Non-parametric Estimation and Uniform Inference of General Treatment Models

Chunrong Ai [*†], Xiaohong Chen [††], Wei Huang [‡‡], and Zheng Zhang [§⋆]

[†]School of Management and Economics, Chinese University of Hong Kong, Shenzhen

[†] Department of Economics, Yale University

[‡] School of Mathematics and Statistics, University of Melbourne

[⋆]Center for Applied Statistics, Institute of Statistics & Big Data, Renmin University of China

Abstract

Ai et al. (2019) proposed an estimation of the general treatment model that encompasses binary, multi-valued, continuous, and a mixture of discrete and continuous treatments with a variety of parameters of interest including the average treatment effect and the quantile treatment effect. They considered a parametric treatment-effect function and proposed a weighted regression with the weight function estimated by the maximum entropy method. This paper extends their work to the nonparametric treatment-effect and the nonparametric heterogenous treatment-effect functions. Under some sufficient conditions, the paper derives large sample properties of the proposed estimator and establishes uniform confidence bands. A small-scale simulation study and an application demonstrate the usefulness of the proposed approach.

Keywords: Conditional density ratio; heterogeneity; quantile treatment effect; uniform confidence band.

[*]E-mail: chunrongai@cuhk.edu.cn

[†]E-mail: xiaohong.chen@yale.edu

[‡]E-mail: wei.huang@unimelb.edu.au

[§]E-mail: zhengzhang@ruc.edu.cn

1 Introduction

Ai et al. (2021) proposed an estimation of the general treatment model that encompasses binary, multi-valued, continuous, and a mixture of discrete and continuous treatments, with a variety of parameters of interest including the average treatment effect, the quantile treatment effect, and other types of treatment effect. They considered a parametric treatment-effect function, a function of the treatment status only, and proposed a weighted regression to estimate the treatment-effect function, with the weight function estimated by the maximum entropy approach. Under some sufficient conditions, they established large sample properties of the estimated treatment-effect function. This paper extends their work to the nonparametric and heterogeneous treatment-effect function, an unknown function of the treatment status, and some covariates. This extension is useful because a false parameterization of the treatment-effect function could lead to an erroneous conclusion and heterogeneous treatment-effect function enables individualized policy design. Under some sufficient conditions, we derive the asymptotic distribution of the estimated treatment-effect function and establish uniform confidence bands. The uniform confidence bands enable more powerful inference.

This paper contributes to the literature by extending the estimation and inference of the conditional average treatment effect (see e.g. Heckman and Vytlacil, 2005; Abrevaya et al., 2015; Nie and Wager, 2021; Fan et al., 2022) and the conditional quantile treatment effect (Giessing and Wang, 2021; Xu et al., 2022) in the binary treatment model to the general treatment model, and extending the estimation and inference of the average treatment-effect function (Hirano and Imbens, 2004; Imai and van Dyk, 2004; Florens et al., 2008; Kennedy et al., 2017; Fong et al., 2018; Huang and Zhang, 2022) and the quantile treatment-effect function (e.g. Galvao and Wang (2015); Ai et al. (2021, 2022)) in the continuous treatment model to the heterogeneous treatment-effect function. Furthermore, the uniform confidence bands in the continuous treatment models appear new and have not been reported in the literature.

The paper is organized as follows. Section 2 sets up the basic framework. Section 3 presents a sequential estimation procedure in which the weights are estimated in the first stage by minimizing the distance between arbitrary weights and uniform weights. The estimated weights are then plugged into the second stage of weighted regression. Section 4 establishes the large sample properties of the estimated heterogeneous treatment-effect function. Section 5 constructs the uniform bands based on the multiplier bootstrap. Section 6.2 suggests a data-driven approach for selecting tuning parameters. Section 7 reports a simulation study, while Section 8 presents an empirical application, followed by some concluding remarks in Section 9. All technical proofs and extra simulation results are relegated to supplemental material.

2 Basic Framework

Let T denote the observed treatment status variable, with support $\mathcal{T} \subset \mathbb{R}$ and a marginal probability distribution function $F_T(t)$. Let $Y^*(t)$ denote the potential response when treatment $T = t$ is assigned. The probability density of $Y^*(t)$ exists, denoted by $f_{Y^*(t)}$, and is continuously differentiable. Let $\mathcal{L}(\cdot)$ denote a nonnegative, strictly convex, and possibly non-smooth loss function satisfying $\mathcal{L}(0) = 0$ and $\mathcal{L}(v) \geq 0$ for all $v \in \mathbb{R}$. The derivative of $\mathcal{L}(\cdot)$, denoted by $\mathcal{L}'(\cdot)$, exists almost everywhere and is non-constant. Let $\boldsymbol{X} \in \mathbb{R}^{d_X}$, for some integer $d_X \geq 1$, denote a vector of observable covariates with support $\mathcal{X}$. Let $Z = H(\boldsymbol{X})$ denote a known transformation of $\boldsymbol{X}$, with dimension $d_Z \leq d_X$ and support $\mathcal{Z}$. Some examples of the transformation are $Z = \boldsymbol{X}$ or $Z = X_1$, a subset of $\boldsymbol{X}$, or $Z = \sum_{j=1}^{J} \mathbb{1}(\boldsymbol{X} \in \mathcal{X}_j) \cdot j$, with $\cup_{j=1}^{J} \mathcal{X}_j = \mathcal{X}$. The covariates Z shall be used to model the heterogeneity in the treatment-effect function across subpopulations. The *conditional-response function*, $g_0(\cdot,\cdot)$, is the unique solution to the following convex optimization problem:

$$g_0(\cdot,\cdot) = \underset{g(\cdot,\cdot)}{\operatorname{argmin}} \int_{\mathcal{T}} \mathbb{E}\left[\mathcal{L}\left\{Y^*(t) - g(t,Z)\right\}\right] dF_T(t). \tag{2.1}$$

The *heterogeneous treatment-effect function,* denoted by HTE, is defined as $\tau_0(t_1, t_0|z) = g_0(t_1, z) - g_0(t_0, z)$ when the treatment status is changed discretely from $T = t_0$ to $T = t_1$ or $\tau_0(t|z) = \partial_t g_0(t,z) = \partial g_0(t,z)/\partial t$ when the treatment status is changed marginally.

Model (2.1) encompasses many popular models with a variety of parameters of interest. For example, with $\mathcal{L}(v) = v^2$ and $\mathcal{T} = \{0, 1\}$, model (2.1) gives the conditional average response function for the binary treatment, $g_0(t,z) = \mathbb{E}\{Y^*(t)|Z = z\}$. The conditional average treatment effect, $g_0(1,z) - g_0(0,z)$, is studied in Heckman et al. (1998); Hahn (1998); Heckman and Vytlacil (2005); Crump et al. (2008); Wager and Athey (2018); Kennedy (2020) for the case $Z = \boldsymbol{X}$, and in Abrevaya et al. (2015); Fan et al. (2022) for the case Z = a strict subset of $\boldsymbol{X}$. With $\mathcal{L}(v) = v\{\tau - \mathbb{1}(v \leq 0)\}$ for some $\tau \in (0, 1)$ and $\mathcal{T} = \{0, 1\}$, model (2.1) gives the conditional quantile response function for the binary treatment, $g_0(t,z) = F^{-1}_{Y^*(t)|Z}(\tau|z) = \inf\{q : \mathbb{P}\{Y^*(t) \geq q|Z = z\} \leq \tau\}$. The conditional quantile treatment effect, $g_0(1,z) - g_0(0,z)$, is studied in Chernozhukov and Hansen (2005).

Let $Y = Y^*(T)$ denote the observed response. Throughout the paper, we impose the following condition to identify the conditional response function (Hirano and Imbens, 2004; Ai et al., 2021).

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

Let $f_{T|X}$ denote the conditional density of T given the observed covariates $\boldsymbol{X}$. Under Assumption 1, the conditional-response function, $g_0(t,z)$, solves

$$g_0(\cdot,\cdot) = \underset{g(\cdot,\cdot)}{\operatorname{argmin}}\, \mathbb{E}\left[\pi_0(T,\boldsymbol{X})\mathcal{L}\left\{Y - g(T,Z)\right\}\right].$$

where

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

is the *weight function*(Robins et al., 2000; Ai et al., 2021). We shall discuss the estimation of $\pi_0(t,\boldsymbol{x})$, $g_0(t,z)$, and the partial derivative $\partial_t g_0(t,z)$ in the next section.

Remark 1. *Ai et al. (2021) studied a parametric-response function, $g_0(t,\boldsymbol{\beta}_0)$, with the finite dimensional parameter $\boldsymbol{\beta}_0 \in \mathbb{R}^p$ solving the following optimization problem:*

$$\boldsymbol{\beta}_0 = \underset{\boldsymbol{\beta}}{\operatorname{argmin}}\, \mathbb{E}\left[\pi_0(T,\boldsymbol{X})\mathcal{L}\{Y - g(T;\boldsymbol{\beta})\}\right],$$

where $g(\cdot,\cdot)$ is a known function. The parametric-response function cannot capture the heterogeneity of the treatment effects across subpopulations. Moreover, the parametric form could be mis-specified, leading to a false conclusion.

3 Estimation

We propose an estimation of the conditional-response function by adapting the sequential estimation approach of Ai et al. (2021). We notice that, for any suitable function $u(t, \boldsymbol{x})$, the weight function $\pi_0(t, \boldsymbol{x})$ satisfies the following moment restriction,

$$\mathbb{E}\left\{\pi_0(T, \boldsymbol{X})u(T, \boldsymbol{X})\right\} = \int u(t, \boldsymbol{x})f_T(t)f_X(\boldsymbol{x})dtd\boldsymbol{x}. \tag{3.1}$$

Equation (3.1) identifies $\pi_0(t, \boldsymbol{x})$ so we can estimate $\pi_0(T_i, \boldsymbol{X}_i)$ by solving the sample analogue of (3.1). The challenge is that (3.1) has an infinite number of restrictions. It is impossible to impose an infinite number of restrictions on a finite number of sample observations. To overcome this difficulty, we approximate a functional space by a sequence of finite-dimensional sieve spaces. Specifically, let $u_K(T, \boldsymbol{X}) = (u_{K,1}(T, \boldsymbol{X}), \ldots, u_{K,K}(T, \boldsymbol{X}))^\top$ denote the *approximation sieves*, such as B-splines and power series (see Newey, 1997; Chen, 2007, for more discussion on sieve approximation). $\pi_0(T, \boldsymbol{X})$ also satisfies

$$\mathbb{E}\left\{\pi_0(T, \boldsymbol{X})u_K(T, \boldsymbol{X})\right\} = \int u_K(t, \boldsymbol{x})f_T(t)f_X(\boldsymbol{x})dtd\boldsymbol{x}\,. \tag{3.2}$$

Let $D(v, v_0)$ denote a distance measure between the weight v and the design weight v_0. $D(v, v_0)$ is continuously differentiable in v, nonnegative, and strictly convex in v, and satisfies $D(v_0, v_0) = 0$. The calibration idea put forward by Deville and Särndal (1992) is to minimize the distance between the final weights and the design weights subject to the sample analog of (3.2). Since $\mathbb{E}[\pi_0(T_i, \boldsymbol{X}_i)] = 1$, we set $v_0 = 1$ and estimate $\pi_0(T_i, \boldsymbol{X}_i)$ by $\widehat{\pi}_i$:

$$\begin{cases} \{\widehat{\pi}_i\}_{i=1}^N = \operatorname{argmin} \sum_{i=1}^N D(\pi_i, 1) \\ \text{subject to } \frac{1}{N}\sum_{i=1}^N \pi_i u_K(T_i, \boldsymbol{X}_i) = \frac{1}{N(N-1)}\sum_{j=1, j\neq i}^N \sum_{i=1}^N u_K(T_i, \boldsymbol{X}_j). \end{cases} \tag{3.3}$$

The dual solution of the primal problem (3.3) is

$$\widehat{\pi}_K(T_i, \boldsymbol{X}_i) = \rho'\left\{\widehat{\boldsymbol{\lambda}}_K^\top u_K(T_i, \boldsymbol{X}_i)\right\},$$

with $\widehat{\boldsymbol{\lambda}}_K$ maximizing the strictly concave function,

$$\widehat{G}_K(\boldsymbol{\lambda}) = \frac{1}{N}\sum_{i=1}^N \rho\left\{\boldsymbol{\lambda}^\top u_K(T_i, \boldsymbol{X}_i)\right\} - \frac{1}{N(N-1)}\sum_{j=1, j\neq i}^N \sum_{i=1}^N \boldsymbol{\lambda}^\top u_K(T_i, \boldsymbol{X}_j),$$

where $\rho(v) = D\{(D'^{-1}(1-v, 1), 1\} + v - v\cdot(D'^{-1}(1-v, 1)$ is a strictly increasing and concave function and $\rho'(\cdot)$ is the first derivative of $\rho(\cdot)$.

