} \quad Y(t) > 0, \\ 0 & \text{if} \quad Y(t) \leq 0\,. \end{cases}$$

The details of the estimation of this model and the test statistics can be found in the supplementary material of this paper. We then tested the Tobit model with several different transformations of the treatment data #ads and the polynomial order p. We found the Tobit model with $T = \log(\log(\log(\#\text{ads} + 1) + 1) + 2)$ and $p = 5$ gives the most reasonable results. The corresponding p-values are shown in Table 3, and the estimated model is depicted in Figure 2. It indicates that the campaign contribution increases rapidly with a relatively small increase in the number of advertisements from 0, and then the improvement gradually becomes marginal.

SUPPLEMENTARY MATERIAL

Supplementary materials are only for online publication. The supplementary file contains the simulation results of the KS-type statistic, the details of the estimation and the test statistics for the Tobit model used in section 6.3, the proofs of Theorems 1, 2, 5, and the asymptotic properties of $\widehat{J}_N(t; \widehat{\boldsymbol{\theta}}_{opt})$ and $\widehat{CM}_N(\widehat{\boldsymbol{\theta}}_{opt})$.

Acknowledgments

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 final 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. and X. Chen (2003). Efficient estimation of models with conditional moment restrictions containing unknown functions. *Econometrica 71*(6), 1795–1843.

Ai, C., L. Huang, and Z. Zhang (2020). A mann–whitney test of distributional effects in a multivalued treatment. *Journal of Statistical Planning and Inference 209*, 85–100.

Ai, C., L. Huang, and Z. Zhang (2022). A simple and efficient estimation of average treatment effects in models with unmeasured confounders. *Statistica Sinica 32*(3).

Ai, C., O. Linton, K. Motegi, and Z. Zhang (2021). A unified framework for efficient estimation of general treatment models. *Quantitative Economics 12*(3), 779–816.

Ai, C., O. Linton, and Z. Zhang (2020). A simple and efficient estimation method for models with nonignorable missing data. *Statistica Sinica 30*, 1949–1970.

Ai, C., O. Linton, and Z. Zhang (2021). Estimation and inference for the counterfactual distribution and quantile functions in continuous treatment models. *Journal of Econometrics*.

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 R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 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.

Bierens, H. J. (1990). A consistent conditional moment test of functional form. *Econometrica: Journal of the Econometric Society*, 1443–1458.

Bierens, H. J. and W. Ploberger (1997). Asymptotic theory of integrated conditional moment tests. *Econometrica: Journal of the Econometric Society*, 1129–1151.

Blundell, R. and J. L. Powell (2003). Endogeneity in nonparametric and semiparametric regression models. *8*.

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 non-parametric 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, X., O. Linton, and I. Van Keilegom (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica 71*(5), 1591–1608.

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.

de Castro, L., A. F. Galvao, D. M. Kaplan, and X. Liu (2019). Smoothed gmm for quantile models. *Journal of Econometrics 213*(1), 121–144.

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*, 1–15.

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

Fan, Y. and Q. Li (2000). Consistent model specification tests: Kernel-based tests versus bierens' icm tests. *Econometric Theory*, 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.

Hansen, L. (1982). Large sample properties of generalized method of moments estimators. *Econometrica 50*, 1029–1054.

Hirano, K. and G. W. Imbens (2004). The propensity score with continuous treatments. In A. Gelman and X.-L. Meng (Eds.), *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, Chapter 7, pp. 73–84. John Wiley & Sons Ltd.

Hirano, K., G. W. Imbens, and G. Ridder (2003, July). 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*, 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 M. Ratkovic (2014). Covariate balancing propensity score. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 76*(1), 243–263.

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 W. K. Newey (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica 77*(5), 1481–1512.

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, July). 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.

Sant'Anna, P. H., X. Song, and Q. Xu (2020). Covariate distribution balance via propensity scores. *arXiv preprint arXiv:1810.01370*.

Shaikh, A. M., M. Simonsen, E. J. Vytlacil, and N. Yildiz (2009). A specification test for the propensity score using its distribution conditional on participation. *Journal of Econometrics 151*(1), 33–46.

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*, 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. 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.

Supplementary Material for " A Unified Framework for Specification Tests of Continuous Treatment Effect Models"

Wei Huang
School of Mathematics and Statistics and Australian Research Council Centre of Excellence for Mathematical and Statistical Frontiers (ACEMS), University of Melbourne, Australia
and
Oliver Linton
Faculty of Economics, University of Cambridge
and
Zheng Zhang
Institute of Statistics & Big Data, Renmin University of China

A Some preliminary results

We recall some preliminary results which have been established in Ai et al. (2021). The following conditions are inherited from Ai et al. (2021):

Assumption A.1. *(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 A.2. *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 A.3. *(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 $\sup_{t \in \mathcal{T}} \|u_{K_1}(t)\| \leq \zeta_1(K_1)$ and $\sup_{\boldsymbol{x} \in \mathcal{X}} \|v_{K_2}(\boldsymbol{x})\| \leq \zeta_2(K_2)$, $K = K_1(N) K_2(N)$ and $\zeta(K) := \zeta_1(K_1) \zeta_2(K_2)$, such that $\zeta(K) K^{-\alpha} \to 0$ and $\zeta(K) \sqrt{K/N} \to 0$ as $N \to \infty$.*

Assumption A.4. $\zeta(K)\sqrt{K^2/N} \to 0$ *and* $\sqrt{N}K^{-\alpha} \to 0$.

See Ai et al. (2021) for a detailed discussion on Assumptions A.1 -A.4. Under these conditions, Ai et al. (2021, Theorem 3) established the following results:

Proposition 1. *Suppose that Assumptions A.1-A.3 hold. Then, we obtain the following:*

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

$$\int_{\mathcal{T}\times\mathcal{X}} |\hat{\pi}_K(t,\boldsymbol{x}) - \pi_0(t,\boldsymbol{x})|^2 dF_{T,X}(t,\boldsymbol{x}) = O_p\left\{\max\left(K^{-2\alpha}, \frac{K}{N}\right)\right\},$$

$$\frac{1}{N}\sum_{i=1}^{N} |\hat{\pi}_K(T_i,\boldsymbol{X}_i) - \pi_0(T_i,\boldsymbol{X}_i)|^2 = O_p\left\{\max\left(K^{-2\alpha}, \frac{K}{N}\right)\right\}.$$

Furthermore, for any estimand with the form of $\mathbb{E}\{\pi_0(T,\boldsymbol{X})R(T,\boldsymbol{X},Y)\}$, where $R(T,\boldsymbol{X},Y) \in L^1(dF_{T,X,Y})$, Proposition 2 of Ai et al. (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 2. *Suppose that Assumptions A.1-A.4 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,*

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

The following conditions are restatements of Assumptions 1-5 listed in the main paper:

Assumption 1. *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* $t \in \mathcal{T}$*.*

Assumption 2. *Under* H_0*, (i)* $\boldsymbol{\theta}^*$ *is an interior point of* Θ*, where* Θ *is a compact set in* $\mathbb{R}^p$*; (ii)* $\|M_N(\widehat{\boldsymbol{\theta}},\widehat{\pi}_K)\| = \inf_{\boldsymbol{\theta}\in\Theta_\delta}\|M_N(\boldsymbol{\theta},\widehat{\pi}_K)\| + o_P(N^{-1/2})$*, where* $\Theta_\delta := \{\boldsymbol{\theta}\in\Theta : \|\boldsymbol{\theta}-\boldsymbol{\theta}^*\| \leq \delta\}$*.*

Assumption 3. *Let* $\eta(T,\boldsymbol{X},Y;t)$ *be defined in (4.2),* $Var\{\eta(T,\boldsymbol{X},Y;t)\} < \infty$ *for all* $t \in \mathcal{T}$*.*

Assumption 4. *(i)* $w(t;\boldsymbol{\theta})$ *is continuously differentiable in* $\boldsymbol{\theta} \in \Theta$ *and continuous in* $t \in \mathcal{T}$*;*

(ii) $g(t;\boldsymbol{\theta})$ *is twice continuously differentiable in* $\boldsymbol{\theta} \in \Theta$ *and* $\nabla_{\boldsymbol{\theta}} g(t;\boldsymbol{\theta})$ *is continuous in* $t \in \mathcal{T}$*;*

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

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

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* $\Big\{m\{Y;g(T;\boldsymbol{\theta})\}:\boldsymbol{\theta}\in\Theta\Big\}$ *satisfies:*

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

for any $\boldsymbol{\theta}\in\Theta$ *and any small* $\delta>0$ *and for some finite positive constant* C*.*

B Proof of Theorem 1

Proof. We first show that, under H_0, $\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|\xrightarrow{P}0$. Because, under H_0, $\widehat{\boldsymbol{\theta}}$ (resp. $\boldsymbol{\theta}^*$) is a unique minimizer of $\|N^{-1}\sum_{i=1}^N\widehat{\pi}_K(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta})\}w(T_i;\boldsymbol{\theta})\|$ (resp. $\|\mathbb{E}[\pi_0(T,\boldsymbol{X})m\{Y;g(T;\boldsymbol{\theta})\}w(T;\boldsymbol{\theta})]\|$), from the theory of M-estimation (van der Vaart, 1998, Theorem 5.7), if the following condition holds:

$$\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{N}\sum_{i=1}^N\widehat{\pi}_K(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta})\}w(T_i;\boldsymbol{\theta})-\mathbb{E}\left[\pi_0(T,\boldsymbol{X})m\{Y;g(T;\boldsymbol{\theta})\}w(T;\boldsymbol{\theta})\right]\right\|\xrightarrow{p}0.$$