Remark 2. *Our approach to estimating the weight function differs slightly from the one in Ai et al. (2021). Ai et al. (2021) considered the product sieve* $u_K(T, \boldsymbol{X}) = w_{K_1}(T) \otimes m_{K_2}(\boldsymbol{X})$ *and exploited the product-moment condition,*

$$\mathbb{E}\left\{\pi_0(T, \boldsymbol{X}) w_{K_1}(T) m_{K_2}^\top(\boldsymbol{X})\right\} = \mathbb{E}\left\{w_{K_1}(T)\right\} \cdot \mathbb{E}\left\{m_{K_2}^\top(\boldsymbol{X})\right\},$$

where $w_{K_1}(T)$ *and* $m_{K_2}(\boldsymbol{X})$ *are approximation sieves of* T *and* $\boldsymbol{X}$ *respectively. They estimated* $\pi_0(T_i, \boldsymbol{X}_i)$ *by minimizing the KL distance from the uniform weight, subject to the product-moment condition,*

$$\begin{cases} \{\widehat{\pi}_i\}_{i=1}^N = \operatorname{argmin} \sum_{i=1}^N \pi_i \log \pi_i \\ \textit{subject to } \frac{1}{N}\sum_{i=1}^N \pi_i w_{K_1}(T_i) m_{K_2}^\top(\boldsymbol{X}_i) = \left\{\frac{1}{N}\sum_{i=1}^N w_{K_1}(T_i)\right\}\left\{\frac{1}{N}\sum_{j=1}^N m_{K_2}^\top(\boldsymbol{X}_j)\right\} \end{cases}.$$

Their dual solution is $\widehat{\pi}(T_i, \boldsymbol{X}_i) = \exp(-w_{K_1}^\top(T_i)\widehat{\Lambda}_{K_1\times K_2}\, m_{K_2}(\boldsymbol{X}_i) - 1)$*, with* $\widehat{\Lambda}_{K_1\times K_2}$ *maximizing the following objective function*

$$\widehat{\Lambda}_{K_1\times K_2} = \underset{\Lambda\in\mathbb{R}^{K_1\times K_2}}{\operatorname{argmax}} \Bigg\{ -\frac{1}{N}\sum_{i=1}^N \exp\left(-w_{K_1}^\top(T_i)\widehat{\Lambda}_{K_1\times K_2} m_{K_2}(\boldsymbol{X}_i) - 1\right) - \left(\frac{1}{N}\sum_{i=1}^N w_{K_1}^\top(T_i)\right)\Lambda\left(\frac{1}{N}\sum_{j=1}^N m_{K_2}(\boldsymbol{X}_j)\right)\Bigg\}.$$

Their approach is technically difficult to extend to other types of distance measures. In contrast, our approach does not exploit the product-moment condition but permits general distance measure $D(\cdot,\cdot)$*, which encompasses the exponential tilting,* $D(v, 1) = v\log v$ *(Imbens et al. (1998)), the empirical likelihood,* $D(v, 1) = \log v$ *(Owen (1988)), and the quadratic programming,* $D(v, 1) = (v-1)^2/2$ *(Yiu and Su (2018)).*

Remark 3. *The advantage of the quadratic distance,* $D(v, 1) = (v-1)^2/2$*, is that it gives a closed-form solution,*

$$\widehat{\pi}_K(t, \boldsymbol{x}) = \left(\frac{1}{N(N-1)}\sum_{i=1}^N\sum_{j=1, i\neq j}^N u_K(T_i, \boldsymbol{X}_j) - \frac{1}{N}\sum_{i=1}^N u_K(T_i, \boldsymbol{X}_i)\right)^\top \times \left(\frac{1}{N}\sum_{i=1}^N u_K(T_i, \boldsymbol{X}_i)\, u_K^\top(T_i, \boldsymbol{X}_i)\right)^{-1} u_K(t, \boldsymbol{x}) + 1.$$

Remark 4. *By construction,* $\widehat{\pi}_K(T_i, \boldsymbol{X}_i)$ *is positive for all* i*. Moreover, by the first order condition,* $\widehat{\pi}_K(T_i, \boldsymbol{X}_i)$ *satisfies the sample analog of (3.2):*

$$\frac{1}{N}\sum_{i=1}^N \widehat{\pi}_K(T_i, \boldsymbol{X}_i) u_K(T_i, \boldsymbol{X}_i) = \frac{1}{N(N-1)}\sum_{i=1, i\neq j}^N \sum_{j=1}^N u_K(T_i, \boldsymbol{X}_j).$$

With the constant one included in the sieve basis functions, $\widehat{\pi}_K(T_i, \boldsymbol{X}_i)$ *also satisfies*

$$\frac{1}{N}\sum_{i=1}^N \widehat{\pi}_K(T_i, \boldsymbol{X}_i) = 1.$$

These restrictions prevent the extreme weights.

In the second stage, we substitute the weight function with the estimates and apply the local linear regression to estimate the conditional-response function, $g_0(t,z)$, and its partial derivative with respect to t,

$$(\widehat{g}(t,z), \widehat{\partial_t g}(t,z)) = \underset{(\alpha,\beta)}{\operatorname{argmin}} \sum_{i=1}^{N} \widehat{\pi}_K(T_i, \boldsymbol{X}_i)\mathcal{L}\left\{Y_i - \alpha - (T_i - t)\beta\right\}\mathcal{K}_{\boldsymbol{h}}\left(T_i - t, Z_i - z\right),$$

where $\boldsymbol{h} = (h_0, h_1, ..., h_{d_Z})$ denotes the bandwidth, $\mathcal{K}_{\boldsymbol{h}}\left(t,z\right) = \mathcal{K}(t/h_0) \cdot \prod_{k=1}^{d_Z} \mathcal{K}(z_k/h_k)$, and $\mathcal{K}(\cdot)$ is a univariate kernel function.

Remark 5. *If T is discrete, the estimate of $g_0(t,z)$ simplifies to*

$$\widehat{g}(t,z) = \underset{a\in\mathbb{R}}{\operatorname{argmin}} \sum_{i=1}^{N} \widehat{\pi}_K(T_i, \boldsymbol{X}_i)\mathcal{L}\left(Y_i - a\right)\mathbb{1}(T_i = t)\mathcal{K}_{\boldsymbol{h}_Z}\left(Z_i - z\right) \text{ for all } (t,z) \in \mathcal{T}\times\mathcal{Z},$$

where $\boldsymbol{h}_Z = (h_1, ..., h_{d_Z})^\top$ is the bandwidth, $\mathcal{K}_{\boldsymbol{h}_Z}\left(z\right) = \prod_{k=1}^{d_Z} \mathcal{K}(z_k/h_k)$.

Remark 6. *The sieve generalized empirical likelihood (GEL) approach proposed in Chen et al. (2019) can be adapted to estimate the conditional-response function. Note that $g_0(T,Z)$ solves the following conditional-moment restriction,*

$$\mathbb{E}\left[\pi_0(T,\boldsymbol{X})\mathcal{L}'(Y - g_0(T,Z))|T,Z\right] = 0,$$

which implies

$$\mathbb{E}\left[\pi_0(T,\boldsymbol{X})\mathcal{L}'(Y - g_0(T,Z))v_{K_1}(T,Z)\right] = 0,$$

where $\{v_{K_1}(T,Z)\}$ is a K_1-dimensional approximating sieve. Let $s(\cdot)$ be a strictly concave and twice-continuously differentiable function, with Lipschitz-continuous second derivative and $s'(0) = s''(0) = -1$. Following Chen et al. (2019), the sequential GEL estimator of $g_0(t,z)$ is given by

$$\widehat{g}_{GEL}(t,z) = \widehat{\alpha}_{K_0}^\top v_{K_0}(t,z),$$

where $\widehat{\alpha}_{K_0} \in \mathbb{R}^{K_0}$ solves the following saddle point optimization problem:

$$\widehat{\alpha}_{K_0} = \arg \min_{\alpha\in\mathbb{R}^{K_0}} \max_{\lambda\in\mathbb{R}^{K}} \frac{1}{N}\sum_{i=1}^{N} s\left\{\widehat{\pi}_K(T_i, \boldsymbol{X}_i)\mathcal{L}'(Y_i - \beta^\top v_{K_0}(T_i, Z)) \cdot \lambda^\top v_{K_1}(T_i, Z_i)\right\} - s(0).$$

The large sample properties of this estimator can be established but the derivation is more complicated than that in Chen et al. (2019).

4 Large Sample Properties

4.1 Notation

To aid exposition, we introduce some notation. For any matrix $\boldsymbol{A}$, $\|\boldsymbol{A}\|$ denotes the Euclidean norm of $\boldsymbol{A}$. For any measurable function f, $\|f\|_{L^2(X)} = (\mathbb{E}[|f(\boldsymbol{X})|^2])^{1/2}$ denotes the $L^2(\boldsymbol{X})$ norm. For any sequences $a_N, b_N \in \mathbb{R}^+$, $a_N \lesssim b_N$ means $a_N \leq Cb_N$ for some constant $C > 0$ for all N, $a_N \prec b_N$ means $a_N/b_N \to 0$ as $N \to \infty$, and $a_N \sim b_N$ means a_N/b_N bounded and bounded away from zero for all N. For any $a, b \in \mathbb{R}$, denote $a \wedge b = \min(a,b)$ and $a \vee b = \max(a,b)$.

4.2 Assumptions

The following conditions are maintained throughout the remainder of the paper.

Assumption 2. *$\{T_i, \boldsymbol{X}_i, Y_i\}_{i=1}^N$ are observations drawn independently from the distribution of $(T, \boldsymbol{X}, Y)$. The support $\mathcal{T} \times \mathcal{X}$ of $(T, \boldsymbol{X})$ is compact.*

Assumption 3. *There exist two positive constants $\underline{\eta}$ and $\overline{\eta}$ such that $0 < \underline{\eta} \leq \pi_0(t, \boldsymbol{x}) \leq \overline{\eta} < \infty$ for all $(t, \boldsymbol{x}) \in \mathcal{T} \times \mathcal{X}$.*

Assumption 4. *$\rho(v)$ is strictly concave in $v \in \mathbb{R}$.*

Assumption 5. *With $h_{pro} = \prod_{k=0}^{d_Z} h_k$, suppose that $h_{pro} \sim N^{-c_1}$ and $K \sim N^{c_2}$ hold for some constants $c_1, c_2 > 0$.*

Assumption 6. *Let $\rho'^{-1}(\cdot)$ be the inverse function of the derivative of $\rho(\cdot)$. There exist $\boldsymbol{\lambda}_K \in \mathbb{R}^K$ and a positive constant $\omega > 0$ such that $\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} |\{\rho'^{-1}(\pi_0(t, \boldsymbol{x})) - \ \boldsymbol{\lambda}_K^\top u_K(t, \boldsymbol{x})| = O(K^{-\omega})$.*

Assumption 7. *The eigenvalues of $\mathbb{E}[u_K(T, \boldsymbol{X}) u_K^\top(T, \boldsymbol{X})]$ are bounded from above and away from zero uniformly in K.*

Assumption 8. *With $\zeta(K) = \sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} \|u_K(t, \boldsymbol{x})\|$, suppose that $\zeta(K)(K^{-\omega} + \sqrt{K/N}) \to 0$, $\sum_{s=0}^{d_Z} h_s^2 \to 0$, and $N h_{pro} h_0^2 \to \infty$ as $N \to \infty$.*

Assumption 2 restricts the random variables to be bounded. This condition is restrictive but convenient for large-sample derivations. Assumption 3 requires the weight function to be bounded and bounded away from zero. This condition is commonly imposed in the covariate-balancing literature. Assumption 4 permits a wide class of estimators, including the exponential tilting, $\rho(v) = -\exp(-v - 1)$, as a special case. Assumption 6 requires the sieve approximation errors to shrink to zero at polynomial rates. Assumption 7 rules out near multicollinearity in the approximating basis functions. This condition is familiar in the sieve regression literature(Chen, 2007). Assumption 8 restricts the growth rate of K to ensure consistency of the proposed estimator. Newey (1997) derived $\zeta(K) = O(K)$ for the power series and $\zeta(K) = O(\sqrt{K})$ for the B-splines.