Then $\widehat{\boldsymbol{\theta}}\xrightarrow{p}\boldsymbol{\theta}^*$. Note that

$$\begin{aligned}
&\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{N}\sum_{i=1}^N\widehat{\pi}_K(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta})\}w(T_i;\boldsymbol{\theta})-\mathbb{E}\left[\pi_0(T,\boldsymbol{X})m\{Y_i;g(T_i;\boldsymbol{\theta})\}w(T_i;\boldsymbol{\theta})\right]\right\|\\
&\leq\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{N}\sum_{i=1}^N\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}\,m\{Y_i;g(T_i;\boldsymbol{\theta})\}w(T_i;\boldsymbol{\theta})\right\| \qquad (B.1)\\
&+\sup_{\boldsymbol{\theta}\in\Theta}\left\|\frac{1}{N}\sum_{i=1}^N\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta})\}w(T_i;\boldsymbol{\theta})-\mathbb{E}\left[\pi_0(T,\boldsymbol{X})m\{Y;g(T;\boldsymbol{\theta})\}w(T;\boldsymbol{\theta})\right]\right\|. \qquad (B.2)
\end{aligned}$$

We first show (B.1) is of $o_p(1)$. Using Assumptions 4 and 5, the Cauchy-Schwarz inequality and Proposition 1, we have that

$$\begin{aligned}
|(B.1)|&\leq\left\{\frac{1}{N}\sum_{i=1}^N\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}^2\right\}^{1/2}\cdot\sup_{\boldsymbol{\theta}\in\Theta}\left\{\frac{1}{N}\sum_{i=1}^N\|m\{Y;g(T;\boldsymbol{\theta})\}w(T;\boldsymbol{\theta})\|^2\right\}^{1/2}\\
&\leq o_p(1).
\end{aligned}$$

We next show (B.2) is of $o_p(1)$. Note that, by the law of large numbers, for every $\boldsymbol{\theta}\in\Theta$, $\left\|N^{-1}\sum_{i=1}^N\pi_0(T_i,\boldsymbol{X}_i)m\{Y_i;g(T_i;\boldsymbol{\theta})\}w(T_i;\boldsymbol{\theta})-\mathbb{E}\left[\pi_0(T,\boldsymbol{X})m\{Y;g(T;\boldsymbol{\theta})\}w(T;\boldsymbol{\theta})\right]\right\|\xrightarrow{P}$

0 holds. By Assumptions A.1, 4 (i) and 5, we have

$$\begin{aligned}&\mathbb{E}\left[\sup_{\boldsymbol{\theta}_1:\|\boldsymbol{\theta}_1-\boldsymbol{\theta}\|<\delta}\left\|\pi_0(T,\boldsymbol{X})m\{Y;g(T;\boldsymbol{\theta}_1)\}w(T;\boldsymbol{\theta}_1)-\pi_0(T,\boldsymbol{X})m\{Y;g(T;\boldsymbol{\theta})\}w(T;\boldsymbol{\theta})\right\|^2\right]\\ \leq& O(1)\cdot\mathbb{E}\left[\sup_{\boldsymbol{\theta}_1:\|\boldsymbol{\theta}_1-\boldsymbol{\theta}\|<\delta}\left\|m\{Y;g(T;\boldsymbol{\theta}_1)\}-m\{Y;g(T;\boldsymbol{\theta})\}\right\|^2\right]+O(1)\cdot\delta^2\\ \leq& O(1)\cdot\delta^2. \end{aligned} \tag{B.3}$$

With (B.3), Assumptions 2 (i), and Andrews (1994, Theorems 4 and 5), the class $\{\pi_0(T,\boldsymbol{X})m\{Y;g(T;\boldsymbol{\theta})\}w(T;\boldsymbol{\theta}):\boldsymbol{\theta}\in\Theta\}$ is stochastically equicontinuous. Then we have that (B.2) is of $o_p(1)$. Hence we have that $\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\|\xrightarrow{P}0$ under H_0.

We start to derive the asymptotic distribution of $\widehat{J}_N(t)$. Note that

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

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

$$\widehat{J}_N(t)=\frac{1}{\sqrt{N}}\sum_{i=1}^N\widehat{U}_i\mathscr{H}(T_i,t)=\frac{1}{\sqrt{N}}\sum_{i=1}^N U_i\mathscr{H}(T_i,t) \tag{B.4}$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^N\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\mathscr{H}(T_i,t) \tag{B.5}$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^N\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t) \tag{B.6}$$

$$+\frac{1}{\sqrt{N}}\sum_{i=1}^N\{\widehat{\pi}_K(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\}\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t). \tag{B.7}$$

The subsequent proof consists of the following key steps:

Step 1. Establishing the asymptotically equivalent representation for (B.5) in terms of *i.i.d.* summations;

Step 2. Establishing the asymptotically equivalent representation for (B.6) in terms of *i.i.d.* summations;

Step 3. Showing (B.7) is of $o_P(1)$.

Using Proposition 2, 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

$$
\begin{aligned}
(\text{B.5}) = & -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\pi_0(T_i, \boldsymbol{X}_i)\cdot\mathbb{E}[m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}\cdot\mathscr{H}(T_i, t)|T_i, \boldsymbol{X}_{\boldsymbol{i}}] \\
& + \frac{1}{\sqrt{N}}\sum_{i=1}^{N}\mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}\cdot\mathscr{H}(T_i, t)|\boldsymbol{X}_i] + o_P(1) \\
= & -\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\phi(T_i, \boldsymbol{X}_i; t) + o_P(1). \qquad \text{(B.8)}
\end{aligned}
$$

We next find the expression for $\sqrt{N}\{\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\}$ by applying Pakes and Pollard (1989, Theorem 3.3). We begin to verify the Conditions (i)-(v) imposed in Pakes and Pollard (1989, Theorem 3.3).

- For Condition (i) of Pakes and Pollard (1989, Theorem 3.3). By Assumption 2,

$$\|M_N(\widehat{\boldsymbol{\theta}}, \widehat{\pi}_K)\| = \inf_{\boldsymbol{\theta}\in\Theta_\delta}\|M_N(\boldsymbol{\theta}, \widehat{\pi}_K)\| + o_P(N^{-1/2}),$$

  where

$$M_N(\boldsymbol{\theta}, \pi) := \frac{1}{N}\sum_{i=1}^{N}\pi(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta}).$$

  By Proposition 2, we have

$$
\begin{aligned}
& M_N(\boldsymbol{\theta}, \widehat{\pi}_K) \\
= & \frac{1}{N}\sum_{i=1}^{N}\Big[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta}) - \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta})|T_i, \boldsymbol{X}_i] \\
& + \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta})|T_i] + \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta})|\boldsymbol{X}_i] \\
& - \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta})]\Big] + o_P(N^{-1/2}) \\
=: & G_N(\boldsymbol{\theta}) + o_P(N^{-1/2}).
\end{aligned}
$$

  where the definition of $G_N(\boldsymbol{\theta})$ is obvious and the equation holds uniformly in $\boldsymbol{\theta}$. Now we have

$$\|G_N(\widehat{\boldsymbol{\theta}})\| = \inf_{\boldsymbol{\theta}\in\Theta_\delta}\|G_N(\boldsymbol{\theta})\| + o_P(N^{-1/2}),$$

  thus Condition (i) of Pakes and Pollard (1989, Theorem 3.3) holds.

- For Condition (ii) of Pakes and Pollard (1989, Theorem 3.3). Let

$$G(\boldsymbol{\theta}) := \mathbb{E}[G_N(\boldsymbol{\theta})] = \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)m\{Y_i; g(T_i; \boldsymbol{\theta})\}w(T_i; \boldsymbol{\theta})],$$

  and Assumption 4 (iii) ensures that the derivative $\nabla_{\boldsymbol{\theta}}G(\boldsymbol{\theta}^*)$ is full rank. Hence, Condition (ii) of Pakes and Pollard (1989, Theorem 3.3) holds.

- For Condition (iii) of Pakes and Pollard (1989, Theorem 3.3). 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 $f(\cdot)$. Assumptions A.1, 4, Assumption 5 and the compactness of Θ imply the empirical processes

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

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

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

and

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

are stochastically equicontinuous (Andrews (1994, Theorems 4 and 5)). Note that

$$\begin{aligned}\sqrt{N}\{G_N(\boldsymbol{\theta}) - G(\boldsymbol{\theta})\} =& \nu_N(\pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i; \boldsymbol{\theta})\right\} w(T_i; \boldsymbol{\theta})) \\ & - \nu_N(\mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i; \boldsymbol{\theta})\right\} w(T_i; \boldsymbol{\theta}) | T_i, \boldsymbol{X}_i\right]) \\ & + \nu_N(\mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i; \boldsymbol{\theta})\right\} w(T_i; \boldsymbol{\theta}) | T_i\right]) \\ & + \nu_N(\mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) m\left\{Y_i; g(T_i; \boldsymbol{\theta})\right\} w(T_i; \boldsymbol{\theta}) | \boldsymbol{X}_i\right])\end{aligned}$$

Then for every sequence $\{\delta_N\}$ of positive numbers that converges to zero,

$$\sup_{\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \le \delta_N} \frac{\sqrt{N}\|G_N(\boldsymbol{\theta}) - G(\boldsymbol{\theta}) - G_N(\boldsymbol{\theta}^*)\|}{1 + \sqrt{N}\{\|G_N(\boldsymbol{\theta})\| + \|G(\boldsymbol{\theta})\|\}} = o_P(1).$$

Thus, Condition (iii) of Pakes and Pollard (1989, Theorem 3.3) holds.

- The Condition (iv) of Pakes and Pollard (1989, Theorem 3.3) is satisfied by noting that under $H_0 : \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\} w(T_i; \boldsymbol{\theta}^*) | T_i] = 0$ and

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

is a sum of i.i.d. random variables of mean zero.