Under the above conditions, we first compute the convergence rates under the mean-squared error norm.

Theorem 1. *Suppose that Assumptions 1-8 hold. We obtain*

$$\int |\widehat{g}(t, z) - g_0(t, z)|^2 \, dF_{T,Z}(t, z) = O_p\left(\left\{\frac{1}{N h_{pro}} + \left(\sum_{s=0}^{d_Z} h_s^2\right)^2\right\} + \left\{K^{-2\omega} + \frac{K}{N}\right\}\right) \text{ and}$$

$$\int \left|\widehat{\partial_t g}(t, z) - \partial_t g_0(t, z)\right|^2 dF_{T,Z}(t, z) = O_p\left(\left\{\frac{1}{N h_{pro} h_0^2} + \left(\sum_{s=0}^{d_Z} h_s^2\right)^2\right\} + \left\{K^{-2\omega} + \frac{K}{N}\right\}\right).$$

We now turn to the large sample properties of $\widehat{g}(t,z) - g_0(t,z)$ and $\widehat{\partial_t g}(t,z) - \partial_t g_0(t,z)$ for any $(t,z) \in \mathcal{T} \times \mathcal{Z}$. Denote $\phi(a|t,z) = \mathbb{E}\{\mathcal{L}(Y-a)|T=t, Z=z\}$ and $\widetilde{\phi}(a|t,z) = \mathbb{E}\{\pi_0(T,\boldsymbol{X})\mathcal{L}(Y-a)|T=t, Z=z\}$.

Assumption 9. •

(i) *$\phi(a|t,z)$ and $\widetilde{\phi}(a|t,z)$ are twice continuously differentiable with respect to a, and all derivatives are continuous in $(t,z) \in \mathcal{T} \times \mathcal{Z}$ for every fixed a. Moreover, $\inf_{(t,z)\in\mathcal{T}\times\mathcal{Z}} \partial_a^2 \widetilde{\phi}(g_0(t,z)|t,z) > 0$.*

(ii) *The kernel function $\mathcal{K}(v) \geq 0$ has a bounded support and $\kappa_{j1} := \int_{-\infty}^{\infty} v^j \mathcal{K}(v)dv$ satisfies $\kappa_{0,1} = 1$, $\kappa_{1,1} = 0$, and $\kappa_{j,1} < \infty$ for $j = 2,3,4$.*

(iii) *The density function $f_{T,Z}(t,z)$ is continuous and satisfies $\inf_{(t,z)\in\mathcal{T}\times\mathcal{Z}} f_{T,Z}(t,z) > 0$.*

(iv) *The conditional density function $f_{Y|T,Z}(y|t,z)$ is continuous in (t,z) for every $y \in \mathbb{R}$. There exists a positive constant $\varepsilon > 0$ and a positive function $G(y|t,z)$ such that $\sup_{\|(t',z')-(t,z)\|\leq\epsilon} f_{Y|T,Z}(y|t',z') \leq G(y|t,z)$ holds for almost all y, and that*

$$\int \{\mathcal{L}(y-\delta) - \mathcal{L}(y) + \mathcal{L}'(y)\cdot\delta\}^2 G(y|t,z)dy = O(\delta^4) \text{ as } \delta \to 0$$

holds uniformly over $(t,z) \in \mathcal{T} \times \mathcal{Z}$.

(v) *The second derivatives $\partial_a^2 \phi(g_0(t,z)|t,z)$ and $\partial_a^2 \widetilde{\phi}(g_0(t,z)|t,z)$, $f_{T,Z}(t,z)$, and $g_0(t,z)$ are continuously differentiable in (t,z), with derivatives uniformly bounded over $(t,z) \in \mathcal{T} \times \mathcal{Z}$.*

Assumption 10. •

(i) *There exists a sufficiently large constant, M, such that*

$$\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}} \{|g_0(t,z) - t\cdot\partial_t g_0(t,z)| \vee |\partial_t g_0(t,z)|\} \leq M.$$

The function classes $\mathcal{F}_1 = \{(y,t) \mapsto \mathcal{L}(y-\alpha-\beta\cdot t) : \alpha,\beta \in [-M,M]\}$ and $\mathcal{F}_2 = \{(y,t) \mapsto \mathcal{L}'(y-\alpha-\beta\cdot t) : a,b \in [-M,M]\}$ are of VC-type in the sense that, for some constants A and v,

$$\sup_Q \mathcal{N}(\mathcal{F}_j, \|\cdot\|_{Q,2}, \tau\|F_j\|_{Q,2}) \leq \left(\frac{A}{\tau}\right)^v, \ j = 1,2,$$

where $\mathcal{N}(\mathcal{F}_j, \|\cdot\|_{Q,2}, \varepsilon)$ is the covering number of $\mathcal{F}_j$ at scale ε with respect to the norm $\|\cdot\|_{Q,2}$, F_j is an envelope function for $\mathcal{F}_j$, and the supreme of Q is taken over all finitely discrete distributions on $\mathbb{R}^2$.

(ii) *$\mathbb{E}\left[\sup_{\alpha,\beta\in[-M,M]} |\mathcal{L}'(Y-\alpha-\beta\cdot T)|^q\right] < \infty$ for some $q > 2$.*

Assumption 11. *(i) $\zeta(K)^2 K \prec N^{1/2} \prec K^\omega$ and $\{\zeta(K)^2 K^{1-2\omega}\} \vee (K^2/N) \prec N^{1-2/q} h_{pro} h_0^2$; (ii) $\sum_{s=0}^{d_z} h_s^2 \prec (N h_{pro})^{-1/2}$.*

Assumption 9 is similar to Condition A of Fan et al. (1994). Assumption 10 is needed for applying Corollary 5.1 of Chernozhukov et al. (2014b), and it is satisfied by commonly used loss function. Assumption 11 is an under-smoothing condition. Denote $e(t,z) = \partial_a^2 \widetilde{\phi}(g_0(t,z)|t,z) f_{T,Z}(t,z)$,

$$\psi_{0,\boldsymbol{h}}(T_i, \boldsymbol{X}_i, Y_i; t, z) = \frac{1}{e(t,z)} \cdot \Big[\pi_0(T_i, \boldsymbol{X}_i) \mathcal{L}'(Y_i - g_0(t,z) - \partial_t g_0(t,z)(T_i - t)) - \mathbb{E}\{\pi_0(T_i, \boldsymbol{X}_i) \mathcal{L}'(Y_i - g_0(t,z) - \partial_t g_0(t,z)(T_i - t)) | T_i, \boldsymbol{X}_i\} \Big] \mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z)$$

and $$\psi_{1,\boldsymbol{h}}(T_i, \boldsymbol{X}_i, Y_i; t, z) = \frac{1}{e(t,z)\kappa_{21}} \cdot \Big[\pi_0(T_i, \boldsymbol{X}_i) \mathcal{L}'(Y_i - g_0(t,z) - \partial_t g_0(t,z)(T_i - t)) - \mathbb{E}\{\pi_0(T_i, \boldsymbol{X}_i) \mathcal{L}'(Y_i - g_0(t,z) - \partial_t g_0(t,z)(T_i - t)) | T_i, \boldsymbol{X}_i\} \Big] (T_i - t) \mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z).$$

Under these conditions, we establish the following asymptotic representation:

Theorem 2. *Under Assumptions 2-11, we obtain the linear (Bahadur) representation of* $\widehat{g}(t,z)$ *and* $\widehat{\partial_t g}(t,z)$ *uniformly over* (t,z)*:*

$$\sqrt{Nh_{pro}}\,\{\widehat{g}(t,z) - g_0(t,z)\} = \frac{1}{\sqrt{Nh_{pro}}} \sum_{i=1}^N \psi_{0,\boldsymbol{h}}(T_i, \boldsymbol{X}_i, Y_i; t, z) + R_{0,N}(t,z)$$

and

$$\sqrt{Nh_{pro}} h_0 \left\{ \widehat{\partial_t g}(t,z) - \partial_t g_0(t,z) \right\} = \frac{1}{\sqrt{Nh_{pro}} h_0} \sum_{i=1}^N \psi_{1,\boldsymbol{h}}(T_i, \boldsymbol{X}_i, Y_i; t, z) + R_{1,N}(t,z).$$

Moreover, we show that

$$\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}} |R_{0,N}(t,z)| = o_P\left(\{\log N\}^{-1/2}\right) \text{ and } \sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}} |R_{1,N}(t,z)| = o_P\left(\{\log N\}^{-1/2}\right).$$

Remark 7. *In the proof of Theorem 2, we show that the bias of of the estimators* $\widehat{g}(t,z)$ *and* $\widehat{\partial_t g}(t,z)$ *are of* $O(\sum_{s=0}^{d_Z} h_s^2)$ *uniformly over* $(t,z) \in \mathcal{T} \times \mathcal{Z}$*, which converge to zero faster than* $\{Nh_{pro}\}^{-1/2}$ *under Assumption 11 (ii).*

Denote the variances by

$$\sigma_{0,N}^2(t,z) = \frac{Var\left(\psi_{0,\boldsymbol{h}}(T_i, \boldsymbol{X}_i, Y_i; t, z)\right)}{h_{pro}} \text{ and } \sigma_{1,N}^2(t,z) = \frac{Var\left(\psi_{1,\boldsymbol{h}}(T_i, \boldsymbol{X}_i, Y_i; t, z)\right)}{h_{pro} h_0^2}.$$

Assumption 12. *There exists some* $\underline{C} > 0$ *such that* $\min_{(t,z)\in\mathcal{T}\times\mathcal{Z}} \sigma_{0,N}^2(t,z) \geq \underline{C}$ *and* $\min_{(t,z)\in\mathcal{T}\times\mathcal{Z}} \sigma_{1,N}^2(t,z) \geq \underline{C}$*.*

Applying the Lyapunov central limit theorem (CLT) and Theorem 2, we obtain the pointwise convergence in distribution,

$$\frac{\sqrt{Nh_{pro}}\{\widehat{g}(t,z) - g_0(t,z)\}}{\sigma_{0,N}(t,z)} \xrightarrow{d} \mathcal{N}(0,1) \text{ and } \frac{\sqrt{Nh_{pro}} h_0 \{\widehat{\partial_t g}(t,z) - \partial_t g_0(t,z)\}}{\sigma_{1,N}(t,z)} \xrightarrow{d} \mathcal{N}(0,1)$$

for every $(t, z) \in \mathcal{T} \times \mathcal{Z}$.

The heterogeneous treatment-effect function when the treatment level changed marginally, $\tau_0(t|z) = \partial_t g_0(t, z)$, is estimated by $\widehat{\tau}(t|z) = \widehat{\partial_t g}(t, z)$, whose asymptotic distribution follows from applying Theorem 2,

$$\frac{\sqrt{Nh_{pro}}h_0\{\widehat{\tau}(t|z) - \tau_0(t|z)\}}{\sigma_{1,N}(t, z)} \xrightarrow{d} \mathcal{N}(0, 1).$$

The heterogeneous treatment-effect function when the treatment level changed discretely from t_0 to t_1, $\tau(t_1, t_0|z) = g(t_1, z) - g(t_0, z)$, is estimated by $\widehat{\tau}(t_1, t_0|z) = \widehat{g}(t_1, z) - \widehat{g}(t_0, z)$, whose asymptotic distribution also follows from applying Theorem 2.