- The Condition (v) of Pakes and Pollard (1989, Theorem 3.3), i.e. $\boldsymbol{\theta}^*$ is an interior point of Θ, is satisfied by Assumption 2 (i).

Therefore, all conditions of Pakes and Pollard (1989, Theorem 3.3) hold, and we get

$$
\begin{aligned}
&\sqrt{N}\left\{\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^*\right\} \qquad\qquad (B.9)\\
=&\Bigg\{-\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)w(T_i;\boldsymbol{\theta}^*)^\top\right]\\
&\cdot\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot w(T_i;\boldsymbol{\theta}^*)\nabla_{\boldsymbol{\theta}}^\top g(T_i;\boldsymbol{\theta}^*)\right]\Bigg\}^{-1}\\
&\cdot\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)\cdot\frac{\partial}{\partial g}\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)w(T_i;\boldsymbol{\theta}^*)^\top\right]\\
&\times\frac{1}{\sqrt{N}}\sum_{i=1}^N\Bigg\{\pi_0(T_i,\boldsymbol{X}_i)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}w(T_i;\boldsymbol{\theta}^*)\\
&\qquad-\pi_0(T_i,\boldsymbol{X}_i)w(T_i;\boldsymbol{\theta}^*)\cdot\mathbb{E}[m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|T_i,\boldsymbol{X}_i]\\
&\qquad+\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)w(T_i;\boldsymbol{\theta}^*)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|\boldsymbol{X}_i]\Bigg\}+o_P(1).
\end{aligned}
$$

Consider the term (B.6). Note that

$$
\begin{aligned}
(B.6)=&\nu_N\left\{\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\widehat{\boldsymbol{\theta}})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t)\right\}\\
&+\sqrt{N}\cdot\mathbb{E}\left\{\pi_0(T_i,\boldsymbol{X}_i)\left[m\left\{Y_i;g(T_i;\boldsymbol{\theta})\right\}-m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}\right]\mathscr{H}(T_i,t)\right\}\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}}.
\end{aligned}
$$

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

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

By (B.9), we have

$$
(B.6)=-\frac{1}{\sqrt{N}}\sum_{i=1}^N\psi(T_i,\boldsymbol{X}_i,Y_i;t)+o_p(1), \qquad (B.10)
$$

where

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

For the term (B.7), we have

$$
\begin{aligned}
|(\text{B.7})| =& \left|\frac{1}{\sqrt{N}} \sum_{i=1}^N \{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\} \left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\right] \mathscr{H}(T_i, t)\right| \\
\leq& \sqrt{N} \cdot \sup_{(t, \boldsymbol{x}) \in \mathcal{T} \times \mathcal{X}} |\widehat{\pi}_K(t, \boldsymbol{x}) - \pi_0(t, \boldsymbol{x})| \\
& \cdot \frac{1}{N} \sum_{i=1}^N \left|m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\} \mathscr{H}(T_i, t)\right| \\
=& \sqrt{N} \cdot O_P\left(\zeta(K) K^{-\alpha} + \zeta(K)\sqrt{\frac{K}{N}}\right) \\
& \cdot \left\{\mathbb{E}\left[\left|m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\{Y_i; g(T_i; \boldsymbol{\theta}^*)\}\right| \cdot |\mathscr{H}(T_i, t)|\right] + O_P\left(N^{-1/2}\right)\right\} \\
\leq& O_P\left(\zeta(K) K^{-\alpha} + \zeta(K)\sqrt{\frac{K}{N}}\right) \cdot \sqrt{N} \cdot \left\{O(1) \cdot \|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\| + O_P\left(N^{-1/2}\right)\right\} \\
=& o_P(1), \qquad\qquad (\text{B.11})
\end{aligned}
$$

where the second equality holds by Proposition 1 and the law of large numbers; the second inequality holds by Assumption 5; and the last equality holds by (B.9) and Assumption A.3.

Hence, combining (B.4), (B.8), (B.10), and (B.11), we have

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

where $\mathbb{E}\{\phi(T_i, \boldsymbol{X}_i; t)\} = 0$ and $\mathbb{E}\{\psi(T_i, \boldsymbol{X}_i, Y_i; t)\} = 0$. We know that

$$\mathbb{E}\left[\int_{\mathcal{T}}\left\{\frac{1}{\sqrt{N}}\sum_{i=1}^{N}\eta(T_i, \boldsymbol{X}_i, Y_i; t)\right\}^2\right] = \int_{\mathcal{T}}\mathbb{E}[\{\eta(T_i, \boldsymbol{X}_i, Y_i; t)\}^2]dt < \infty,$$

that is, $N^{-1/2}\sum_{i=1}^{N}\eta(T_i, \boldsymbol{X}_i, Y_i; \cdot)$ is tight. Hence, by the functional central limit theorem for Hilbert-valued random arrays Li et al. (2003, Lemma 2.1), we have that 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

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

Hence, (i) and (ii) are proved.