Corollary 3. *Under Assumptions 2-11, we obtain*

$$\begin{aligned}
&\sqrt{Nh_{pro}}\,\{\widehat{\tau}(t_1, t_0|z) - \tau_0(t_1, t_0|z)\} \\
=&\frac{1}{\sqrt{Nh_{pro}}}\sum_{i=1}^{N}\{\psi_{0,\boldsymbol{h}}(T_i, \boldsymbol{X}_i, Y_i; t_1, z) - \psi_{0,\boldsymbol{h}}(T_i, \boldsymbol{X}_i, Y_i; t_0, z)\} + R_N(t_1, t_0, z),
\end{aligned}$$

where

$$\sup_{(t_0,t_1,z)\in\mathcal{T}\times\mathcal{T}\times\mathcal{Z}} |R_N(t_0, t_1, z)| = o_P\left(\{\log N\}^{-1/2}\right).$$

5 Uniform Inference Bounds

The asymptotic distribution derived above can be used for (pointwise) statistical inference. Specifically, with $\tau_0(t|z)$ and $\tau_0(t_1, t_0|z)$ as hypothesized functions, denote

$$\begin{aligned}
\widehat{Z}_{1,N}(t, z) &= \frac{\sqrt{Nh_{pro}}h_0\{\widehat{\tau}(t|z) - \tau_0(t|z)\}}{\widehat{\sigma}_{1,N}(t, z)} \text{ and} \\
\widehat{Z}_{2,N}(t_1, t_0, z) &= \frac{\sqrt{Nh_{pro}}\,\{\widehat{\tau}(t_1, t_0|z) - \tau_0(t_1, t_0|z)\}}{\widehat{\sigma}_{2,N}(t_1, t_0, z)},
\end{aligned}$$

where $\widehat{\sigma}_{1,N}(t, z)$ and $\widehat{\sigma}_{2,N}(t_1, t_0, z)$ are respectively consistent estimates of $\sigma_{1,N}(t, z)$ and

$$\sigma_{2,N}(t_1, t_0, z) = \sqrt{\frac{Var\left(\psi_{0,\boldsymbol{h}}(T_i, \boldsymbol{X}_i, Y_i; t_1, z) - \psi_{0,\boldsymbol{h}}(T_i, \boldsymbol{X}_i, Y_i; t_0, z)\right)}{h_{pro}}}.$$

For a given level of significance α, let c_α denote the critical value of the standard normal distribution. Then for any given point (t, t_0, t_1, z), we reject $\tau_0(t|z)$ ($\tau_0(t_1, t_0|z)$) if

$$\left|\widehat{Z}_{1,N}(t, z)\right| > c_\alpha \;\left(\left|\widehat{Z}_{2,N}(t_1, t_0, z)\right| > c_\alpha\right).$$

The pointwise tests can be applied to all points. The hypothesized function is rejected if it is rejected by some pointwise tests. These tests are not powerful enough to test the hypothesized functions. A more powerful test is to find the critical values $C_{1\alpha}, C_{2\alpha}$ such that, under the null,

$$\lim_{N\to\infty} \mathbb{P}\left(\sup_{t,z}\left|\widehat{Z}_{1,N}(t,z)\right| < C_{1\alpha}\right) = \lim_{N\to\infty} \mathbb{P}\left(\sup_{t_1,t_0,z}\left|\widehat{Z}_{2,N}(t_1,t_0,z)\right| < C_{2\alpha}\right) = 1-\alpha.$$

To develop such test, we apply the wild-bootstrap algorithm. Specifically, let ξ denote a positive random variable, independent of $(T, \boldsymbol{X}, Y)$ with $\mathbb{E}[\xi] = 1$ and $Var(\xi) = 1$. The distribution of ξ has a sub-exponential tail. Let B be a positive integer and let $\{\xi_i^{(b)}\}_{i=1,b=1}^{N,B}$ be an $i.i.d.$ sequence from the distribution of ξ. For $b = 1, ..., B$ and every $(t, z) \in \mathcal{T} \times \mathcal{Z}$, we compute $\widehat{\pi}_K^{(b)}(T_i, \boldsymbol{X}_i) = \rho'\left\{\left(\widehat{\boldsymbol{\lambda}}_K^{(b)}\right)^\top u_K(T_i, \boldsymbol{X}_i)\right\}$, with $\widehat{\boldsymbol{\lambda}}_K^{(b)} \in \mathbb{R}^K$ the maximizer of

$$\widehat{G}_K^{(b)}(\boldsymbol{\lambda}) = \frac{1}{N}\sum_{i=1}^N \xi_i^{(b)}\rho\left\{\boldsymbol{\lambda}^\top u_K(T_i,\boldsymbol{X}_i)\right\} - \frac{1}{N(N-1)}\sum_{j=1,j\neq i}^N\sum_{i=1}^N \boldsymbol{\lambda}^\top u_K(T_i,\boldsymbol{X}_j).$$

We then compute

$$(\widehat{g}^{(b)}(t,z), \widehat{\tau}^{(b)}(t,z)) = \underset{(\alpha,\beta)}{\operatorname{argmin}} \sum_{i=1}^N \xi_i^{(b)}\widehat{\pi}_K^{(b)}(T_i,\boldsymbol{X}_i)\mathcal{L}\left\{Y_i - \alpha - \beta\cdot(T_i - t)\right\}\mathcal{K}_{\boldsymbol{h}}\left(T_i - t, Z_i - z\right)$$

and $\widehat{\tau}^{(b)}(t_1,t_0,z) = \widehat{g}^{(b)}(t_1,z) - \widehat{g}^{(b)}(t_0,z)$. Denote

$$Z_{1,N}^{(b)}(t,z) = \frac{1}{\sqrt{Nh_{pro}}h_0\sigma_{1,N}(t,z)}\sum_{i=1}^N\{\xi_i^{(b)} - 1\}\psi_{1,\boldsymbol{h}}(T_i,\ \boldsymbol{X}_i, Y_i; t, z)$$

and

$$Z_{2,N}^{(b)}(t_1,t_0,z) = \frac{1}{\sqrt{Nh_{pro}}\sigma_{2,N}(t_1,t_0,z)}\sum_{i=1}^N\{\xi_i^{(b)}-1\}\left(\psi_{0,\boldsymbol{h}}(T_i,\boldsymbol{X}_i,Y_i;t_1,z) - \psi_{0,\boldsymbol{h}}(T_i,\boldsymbol{X}_i,Y_i;t_0,z)\right).$$

The following theorem provides a theoretical foundation for computing the critical values.

Theorem 4. *Under Assumptions 2-11, we have, for* $b = 1, \ldots, B$,

$$\frac{\sqrt{Nh_{pro}}h_0\left\{\widehat{\tau}^{(b)}(t,z) - \widehat{\tau}(t,z)\right\}}{\sigma_{1,N}(t,z)} = Z_{1,N}^{(b)}(t,z) + R_{1,N}^{(b)}(t,z)$$

and

$$\frac{\sqrt{Nh_{pro}}\left\{\widehat{\tau}^{(b)}(t_1,t_0,z) - \widehat{\tau}(t_1,t_0,z)\right\}}{\sigma_{2,N}(t_1,t_0,z)} = Z_{2,N}^{(b)}(t_1,t_0,z) + R_{2,N}^{(b)}(t_1,t_0,z),$$

where

$$\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}|R_{1,N}^{(b)}(t,z)| = \sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}|R_{2,N}^{(b)}(t_1,t_0,z)| = o_P\left(\{\log N\}^{-1/2}\right).$$

Theorem 4 implies

$$\frac{\sqrt{Nh_{pro}}h_0\{\widehat{\tau}^{(b)}(t,z)-\widehat{\tau}(t,z)\}}{\sigma_{1,N}(t,z)}\xrightarrow{d}\mathcal{N}(0,1)$$

and

$$\frac{\sqrt{Nh_{pro}}\{\widehat{\tau}^{(b)}(t_1,t_0,z)-\widehat{\tau}(t_1,t_0,z)\}}{\sigma_{2,N}(t_1,t_0,z)}\xrightarrow{d}\mathcal{N}(0,1)$$

for every $(t,z)\in\mathcal{T}\times\mathcal{Z}$.

The following procedure constructs the uniform inference bands. A similar procedure can be implemented for the uniform inference of $\tau(t,z)$ and $\tau(t_1,t_0,z)$.

Uniform Confidence Band Implementation Procedure

1. Compute $\widehat{\tau}(t,z)$ for a suitably fine grid over $\mathcal{T}\times\mathcal{Z}$ and $\widehat{\tau}(t_1,t_0,z)$ for a suitably fine grid over $\mathcal{T}\times\mathcal{T}\times\mathcal{Z}$.

2. Consider a sufficiently large integer, B. Compute $\widehat{\tau}^{(b)}(t,z)$ and $\widehat{\tau}^{(b)}(t_1,t_0,z)$ over the same grid for $b=1,...,B$, with a new set of i.i.d. $2\times$ Bernoulli$(1/2)$ random variables $\{\xi_i^{(b)}\}_{i=1}^N$ in each step b.

3. Estimate $\sigma_{1,N}(t,z)$ by the sample standard deviation of $\sqrt{Nh_{pro}}h_0\{\widehat{\tau}^{(b)}(t,z)-\widehat{\tau}(t,z)\}_{b=1}^B$, denoted by $\widehat{\sigma}_{1,N}(t,z)$. Estimate $\sigma_{2,N}(t_1,t_0,z)$ by the sample standard deviation of $\sqrt{Nh_{pro}}\{\widehat{\tau}^{(b)}(t_1,t_0,z)-\widehat{\tau}(t_1,t_0,z)\}_{b=1}^B$, denoted by $\widehat{\sigma}_{2,N}(t_1,t_0,z)$.

4. For $b=1,...,B$, compute

$$M_{1b}^{1-sided}=\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}\frac{\sqrt{Nh_{pro}}h_0\{\widehat{\tau}^{(b)}(t,z)-\widehat{\tau}(t,z)\}}{\widehat{\sigma}_{1,N}(t,z)},$$
$$M_{1b}^{2-sided}=\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}\frac{\sqrt{Nh_{pro}}h_0|\widehat{\tau}^{(b)}(t,z)-\widehat{\tau}(t,z)|}{\widehat{\sigma}_{1,N}(t,z)},$$

and

$$M_{2b}^{1-sided}=\sup_{(t_1,t_0,z)\in\mathcal{T}\times\mathcal{T}\times\mathcal{Z}}\frac{\sqrt{Nh_{pro}}\{\widehat{\tau}^{(b)}(t_1,t_0,z)-\widehat{\tau}(t_1,t_0,z)\}}{\widehat{\sigma}_{2,N}(t_1,t_0,z)},$$
$$M_{2b}^{2-sided}=\sup_{(t_1,t_0,z)\in\mathcal{T}\times\mathcal{T}\times\mathcal{Z}}\frac{\sqrt{Nh_{pro}}|\widehat{\tau}^{(b)}(t_1,t_0,z)-\widehat{\tau}(t_1,t_0,z)|}{\widehat{\sigma}_{2,N}(t_1,t_0,z)},$$

where the supremum is approximated by the maximum over the chosen grid point.

5. Given a confidence level $1-\alpha$, find the empirical $1-\alpha$ quantile of the sets of number $\{M_{jb}^{1-sided}:b=1,...,B\}$ and $\{M_{jb}^{2-sided}:b=1,...,B\}$, which are denoted by $\widehat{C}_{j\alpha}^{1-sided}$ and $\widehat{C}_{j\alpha}^{2-sided}$ respectively for $j=1,2$.

6. The uniform confidence bands are constructed as

$$I_{1L} = \left\{ \left(\widehat{\tau}(t,z) - \widehat{C}_{1\alpha}^{1-sided} \frac{\widehat{\sigma}_{1,N}(t,z)}{\sqrt{Nh_{pro}}h_0}, \infty \right) : (t,z) \in \mathcal{T} \times \mathcal{Z} \right\},$$

$$I_{1R} = \left\{ \left(-\infty, \widehat{\tau}(t,z) + \widehat{C}_{1\alpha}^{1-sided} \frac{\widehat{\sigma}_{1,N}(t,z)}{\sqrt{Nh_{pro}}h_0} \right) : (t,z) \in \mathcal{T} \times \mathcal{Z} \right\},$$

$$I_{12} = \left\{ \left(\widehat{\tau}(t,z) - \widehat{C}_{1\alpha}^{2-sided} \frac{\widehat{\sigma}_{1,N}(t,z)}{\sqrt{Nh_{pro}}h_0}, \widehat{\tau}(t,z) + \widehat{C}_{1\alpha}^{2-sided} \frac{\widehat{\sigma}_{1,N}(t,z)}{\sqrt{Nh_{pro}}h_0} \right) : (t,z) \in \mathcal{T} \times \mathcal{Z} \right\},$$

$$I_{2L} = \left\{ \left(\widehat{\tau}(t_1,t_0,z) - \widehat{C}_{2\alpha}^{1-sided} \frac{\widehat{\sigma}_{2,N}(t_1,t_0,z)}{\sqrt{Nh_{pro}}}, \infty \right) : (t_1,t_0,z) \in \mathcal{T} \times \mathcal{Z} \right\},$$

$$I_{2R} = \left\{ \left(-\infty, \widehat{\tau}(t_1,t_0,z) + \widehat{C}_{2\alpha}^{1-sided} \frac{\widehat{\sigma}_{2,N}(t_1,t_0,z)}{\sqrt{Nh_{pro}}} \right) : (t_1,t_0,z) \in \mathcal{T} \times \mathcal{Z} \right\},$$

$$I_{22} = \left\{ \left(\widehat{\tau}(t_1,t_0,z) - \widehat{C}_{2\alpha}^{2-sided} \frac{\widehat{\sigma}_{2,N}(t_1,t_0,z)}{\sqrt{Nh_{pro}}}, \widehat{\tau}(t_1,t_0,z) + \widehat{C}_{2\alpha}^{2-sided} \frac{\widehat{\sigma}_{2,N}(t_1,t_0,z)}{\sqrt{Nh_{pro}}} \right) : (t_1,t_0,z) \in \mathcal{T} \times \mathcal{Z}$$

The following theorem establishes the asymptotic validity of the proposed confidence bands.

Theorem 5. *Under Assumptions 2-11, and suppose that $B \to \infty$ as $N \to \infty$, we have*

$$\lim_{N\to\infty} \mathbb{P}\left(\tau(\cdot,\cdot) \in I_{1L}\right) = \lim_{N\to\infty} \mathbb{P}\left(\tau(\cdot,\cdot) \in I_{1R}\right) = \lim_{N\to\infty} \mathbb{P}\left(\tau(\cdot,\cdot) \in I_{12}\right) = 1-\alpha$$

and

$$\lim_{N\to\infty} \mathbb{P}\left(\tau(\cdot,\cdot,\cdot) \in I_{2L}\right) = \lim_{N\to\infty} \mathbb{P}\left(\tau(\cdot,\cdot,\cdot) \in I_{2R}\right) = \lim_{N\to\infty} \mathbb{P}\left(\tau(\cdot,\cdot,\cdot) \in I_{22}\right) = 1-\alpha.$$

We can use the uniform confidence bands to test the hypothesized treatment-effect functions. For example, with $T \in \{0,1\}$, $Z = \boldsymbol{X}$, and $\mathcal{L}(v) = v^2$, we use the uniform confidence band to test the null,

$$H_0 : \tau(1,1,x) = \mathbb{E}[Y^*(1) - Y^*(0) | \boldsymbol{X} = x] = 0 \text{ for all } x \in \mathcal{X},$$

which is studied by Crump et al. (2008). For multivalued treatments, $T \in \mathcal{T} = \{0,1,...,J\}$, we use the uniform confidence band to test the null,

$$H_0 : \tau(t_1,t_0,z) = 0 \text{ for all } z \in \mathcal{Z} \text{ and all } t_1 \neq t_0 \in \{0,1,...,J\}.$$

For continuous treatments, we use the uniform confidence band to test the null,

$$H_0 : \tau(t,z) = 0 \text{ for all } (t,z) \in \mathcal{T} \times \mathcal{Z}. \quad (5.1)$$

The null is rejected at α significance if there exists a pair of $(t,z) \in \mathcal{T} \times \mathcal{Z}$ such that 0 is not contained in the $(1-\alpha)$ uniform confidence region.

6 Numerical Details

6.1 Basis function

In our numerical studies, we set the sieve basis u_K to be the product power series. Specifically, we define $u_K(T, \boldsymbol{X}) = w_{K_1}(T) \otimes w_{K_2}(\boldsymbol{X})$ and $K = (K_1+1)\cdot(K_2+1)$, where, for any positive integers p and r, $w_p : \mathbb{R}^r \mapsto \mathbb{R}^{pr+1}$:

$$w_p(\boldsymbol{v}) = (1, \boldsymbol{v}_1^{1:p}, \ldots, \boldsymbol{v}_r^{1:p})^\top ,$$

and $\boldsymbol{v}_j^{1:p} = (\boldsymbol{v}_j, \boldsymbol{v}_j^2, \ldots, \boldsymbol{v}_j^p)$ for the jth element of $\boldsymbol{v}$ for any variable $\boldsymbol{v} \in \mathbb{R}^r$.

6.2 Selection of tuning parameters

The tuning parameters K and $\boldsymbol{h}$ must satisfy some sufficient conditions for the asymptotic derivation. However, those conditions do not give a unique selection in practice. This section presents a data-driven approach to selecting the tuning parameters. Specifically, we randomly split the dataset into F sets, for an integer $2 \le F \le N$. In principle, we choose the tuning parameters for estimating g_0, $\widehat{K} = (\widehat{K}_1 + 1)\cdot(\widehat{K}_2 + 1)$ and $\widehat{\boldsymbol{h}}$, to minimize the F-fold GCV criteria

$$GCV(K_1, K_2, \boldsymbol{h}) = \sum_{j=1}^{F} \frac{1}{|S_j|}\left[\sum_{k\in S_j} \widehat{\pi}_K(T_k, \mathbf{X}_k)\mathcal{L}\{Y_k - \widehat{g}^{(-j)}(T_k, Z_k)\}\right]^2 \Big/ (1-K/N)^2 ,$$

over $(K_1, K_2, \boldsymbol{h}) \in \{1, \ldots, P\}^2 \times \Pi_{i=0}^{d_Z}\mathcal{H}_i$, for some positive integer P and set $\mathcal{H}_i$'s, where S_j denotes the jth set of the dataset, $|S_j|$ is the sample size of the jth set, and $\widehat{g}^{(-j)}$ is computed from the sample observations excluding the observations in S_j.

Remark 7 suggests that the rate of the optimal bandwidth, $h_j \asymp N^{-1/(4+d)}$, for $\widehat{g}$, and $N^{-1/(6+d)}$ for $\widehat{\partial_t g}$, where $d = d_Z + 1$, for $j = 0, \ldots, d_Z$. We thus suggest to set $\mathcal{H}_j = \mathrm{sd}_j \cdot N^{-1/(4+d)} \cdot [(\log(N)/3)^{-1}, \log(N)/3]$, where sd_j denotes the sample standard deviation of the corresponding covariate, $j = 0, \ldots, d_Z$. For estimating the derivative $\partial_t g_0$, we take $\widehat{K}$ the same as for g_0, but the bandwidths $\widetilde{\boldsymbol{h}} = \widehat{\boldsymbol{h}} \cdot N^{1/(4+d)} \cdot N^{-1/(6+d)}$. Finally, the sufficient conditions in Theorem 5 require under-smoothing for the uniform inference so we take $\widehat{h}_j \cdot N^{1/(4+d)} \cdot N^{-1/(4+d-c_j)}$ for the uniform CI of g_0, and $\widetilde{h}_j \cdot N^{1/(6+d)} \cdot N^{1/(6+d-e_j)}$ for $\partial_t g_0$, for some constants $0 \le c_j < 4+d$ and $0 \le e_j < 6+d$. The choice of the c_j's and e_j's are discussed in the next section.

7 Numerical Studies

7.1 Simulation study

We conduct a small-scale study on a continuous treatment model to assess the finite performance of the proposed estimators $\widehat{g}$ and $\widehat{\partial_t g}$. We consider three response functions. Specifically, let U_w, U_t and U_y be three independent standard normal random variables. The models are given by

$$Z \sim \text{Uniform}(-0.65, 0.65),\ W = 0.5Z + 0.5U_w,\ T = 0.1W + 0.1Z + 0.4U_t.$$
$$\textbf{DGPC0}: \quad Y^*(t) = Z + W + 0.5U_y,$$

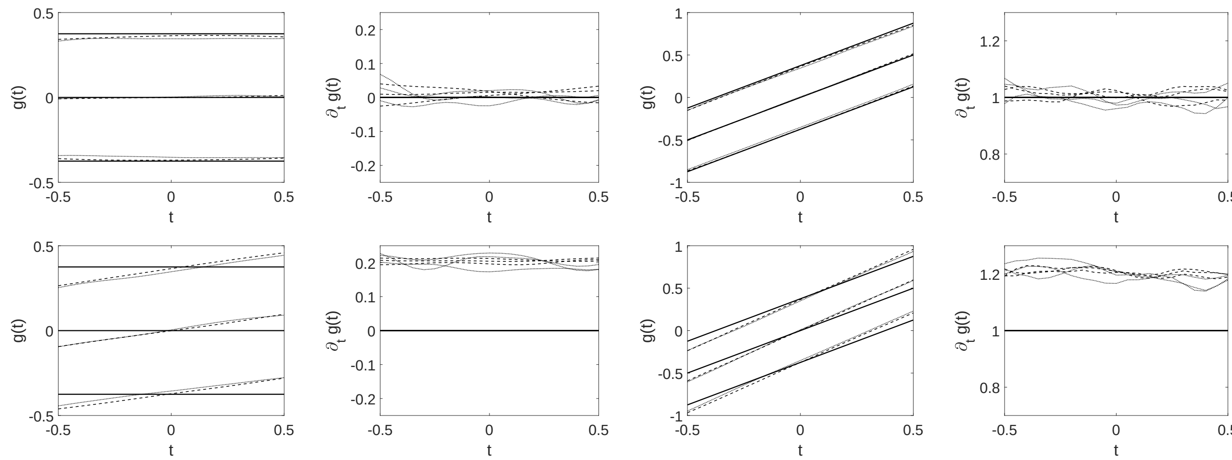

Figure 1: Plots of our estimators (row 1) and the naive estimators (row 2) of the conditional mean dose-response from DGPC0 (first two columns) and the conditional median dose-response from DGPC1L (last two columns), and their derivatives, $N = 200$ (dotted line) and $N = 600$ (dash-dotted line), for $z = \{-0.25, 0, 0.25\}$.

$$\begin{aligned} &\textbf{DGPC1L}: \quad Y^*(t) = t + Z + W + 0.5U_y \text{ (linear in } t\text{)},\\ &\textbf{DGPC1NL}: \quad Y^*(t) = t^2 + Z + W + 0.5U_y \text{ (non-linear in } t\text{)}. \end{aligned}$$

We generate $J = 200$ samples of size $N = 200$, 400 and 600 respectively. With $\mathcal{L}(v) = v^2$ and $\mathcal{L}(v) = v\{0.5 - \mathbb{1}(v \leq 0)\}$ respectively, we compute the estimators $\widehat{g}$ and $\widehat{\partial_t g}$ for the conditional average response $g_0(t, z) = \mathbb{E}\{Y^*(t)|Z = z\}$ and the conditional median response $g_0(t, z) = F^{-1}_{Y^*(t)|Z}(0.5|z)$, over the grid $(t, z) \in \mathcal{V} = \{-0.5, -0.45, \ldots, 0.45, 0.5\}^2$. Note that our estimators $\widehat{g}$ and $\widehat{\partial_t g}$ have closed-form solutions for the conditional average dose-response, but not for the conditional quantile dose-response function. The numerical minimization for estimating $g_0(t, z) = F^{-1}_{Y^*(t)|Z}(0.5|z)$ is time-consuming. To speed up the computation, we use the iteratively reweighted least squares algorithm proposed by Lejeune and Sarda (1988). We estimate π_0 using the closed-form solution in Remark 3 with $\rho(v) = -(v-1)^2/2$. Note that such a ρ choice does not guarantee positive $\widehat{\pi}_K$. We thus truncate any non-positive results to be the minimum positive value in $\{\widehat{\pi}_K(T_i, \boldsymbol{X}_i)\}_{i=1}^N$.