(iii) 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 $h(\widehat{J}_N) = \int\{\widehat{J}_N(t)\}^2 dF_T(t)$ converges to $\int\{J_\infty(t)\}^2 dF_T(t)$ in distribution. By applying a similar argument to the proof of Theorem 2.2 (ii) of Li et al. (2003), we have $|\widehat{CM}_N - h(\widehat{J}_N)| = o_P(1)$. This completes the proof of Theorem 1 (iii). Part (iv) follows from Theorem 1 (i) and the continuous mapping theorem. □

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

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

.

Then

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

and

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

We have

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

where the second 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

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

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

$$\widehat{U}_i = U_{iN} + \{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\} m\{Y_i; g_N(T_i; \boldsymbol{\theta}_N^*)\}$$

$$+ \pi_0(T_i, \boldsymbol{X}_i) \left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\left\{Y_i; g_N(T_i; \boldsymbol{\theta}_N^*)\right\}\right]$$
$$+ \{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\} \left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\left\{Y_i; g_N(T_i; \boldsymbol{\theta}_N^*)\right\}\right].$$

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 2, 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

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

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

We consider the term (D.3). We first find the expression for $\sqrt{N}\{\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_N^*\}$. Similar to (B.9) in the proof of Theorem 1, by applying Pakes and Pollard (1989, Theorem 3.3), we have

$$\sqrt{N}\left\{\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}^*\right\} \tag{D.5}$$

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

We next find the expression for $\sqrt{N}\{\boldsymbol{\theta}^* - \boldsymbol{\theta}^*_N\}$. Note that under the local alternative $H_L : \mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i) m\{Y_i; g_N(T_i; \boldsymbol{\theta}^*_N)\} | T_i] = 0$, using the mean value theorem, we have

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

where $\widetilde{\boldsymbol{\theta}}$ lies on the line joining from $\boldsymbol{\theta}^*$ and $\boldsymbol{\theta}^*_N$, and $\widetilde{g}_N(T_i; \boldsymbol{\theta}) := g(T_i; \boldsymbol{\theta}) + \gamma \cdot \delta(T)/\sqrt{N}$ for some $\gamma \in (0, 1)$. Then

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

Hence, we get

$$
\begin{aligned}
&\sqrt{N}\left\{\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_N^*\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}^*)w(T_i;\boldsymbol{\theta}^*)^\top\right] \\
&\quad\cdot\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 w(T_i;\boldsymbol{\theta}^*)\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)^\top\right]\Bigg\}^{-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\}|T_i,\boldsymbol{X}_i\right]\cdot\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)w(T_i;\boldsymbol{\theta}^*)^\top\right] \\
&\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 w(T_i;\boldsymbol{\theta}^*)\right] \\
&-\Bigg\{\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}^*)w(T_i;\boldsymbol{\theta}^*)^\top\right] \\
&\quad\cdot\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 w(T_i;\boldsymbol{\theta}^*)\nabla_{\boldsymbol{\theta}}g(T_i;\boldsymbol{\theta}^*)^\top\right]\Bigg\}^{-1} \\
&\quad\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}^*)w(T_i;\boldsymbol{\theta}^*)^\top\right] \\
&\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\}w(T_i;\boldsymbol{\theta}^*) \\
&\qquad\qquad-\pi_0(T_i,\boldsymbol{X}_i)w(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)w(T_i;\boldsymbol{\theta}^*)m\left\{Y_i;g(T_i;\boldsymbol{\theta}^*)\right\}|\boldsymbol{X}_i]\Big\}+o_P(1).
\end{aligned}
$$

Then similar to (B.10), 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

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

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

Similar to (B.7), we have that (D.4) is of $o_P(1)$.

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

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

$$= \frac{1}{N}\sum_{i=1}^{N} U_i \mathscr{H}(T_i, t) \tag{D.6}$$

$$+ \frac{1}{N}\sum_{i=1}^{N}\{\widehat{\pi}_K(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i)\} m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\} \mathscr{H}(T_i, t) \tag{D.7}$$

$$+ \frac{1}{N}\sum_{i=1}^{N} \pi_0(T_i, \boldsymbol{X}_i)\left[m\left\{Y_i; g(T_i; \widehat{\boldsymbol{\theta}})\right\} - m\left\{Y_i; g(T_i; \boldsymbol{\theta}^*)\right\}\right] \mathscr{H}(T_i, t) \tag{D.8}$$

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

By applying a similar argument for (B.5)-(B.7), we have that (D.7)-(D.9) are of $o_P(1)$. Under H_1, the law of large numbers implies (D.6) $= \mu_1(t) + o_P(1)$. Hence, we conclude the proof.