We compute the naive estimators, $\widehat{g}_{\mathrm{N}}$ and $\widehat{\partial_t g}_{\mathrm{N}}$, by setting $\widehat{\pi}_K \equiv 1$. We evaluate the performance of the estimators using the average squared errors (ASEs) and report the mean and standard deviation of the ASEs for each estimator from the 200 simulation samples in Table 1. The simulation results reveal that the proposed estimator consistently outperforms the naive estimator. Moreover, as the sample size increases, both the mean and standard deviation of the ASEs for the response function and its partial derivative decrease.

Figure 1 displays the bias of the proposed estimator and the naive estimator for the conditional average dose-response in model DGPC0 and the conditional median dose-response in model DGPC1L with sample sizes $N = 200$ and $N = 600$. The bias of the naive estimator is substantial since it does not control all the confounding factors. In contrast, the proposed estimator controls all the confounding factors and consequently, the bias decreases as the sample size increases.

Figure 2 plots the proposed estimator and the naive estimator of the conditional average re-

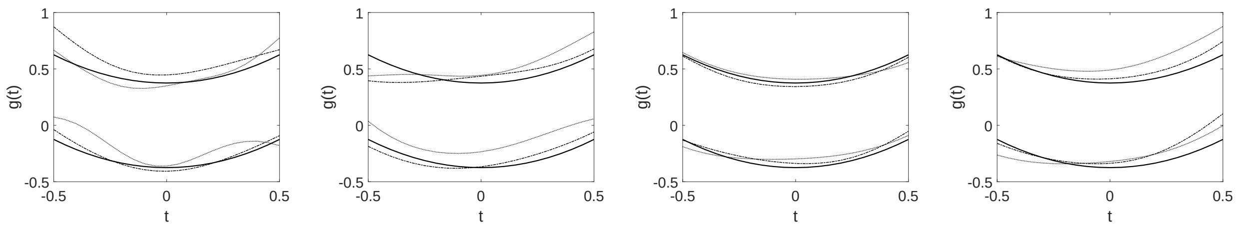

Figure 2: Plots of our estimators (1st & 3rd) and the naive estimators (2nd & 4th) of the conditional average dose-response $g_0(t, z)$ for $z = -0.25, 0.25$, from DGPC1NL, $N = 400$ (1st & 2nd) and $N = 600$ (3rd & 4th), corresponding to the 1st quartile (dash-dotted) and the 3rd quartile (dotted) of the 200 ASEs against the true curve (solid).

sponse in model DGPC1NL, with sample sizes $N = 400$ and $N = 600$. We present the estimators obtained from two random samples, corresponding to the 1st and the 3rd quartiles of the ASEs from the 200 simulated samples. The plot reveals that the variabilities of all the estimators decrease as the sample size increases. The proposed estimator converges to the true curve while the naive one shows a bias. In particular, the naive estimator tends to have a positive linear relationship between t and the true curve due to the confounding.

We also compute the uniform confidence bands for g_0 and $\partial_t g_0$ with the wild-bootstrap method with $B = N$. We take the under-smoothing bandwidth constants in Section 6.2 to be $c_0 = e_0 = 0$, $c_1 = 3.5, e_1 = 4.5$ for conditional average dose-response and $c_1 = 2.5, e_1 = 0$ for conditional median dose-response. Tables 2 reports the empirical coverage probability of the confidence band for g_0 at confidence levels $0.99, 0.95$ and 0.90, respectively. Table 3 shows the empirical rejection probability of H_0 in (5.1) at significance levels $0.01, 0.05$ and 0.1 respectively. The null hypothesis H_0 holds under DGPC0, but does not under DGPC1L and DGPC1NL. All three tables reveal that the size of the uniform inferences and the power of the tests are reasonably good.

Figure 3 depicts the uniform 95% confidence bands for each model computed from random samples with sizes $N = 200$ and $N = 600$ respectively. The simulation results also show that the selected bandwidths decrease as the sample size increases and the uniform confidence bands cover the true curves.

8 Empirical Study

In this section, we apply the proposed estimation method to analyze a dataset from the U.S. presidential campaign to understand how the number of political advertisements aired causally affects campaign contributions in non-competitive states. The dataset, commonly utilized in continuous treatment effect literature, covers a range of 0 to 22379 across $N = 16265$ zip codes in non-competitive states (see e.g. Urban and Niebler, 2014; Fong et al., 2018; Ai et al., 2021; Huang et al., 2022).

The covariates $\boldsymbol{X}$ considered include $\log$(Population), the percentage of the population that is over 65 years old, $\log$(Median Family Income $+ 1$), the percentage of the Hispanic population, the percentage of the black population, $\log$(Population density $+ 1$), the percentage of college graduates and the indicator whether the area can commute to a competitive state. Additional

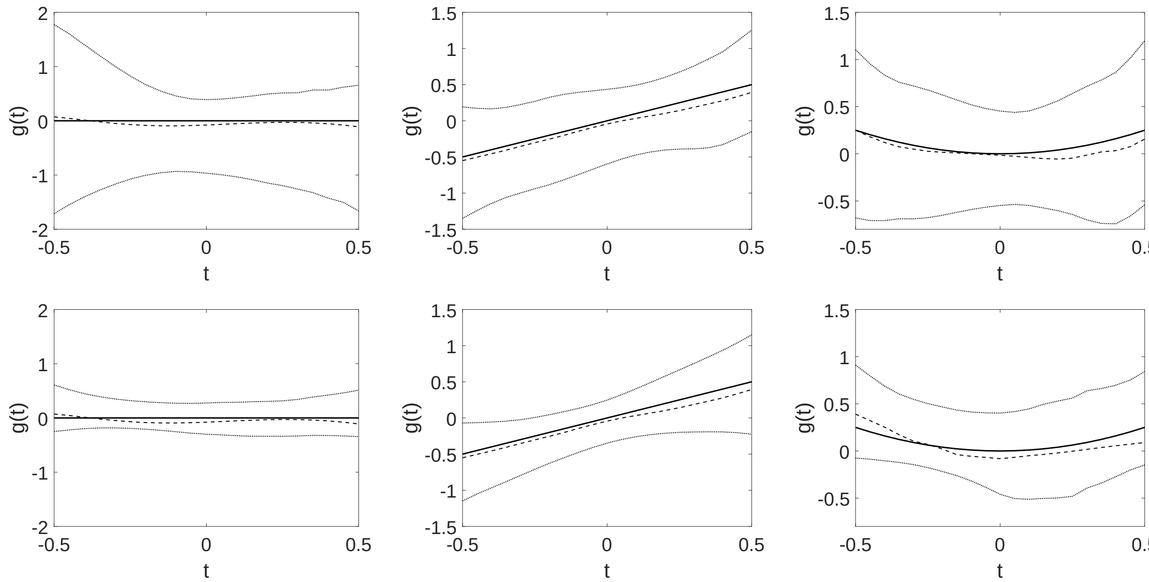

Figure 3: Plots of the true curve (solid), our estimator (dashed) of $g_0(t,z)$ for $z = 0$ with the uniform confidence band (dotted) from a simulated sample of DGPC0 (left), DGPC1L (centre) and DGPC1NL (right) with $N = 200$ (top) and $N = 600$ (bottom).

information can be found in Fong et al. (2018).

Urban and Niebler (2014) analyzed the causal relationship between advertising and campaign contributions from this data using a binary model. They compared the campaign contributions from the 5230 zip codes that received more than 1000 advertisements with those from the remaining 11035 zip codes that received fewer than 1000 advertisements. Their analysis revealed a significant causal effect of advertising in non-competitive states on the level of campaign contributions.

In contrast, Ai et al. (2021) treated the number of political advertisements as continuous and assumed a linear model on the average response function. They applied log transformations to both the outcome and treatment variables: $\log(\text{Contribution}+1)$ and $\log(\#\text{ads}+1)$, respectively, where #ads is the number of advertisements. Their analysis finds no significant causal effect of advertising on campaign contributions, which is consistent with the finding in Fong et al. (2018).

Huang et al. (2022) proposed a unified framework for the specification test of the continuous treatment effect models and rejected the linear model assumed in Ai et al. (2021). They recommended a Tobit model, combined with a Box-Cox transformation of the observed contributions and a composite log transformation of #ads, as a better fit for the data. Specifically, their Box-Cox transformation is defined as $\text{BoxCox}(\text{Contribution}, \lambda_1, \lambda_2) = \{(\text{Contribution}+\lambda_2)^{\lambda_1}-1\}/\lambda_1$ with $(\tilde{\lambda}_1, \tilde{\lambda}_2) = (0.1397, 0.0176)$. They then considered the observed outcome variable,

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

and the treatment variable, $T = \log(\log(\log(\#\text{ads}+1)+1)+2)$. Their estimated Tobit model showed that campaign contributions increase rapidly when $\#\text{ads} \in [0, 20]$, with improvements becoming marginal thereafter. However, Huang et al. (2022) did not give any inferential results for the study.

Table 1: $100 \times$ mean (standard deviation) of the ASEs from 200 Monte-Carlo simulations

	N	Method	DGPC0		DGPC1L		DGPC1NL	
			g	$\partial_t g$	g	$\partial_t g$	g	$\partial_t g$
Average	200	ours	1.94 (1.02)	10.06 (14.70)	1.93 (1.03)	10.30 (15.70)	2.17 (1.08)	17.88 (14.93)
		naive	2.17 (1.13)	14.08 (16.11)	2.23 (1.14)	14.23 (15.91)	2.40 (1.25)	21.57 (16.84)
	400	ours	1.00 (0.51)	4.92 (10.09)	1.03 (0.50)	4.91 (10.10)	1.22 (0.59)	13.55 (10.67)
		naive	1.25 (0.60)	8.30 (5.88)	1.24 (0.58)	8.33 (5.79)	1.49 (0.68)	16.42 (6.92)
	600	ours	0.70 (0.35)	2.58 (2.13)	0.69 (0.34)	2.48 (2.06)	0.89 (0.45)	11.31 (7.10)
		naive	0.97 (0.43)	6.56 (3.94)	0.98 (0.42)	6.66 (3.92)	1.18 (0.53)	14.68 (5.34)
Median	200	ours	2.72 (1.50)	22.01 (37.46)	2.73 (1.49)	22.05 (36.59)	2.94 (1.53)	28.81 (34.88)
		naive	2.92 (1.47)	23.89 (29.43)	2.94 (1.52)	24.03 (30.03)	3.12 (1.51)	31.60 (29.11)
	400	ours	1.46 (0.90)	13.07 (30.40)	1.47 (0.91)	12.98 (30.35)	1.67 (0.94)	20.80 (29.26)
		naive	1.70 (0.95)	15.88 (27.78)	1.72 (0.96)	15.78 (27.87)	1.97 (1.02)	24.32 (26.59)
	600	ours	1.02 (0.60)	8.52 (23.04)	1.01 (0.59)	8.50 (22.94)	1.20 (0.61)	16.36 (21.88)
		naive	1.29 (0.63)	12.05 (22.23)	1.30 (0.61)	11.98 (23.13)	1.50 (0.68)	19.74 (23.28)

Table 2: Empirical coverage probability of confidence band for g calculated from 200 Monte-Carlo simulations

	N	DGPC0			DGPC1L			DGPC1NL		
		99%	95%	90%	99%	95%	90%	99%	95%	90%
Average	200	1.000	0.970	0.930	1.000	0.975	0.935	0.995	0.955	0.900
	400	0.990	0.965	0.905	0.995	0.955	0.900	0.995	0.930	0.870
	600	0.995	0.955	0.895	0.995	0.950	0.875	0.985	0.935	0.865
Median	200	0.995	0.945	0.920	0.995	0.965	0.915	0.990	0.970	0.905
	400	1.000	0.970	0.930	1.000	0.980	0.920	0.990	0.950	0.885
	600	0.995	0.955	0.870	0.995	0.950	0.880	0.990	0.940	0.850

These conflicting findings from the previous studies suggest a complicated causal relationship between advertising and campaign contributions. We use the same transformations on the variables as Huang et al. (2022) and apply the proposed nonparametric method to estimate the conditional average response, $g_0(t,z) = \mathbb{E}\{Y^*(t)|Z = z\}$, where Z denotes either the Median family Income or the percentage of the black population. We estimate g_0 and $\partial_t g_0$, along with 95% uniform confidence bands.

We plot the estimated functions in Figure 4 using the median family income $z \in \$\{25, 30, 35, \ldots, 120\}$ thousands. For median family income above \$50 thousand, the estimated response function displays patterns resembling those of the unconditional average response function estimated by Huang et al. (2022). Specifically, advertising rapidly boosts campaign contributions when #ads≤ 20, with only marginal improvements beyond that point. The estimated derivatives and the associated 95% uniform confidence bands indicate significant causal effect of advertising on the campaign contributions for 13 out 20 family median income levels, mainly when the family median income is above \$50 thousands and $\#ad < 50$ (see e.g. the figure's second row). Furthermore, the figure's top second plot demonstrates that the estimated contributions tend to increase almost monotonically with the area's median family income.

However, Figure 5 shows the causal link between advertising and campaign contributions is not affected by the proportion of the black population.

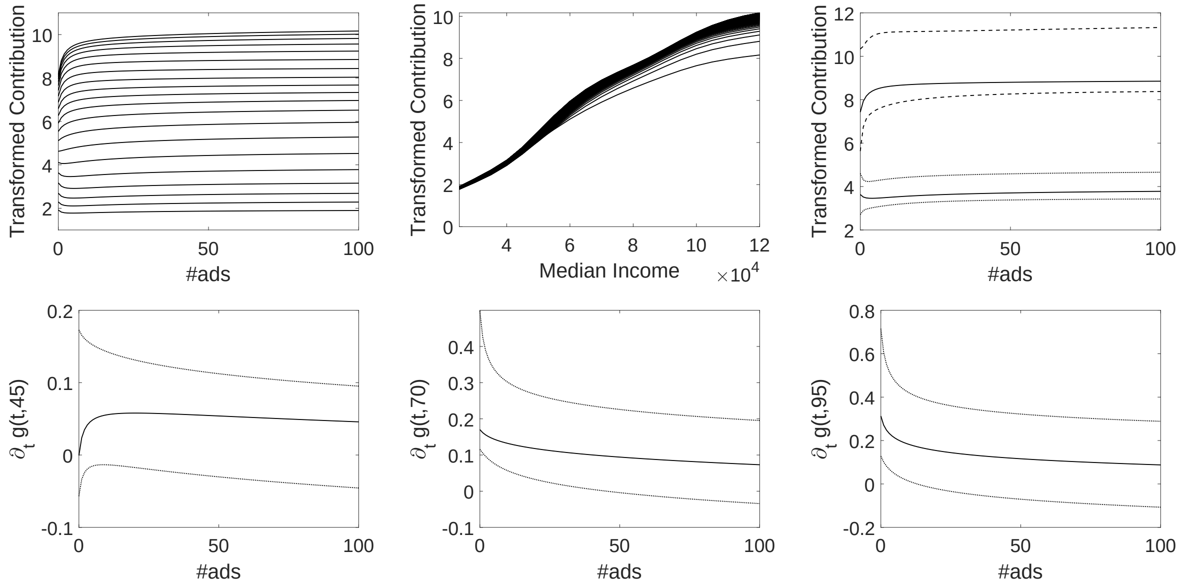

Figure 4: Plots of our estimators of $g_0(t, z)$ (top) for $\#$ads $\in [0, 100]$ and Median Income $=$ $\$\{25, 30, \ldots, 120\}$ thousands (top 1 and 2), and for Median Income $= \$45$ thousands (solid) and $\$95$ thousands (dash-dotted), associated with 95% uniform confidence bands (UCB) (top 3); Plots of our estimators of $\partial_t g_0(t, z)$ with 95% UCB (bottom) for Median Income $= \$\{45, 70, 95\}$ thousands (corresponding to bottom 1, 2 and 3).

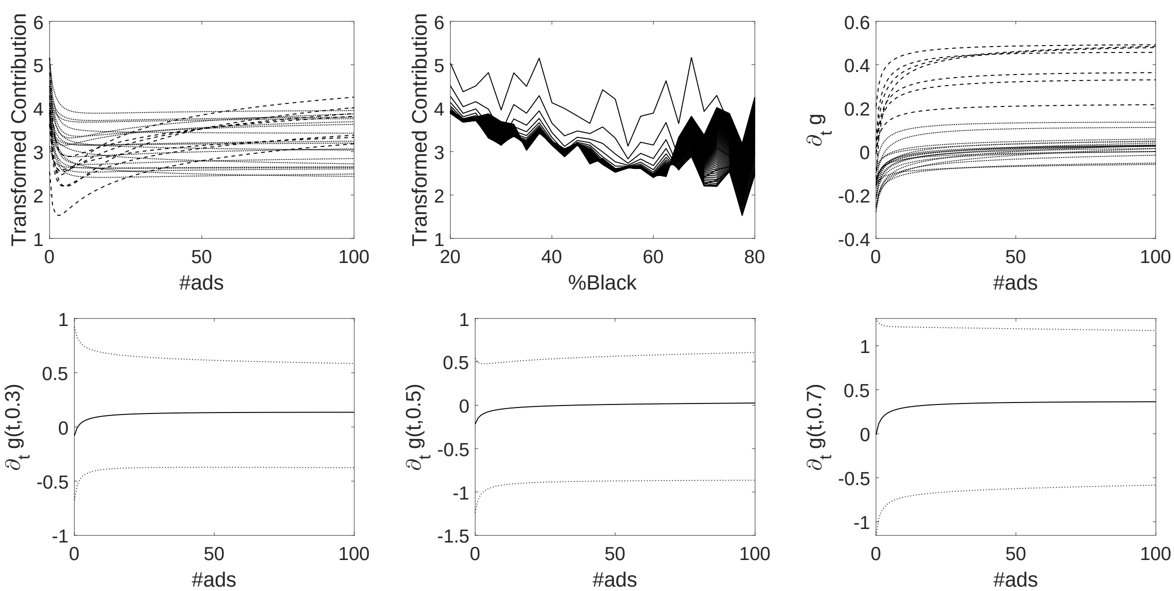

Figure 5: Plots of our estimators of $g(t, z)$ (top) (top 1 and 2), and our estimator of $\partial_t g_0(t, z)$ (top 3) for $\#$ads $\in [0, 100]$ and $\%$Black $= \$\{0.2, 0.225, \ldots 0.65\}$ (dotted) and $\{0.675, 0.7, \ldots, 0.8\}$ (dashed); Plots of our estimators of $\partial_t g(t, z)$ with 95% UCB (bottom) for $\%$Black $= \$\{0.3, 0.5, 0.7\}$ (corresponding to bottom 1, 2 and 3).

Table 3: Empirical rejection probability for $H_0 : \partial_t g(t) = 0$ for all $t \in \mathcal{T}$,calculated from 200 Monte-Carlo simulations

		DGPC0			**DGPC1L**			**DGPC1NL**		
	N	0.01	0.05	0.10	0.01	0.05	0.10	0.01	0.05	0.10
Average	200	0.005	0.015	0.060	0.780	0.945	0.975	0.075	0.245	0.395
	400	0.015	0.065	0.120	0.965	0.990	1.000	0.190	0.505	0.700
	600	0.010	0.045	0.095	0.995	1.000	1.000	0.410	0.695	0.850
Median	200	0.000	0.010	0.025	0.715	0.820	0.905	0.030	0.150	0.275
	400	0.000	0.055	0.100	0.930	0.965	0.980	0.260	0.470	0.570
	600	0.005	0.040	0.095	0.980	0.990	0.995	0.380	0.700	0.805

9 Concluding Remarks

To be added...

Acknowledgments

The first author, Chunrong Ai.... The second author, Wei Huang's research was supported by the Professor Maurice H. Belz Fund of the University of Melbourne. The last author, Zheng Zhang, acknowledges financial support from National Key R&D Program of China (No. 2022YFA1008300), and the fund for building world-class universities (disciplines) of Renmin University of China [project number KYGJC2022014]. We thank Haoze Hou for his great help on revising this manuscript.

Appendix

A Verification of Assumption 10

The existence of M is guaranteed by Assumption 2 combined with the continuity of $g_0(t, \boldsymbol{x})$ and $\partial_t g_0(t, \boldsymbol{x})$. When $\mathcal{L}(v) = v^2$, we verify the VC-type conditions for $\mathcal{F}_1$ and $\mathcal{F}_2$ imposed in Assumption 10(i). For fixed a and b, the subgraph of the function $(y, t) \mapsto \mathcal{L}(y - a - bt)$ satisfies

$$
\begin{aligned}
&\{(y, t, u) : u \leq \mathcal{L}(y - a - bt)\} = \{(y, t, u) : u \leq (y - a - bt)^2\} \\
=&\{(y, t, u) : y - a - bt \geq \sqrt{u \vee 0}\} \cup \{(y, t, u) : y - a - bt \leq -\sqrt{u \vee 0}\} \in \mathcal{G} \cup \mathcal{G},
\end{aligned}
$$

where $\mathcal{G}$ is the collection of the negativity set of the vector space spanned by the four measurable functions $(y, t, u) \mapsto y$, $(y, t, u) \mapsto t$, $(y, t, u) \mapsto \sqrt{u \vee 0}$ and $(y, t, u) \mapsto 1$, i.e.

$$
\mathcal{G} = \{\{(y, t, u) : x_1 \cdot y + x_2 \cdot t + x_3 \cdot \sqrt{u \vee 0} + x_4 \leq 0\} : x_1, x_2, x_3, x_4 \in \mathbb{R}\}.
$$

By Van Der Vaart and Wellner (1996, Lemma 2.6.15, Lemma 2.6.18 (iii)), $\mathcal{G}$ is a VC class of set. Then by Van Der Vaart and Wellner (1996, Lemma 2.6.17 (iii)), $\mathcal{G} \cup \mathcal{G}$ is also a VC class of set. Hence, $\mathcal{F}_1$ in Assumption10(i) is a VC-subgraph class. By Van Der Vaart and Wellner (1996, Theorem 2.6.7), the VC-type condition for $\mathcal{F}_1$ imposed in Assumption 10(i) holds. Similarly, the VC-type condition for $\mathcal{F}_2$ imposed in Assumption 10(i) holds.

When $L(v) = v\{\tau - \mathbb{1}(v \leq 0)\}$, the VC-type conditions for $\mathcal{F}_1$ and $\mathcal{F}_2$ can be similarly verified.

B Key Lemmas

In this section, we list the key lemmas that will help establish our main theorems. The proofs of these lemmas will be presented in the next section.

B.1 Notations

To aid the presentation of the lemmas, we define the following quantities and notations. Define

$$G_K^*(\boldsymbol{\lambda}) := \mathbb{E}\left[\widehat{G}_K(\boldsymbol{\lambda})\right] = \mathbb{E}\left[\rho\left\{\boldsymbol{\lambda}^\top u_K(T,\boldsymbol{X})\right\}\right] - \mathbb{E}\left[\boldsymbol{\lambda}^\top u_K(T_1,\boldsymbol{X}_2)\right], \ \boldsymbol{\lambda}_K^* := \underset{\boldsymbol{\lambda}}{\operatorname{argmax}}\, G_K^*(\boldsymbol{\lambda}),$$
$$\pi_K^*(T,\boldsymbol{X}) := \rho'\left(\{\boldsymbol{\lambda}_K^*\}^\top u_K(T,\boldsymbol{X})\right), \ \pi_K(T,\boldsymbol{X}) := \rho'\left(\{\boldsymbol{\lambda}_K\}^\top u_K(T,\boldsymbol{X})\right),$$

where $\boldsymbol{\lambda}_K$ is given in Assumption 6.