□

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

Theorem 3. *Suppose that $m(y;g)$ is differentiable with respect to g, Assumptions 1-5 and Assumptions A.1-A.4 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 1, $|N^{-1}\cdot\widehat{CM}_N(\boldsymbol{\theta}) - CM(\boldsymbol{\theta})| \xrightarrow{P} 0$ 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 1, under H_0, it is easy to obtain

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

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 2, 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

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

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	0.021	0.067	0.124	0.010	0.064	0.136	0.008	0.058	0.117
		200	0.015	0.056	0.111	0.014	0.062	0.135	0.012	0.056	0.107
		500	0.008	0.051	0.108	0.012	0.059	0.117	0.011	0.048	0.107
	DGP0-NL	100	0.025	0.085	0.153	0.014	0.070	0.134	0.007	0.064	0.119
		200	0.021	0.059	0.119	0.013	0.069	0.131	0.012	0.065	0.110
		500	0.012	0.058	0.110	0.011	0.052	0.105	0.011	0.057	0.111
Median	**DGP0-L**	100	0.035	0.107	0.182	0.016	0.065	0.132	0.031	0.106	0.162
		200	0.022	0.081	0.141	0.016	0.068	0.121	0.025	0.065	0.134
		500	0.017	0.064	0.121	0.011	0.066	0.125	0.013	0.052	0.110
	DGP0-NL	100	0.040	0.124	0.196	0.009	0.063	0.109	0.026	0.097	0.172
		200	0.025	0.078	0.133	0.010	0.073	0.127	0.021	0.074	0.140
		500	0.010	0.059	0.119	0.016	0.053	0.116	0.015	0.072	0.126

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 establishing Theorem 1. □

F Additional simulation results of KS-type statistic

We also performed the simulation studies described in section 6.2 of the paper using the KS-type statistic. The results are similar to those of the CM-type one.

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 KS test method.

G Estimating and testing Tobit linear models

Let

$$Y(t) = \boldsymbol{\beta}^\top \boldsymbol{t} + \epsilon ,$$

for some unknown parameter $\boldsymbol{\beta}$ in a compact set in $\mathbb{R}^p$, where $\boldsymbol{t} = (1, t, t^2, \ldots, t^{p-1})^\top$ for some positive integer p and ϵ is a normal random variable with mean 0 and unknown

Table 2: Estimated power

$m(\cdot)$	Model	N	Logistic 1%	Logistic 5%	Logistic 10%	Cosine-Sine 1%	Cosine-Sine 5%	Cosine-Sine 10%	Indicator 1%	Indicator 5%	Indicator 10%
Average	**DGP1-L**	100	0.587	0.807	0.885	0.488	0.714	0.825	0.505	0.748	0.850
		200	0.924	0.982	0.996	0.890	0.970	0.986	0.919	0.984	0.998
	DGP1-NL	100	0.483	0.693	0.800	0.444	0.698	0.803	0.382	0.600	0.732
		200	0.721	0.895	0.945	0.801	0.910	0.960	0.762	0.892	0.930
Median	**DGP1-L**	100	0.277	0.533	0.660	0.164	0.372	0.525	0.263	0.525	0.655
		200	0.606	0.818	0.907	0.505	0.7434	0.834	0.612	0.829	0.902
	DGP1-NL	100	0.209	0.399	0.523	0.167	0.363	0.495	0.161	0.365	0.487
		200	0.356	0.593	0.732	0.350	0.625	0.756	0.380	0.632	0.755

variance σ^2. A Tobit linear model assumes the potential outcome

$$Y^*(t) = \begin{cases} Y(t) & \text{if} \quad Y(t) > 0, \\ 0 & \text{if} \quad Y(t) \leq 0\,. \end{cases}$$

It can be shown that the log-likelihood function of $\boldsymbol{\beta}$ and σ given $Y^*(t)$ is

$$\ln f\{Y^*(t), t, \boldsymbol{\beta}, \sigma\} = \sum_{Y_i^*(t)=0} \ln \left[\Phi\left\{ -\frac{\boldsymbol{\beta}^\top \boldsymbol{t}}{\sigma} \right\} \right] + \sum_{Y_i^*(t)>0} \ln \left[\sigma^{-1} \phi\left\{ \frac{Y_i^*(t) - \boldsymbol{\beta}^\top \boldsymbol{t}}{\sigma} \right\} \right],$$

where $\Phi(\cdot)$ and $\phi(\cdot)$ are respectively the distribution function and density function of a standard normal random variable. Olsen (1978) proposed a reparametrization $\boldsymbol{\beta} = \boldsymbol{\delta}/\gamma$ and $\sigma^2 = \gamma^{-2}$, the resulting transformed log-likelihood of the parameter $\boldsymbol{\theta} = (\boldsymbol{\delta}, \gamma)$ is then

$$\ln f\{Y^*(t), t, \boldsymbol{\theta}\} = \sum_{Y_i^*(t)=0} \ln[\Phi\{-(\boldsymbol{\delta}^\top \boldsymbol{t})\}] + \sum_{Y_i^*(t)>0} \ln(\gamma) + \ln[\phi\{\gamma Y_i^*(t) - \boldsymbol{\delta}^\top \boldsymbol{t}\}],$$

which is globally concave in terms of $\boldsymbol{\theta}$.

Note that in this case, we can test the model by testing

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

against the alternative hypothesis

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

where Θ is a compact set in $\mathbb{R}^{p+1}$.

This is a multi-dimensional moment condition. It is straightforward to extend our test method by taking $m\{Y^*(t); g(t; \boldsymbol{\theta})\} = \nabla_{\boldsymbol{\theta}} \ln f\{Y^*(t), t, \boldsymbol{\theta}\}$ and $w(T; \boldsymbol{\theta})$ in (3.6) to be 1. In particular, our Cramer-von Mises (CM)-type statistic and Kolmogorov-Smirnov(KS)-type statistic are extended to

$$CM_N^0 = \frac{1}{N} \sum_{i=1}^{N} \{J_N^0(T_i)\}^\top \{J_N^0(T_i)\} \quad \text{and} \quad KS_N^0 = \sup_{t \in \mathcal{T}} \|J_N^0(t)\|_\infty,$$

where $\|\cdot\|_\infty$ is the maximum norm of a vector and

$$J_N^0(t) = \frac{1}{N}\sum_{i=1}^{N}\pi_0(T_i,\boldsymbol{X}_i)\nabla_{\boldsymbol{\theta}}\ln f(Y_i,T_i,\boldsymbol{\theta})\mathscr{H}(T_i,t)\,.$$

We can estimate $J_N^0(t)$ by

$$\widehat{J}_N(t) = \frac{1}{N}\sum_{i=1}^{N}\widehat{\pi}_K(T_i,\boldsymbol{X}_i)\nabla_{\boldsymbol{\theta}}\ln f(Y_i,T_i,\widehat{\boldsymbol{\theta}}_{LL})\mathscr{H}(T_i,t)\,,$$

where $\widehat{\pi}_K$ is defined the same as that in section 3 and

$$\widehat{\boldsymbol{\theta}}_{LL} := \arg\min_{\boldsymbol{\theta}\in\Theta}\|M_N(\boldsymbol{\theta},\widehat{\pi}_K)\|\,,$$

where

$$M_N(\boldsymbol{\theta},\pi) := \frac{1}{N}\sum_{i=1}^{N}\pi(T_i,\boldsymbol{X}_i)\nabla_{\boldsymbol{\theta}}\ln f(Y_i,T_i,\boldsymbol{\theta})\,.$$

Theorem 1 can be applied here under Assumptions A.1 to 4 in section 4 and

Assumption G.1. *(i)* $\mathbb{E}\left[\sup_{\boldsymbol{\theta}\in\Theta}\|\nabla_{\boldsymbol{\theta}}\ln f\{Y,T,\boldsymbol{\theta}\}\|^{2+\delta}\right] < \infty$ *for some* $\delta > 0$*; (ii) The function class* $\left\{\nabla_{\boldsymbol{\theta}}\ln f\{Y,T,\boldsymbol{\theta}\} : \boldsymbol{\theta}\in\Theta\right\}$ *satisfies:*

$$\mathbb{E}\left[\sup_{\boldsymbol{\theta}_1:\|\boldsymbol{\theta}_1-\boldsymbol{\theta}\|<\delta}\|\nabla_{\boldsymbol{\theta}}\ln f\{Y,T,\boldsymbol{\theta}_1\} - \nabla_{\boldsymbol{\theta}}\ln f\{Y,T,\boldsymbol{\theta}\}\|^2\right]^{1/2} \le C\cdot\delta$$

for any $\boldsymbol{\theta}\in\Theta$ *and any small* $\delta > 0$ *and for a finite positive constant* C*.*

With

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

and

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

and

$$\eta(T_i,\boldsymbol{X}_i,Y_i;t) := \pi_0(T_i,\boldsymbol{X}_i)\cdot\nabla_{\boldsymbol{\theta}}\ln f(Y_i,T_i,\boldsymbol{\theta})\mathscr{H}(T_i,t) - \phi(T_i,\boldsymbol{X}_i;t) - \psi(T_i,\boldsymbol{X}_i,Y_i;t).$$

The approximation method of the null limiting distribution described in section 5 can be directly applied here.

References

Ai, C., O. Linton, K. Motegi, and Z. Zhang (2021). A unified framework for efficient estimation of general treatment models. *Quantitative Economics 12*(3), 779–816.

Ai, C., O. Linton, and Z. Zhang (2021). Estimation and inference for the counterfactual distribution and quantile functions in continuous treatment models. *Journal of Econometrics*.

Andrews, D. W. K. (1994). Empirical process methods in econometrics. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume 4, Chapter 37, pp. 2247–2294. Citeseer.

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.

Olsen, R. J. (1978). Note on the uniqueness of the maximum likelihood estimator for the tobit model. *Econometrica: Journal of the Econometric Society*, 1211–1215.

Pakes, A. and D. Pollard (1989). Simulation and the asymptotics of optimization estimators. *Econometrica 57*(5), 1027–1057.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.