Let $\phi_N(T,\boldsymbol{X},Y;t)$ be a random variable and $R(T,\boldsymbol{X},Y;\boldsymbol{\theta},t) := \mathcal{L}'(Y - \boldsymbol{\theta}^\top \boldsymbol{S}_t)\phi_N(T,\boldsymbol{X},Y;t)$, for $\boldsymbol{\theta}\in[-M,M]^2$ and $\boldsymbol{S}_t = (1, T-t)^\top$, such that

1. $\phi_N(T,\boldsymbol{X},Y;t)$ is Lipschitz in t such that $|\phi_N(T,\boldsymbol{X},Y;t) - \phi_N(T,\boldsymbol{X},Y;s)| \leq |t-s|\Phi_N(T,\boldsymbol{X},Y)$ for some measurable function Φ_N for every $(T,\boldsymbol{X},Y)\in\mathcal{T}\times\mathcal{X}\times\mathbb{R}$ and N;
2. $\mathbb{E}\{\phi_N(T,\boldsymbol{X},Y;t)|T=t',\boldsymbol{X}=\boldsymbol{x}'\}$ is continuously differentiable w.r.t. $(t',\boldsymbol{x}')\in\mathcal{T}\times\mathcal{X}$;
3. $\mathbb{E}\left\{\sup_{t\in\mathcal{T}}|\phi_N(T,\boldsymbol{X},Y;t)|^q\right\} \sim N^{qc_0}$ for some $0\leq c_0<1$.
4. $\sup_{(\boldsymbol{\theta},t,z)}\mathbb{E}\left[|R(T,\boldsymbol{X},Y;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T-t,Z-z)|^2\right] \sim h_{pro}$.

Furthermore, we let

$$\mathcal{G}_N = \left\{(t',\boldsymbol{x}',y)\mapsto R(t',\boldsymbol{x}',y;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(t'-t,z'-z) : (\boldsymbol{\theta},t,z)\in[-M,M]^2\times\mathcal{T}\times\mathcal{Z}\right\}.$$

B.2 Lemmas

The following lemmas establish the convergence rates of $\widehat{\pi}_K(t,\boldsymbol{x})$ under L^∞ and L^2. This lemma can be proved in the same way as Lemmas 2.1 and 2.2 in Ai et al. (2021) . Thus, the proof will be omitted.

Lemma 1. *Suppose Assumptions 2-8 hold. Then, as $N\to\infty$, we have*

$$\|\boldsymbol{\lambda}_K^* - \boldsymbol{\lambda}_K\| = O\left(K^{-\omega}\right), \quad \|\widehat{\boldsymbol{\lambda}}_K^{(b)} - \boldsymbol{\lambda}_K^*\| = O_p\left(\sqrt{\frac{K}{N}}\right);$$

$$\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}|\pi_0(t,\boldsymbol{x}) - \pi_K^*(t,\boldsymbol{x})| = O\left(\zeta(K)K^{-\omega}\right),$$

$$\int|\pi_0(t,\boldsymbol{x}) - \pi_K^*(t,\boldsymbol{x})|^2 dF_{T,X}(t,\boldsymbol{x}) = O\left(K^{-2\omega}\right),$$

$$\frac{1}{N}\sum_{i=1}^{N}|\pi_0(T_i,\boldsymbol{X}_i)-\pi_K^*(T_i,\boldsymbol{X}_i)|^2=O_p\left(K^{-2\omega}\right)\ ;$$

and

$$\sup_{(t,x)\in\mathcal{T}\times\mathcal{X}}|\widehat{\pi}_K^{(b)}(t,\boldsymbol{x})-\pi_K^*(t,\boldsymbol{x})|=O_p\left(\zeta(K)\sqrt{\frac{K}{N}}\right),$$

$$\int|\widehat{\pi}_K^{(b)}(t,\boldsymbol{x})-\pi_K^*(t,\boldsymbol{x})|^2dF_{T,X}(t,\boldsymbol{x})=O_p\left(\frac{K}{N}\right),$$

$$\frac{1}{N}\sum_{i=1}^{N}|\widehat{\pi}_K^{(b)}(T_i,\boldsymbol{X}_i)-\pi_K^*(T_i,\boldsymbol{X}_i)|^2=O_p\left(\frac{K}{N}\right),$$

for $b=0,1,\ldots,B$, *where* $\widehat{\lambda}_K^{(0)}:=\widehat{\lambda}_K$ *and* $\widehat{\pi}_K^{(0)}:=\widehat{\pi}_K$.

Lemma 2. *Under Assumptions 2-8, 10 and 11, $\mathcal{G}_N$ is a VC-type class. More concretely, there exist constants A and $v\geq e$ independent of N such that*

$$\sup_Q\mathcal{N}(\mathcal{G}_N,\|\cdot\|_{Q,2},\tau\|F_N\|_{Q,2})\leq\left(\frac{A}{\tau}\right)^v\quad 0<\tau<1\,,$$

for $b=0,1,\ldots,B$ and $\sup_Q$ is taken over all finitely discrete probability measures on $\mathcal{T}\times\mathcal{X}\times\mathbb{R}$, where F_N is the envolpe function of $\mathcal{G}_N$:

$$F_N(t',\boldsymbol{x}',y)=H\cdot\sup_{\boldsymbol{\theta}\in[-M,M]^2}|\mathcal{L}'(y-\boldsymbol{\theta}^\top(1,t'^\top)|\cdot\sup_{t\in\mathcal{T}}|\phi_N(t',\boldsymbol{x}',y;t)|\,,$$

where H is the supremum of $\mathcal{K}_{\boldsymbol{h}}(t'-t,z'-z)$ taken over all $(t,t',z,z')\in\mathcal{T}^2\times\mathcal{Z}^2$ and $\boldsymbol{h}>0$.

Lemma 3. *Suppose that Assumptions 2-8, 10 and 11 hold and $\{\zeta(K)^2K^{1-2\omega}\}\vee(K^2/N)\prec N^{1-2/q-2c_0}h_{pro}$. Then for $b=0,1,\ldots,B$, we have*

$$\begin{aligned}&\frac{1}{\sqrt{Nh_{pro}}}\sum_{i=1}^{N}\Big\{\xi_i^{(b)}\widehat{\pi}_K^{(b)}(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)\\&\qquad-\mathbb{E}[\pi_0(T,\boldsymbol{X})R(T_i,\boldsymbol{X}_i,Y_i;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T-t,Z-z)]\Big\}\\=&\frac{1}{\sqrt{Nh_{pro}}}\sum_{i=1}^{N}\xi_i^{(b)}\pi_0(T_i,\boldsymbol{X}_i)\Big[R(T_i,\boldsymbol{X}_i,Y_i;\boldsymbol{\theta},t)-\mathbb{E}\{R(T,\boldsymbol{X},Y;\boldsymbol{\theta},t)|T_i,\boldsymbol{X}_i\}\Big]\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)\\&\qquad+o_p\left\{(\log N)^{-1/2}\right\}\,,\end{aligned}$$

where $o_p\left\{(\log N)^{-1/2}\right\}$ holds uniformly in $(\boldsymbol{\theta},t,z)\in[-M,M]^2\times\mathcal{T}\times\mathcal{Z}$, and M is defined in Assumption 10.

Lemma 4. *Suppose that for every $(t,z)\in\mathcal{T}\times\mathcal{Z}$, $A_N(\theta,t,z)$ is a sequence of convex random functions in θ defined on an open convex set $\mathcal{S}$ in $\mathbb{R}^p$ for $p\in\mathbb{N}$. There exists a function $A(\theta,t,z)$ that is convex in θ such that $\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}|A_N(\theta,t,z)-A(\theta,t,z)|=o_P(1)$ for every fixed θ. Then for each compact set $K\subset\mathcal{S}$, we have*

$$\sup_{\theta\in K}\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}|A_N(\theta,t,z)-A(\theta,t,z)|=o_P(1).$$

C Proof of Lemmas 2 to 4

C.1 Proof of Lemma 2

Applying Proposition 3.6.12 in Giné and Nickl (2021), we have the class of functions, $\mathcal{H}_N = \{(t', z') \mapsto \mathcal{K}_{\boldsymbol{h}}(t'-t, z'-z) : (t,z) \in \mathcal{T} \times \mathcal{Z}\}$, is a VC-type class and there exist constants A_1 and $v_1 \geq e$ such that

$$\sup_Q \mathcal{N}(\mathcal{H}_N, \|\cdot\|_{Q,2}, \tau H) \leq \left(\frac{A_1}{\tau}\right)^{v_1} \quad 0 < \tau < 1\,.$$

Applying Theorem 2.7.11 in Van Der Vaart and Wellner (1996), we have, for the class of functions, $\mathcal{F}_N = \{(t', \boldsymbol{x}', y) \mapsto \phi_N(t', \boldsymbol{x}', y; t) : t \in \mathcal{T}\}$, the bracketing number

$$\sup_Q \mathcal{N}_{[]}(\mathcal{F}_N, \|\cdot\|_{Q,2}, 2\tau\|\Phi_N\|_{Q,2}) \leq \mathcal{N}(\mathcal{T}, |\cdot|, \tau) = \frac{|\mathcal{T}|}{\tau}\,,$$

where $|\mathcal{T}|$ is the length of the compact interval $\mathcal{T}$. Then using the relationship between the covering and the bracketing numbers (see e.g. page 84 of Van Der Vaart and Wellner (1996)), we have

$$\sup_Q \mathcal{N}(\mathcal{F}_N, \|\cdot\|_{Q,2}, \tau\|\Phi_N\|_{Q,2}) \leq \sup_Q \mathcal{N}_{[]}(\mathcal{F}_N, \|\cdot\|_{Q,2}, 2\tau\|\Phi_N\|_{Q,2}) \leq \frac{|\mathcal{T}|}{\tau}\,.$$

Thus, $\mathcal{F}_N$ is a VC-type class. Then using Assumption 10 and Corollary A.1 of Chernozhukov et al. (2014b), we have the results.

C.2 Proof of Lemma 3

Let $\xi_i^{(0)} := 1$ for $i = 1, \ldots, N$ and $\widehat{G}_K^{(0)} := \widehat{G}_K$. The first order condition $\nabla \widehat{G}_K^{(b)}(\widehat{\boldsymbol{\lambda}}_K^{(b)}) = 0$ implies

$$\frac{1}{N}\sum_{i=1}^{N} \xi_i^{(b)} \rho'\left((\widehat{\boldsymbol{\lambda}}_K^{(b)})^\top u_K(T_i, \boldsymbol{X}_i)\right) u_K(T_i, \boldsymbol{X}_i) = \frac{1}{N(N-1)} \sum_{i=1, i\neq j}^{N} \sum_{j=1}^{N} u_K(T_i, \boldsymbol{X}_j).$$

Applying the mean value theorem, we have

$$\begin{aligned}
&\left[-\frac{1}{N}\sum_{i=1}^{N} \xi_i^{(b)} \rho''\left((\widetilde{\boldsymbol{\lambda}}^{(b)})^\top u_K(T_i, \boldsymbol{X}_i)\right) u_K(T_i, \boldsymbol{X}_i) u_K^\top(T_i, \boldsymbol{X}_i)\right] (\widehat{\boldsymbol{\lambda}}_K^{(b)} - \boldsymbol{\lambda}_K^*) \\
&= \frac{1}{N}\sum_{i=1}^{N} \xi_i^{(b)} \rho'\left(\{\boldsymbol{\lambda}_K^*\}^\top u_K(T_i, \boldsymbol{X}_i)\right) u_K(T_i, \boldsymbol{X}_i) - \frac{1}{N(N-1)} \sum_{i=1, i\neq j}^{N} \sum_{j=1}^{N} u_K(T_i, \boldsymbol{X}_j),
\end{aligned}$$

where $\widetilde{\boldsymbol{\lambda}}^{(b)}$ lies on the line joining from $\widehat{\boldsymbol{\lambda}}_K^{(b)}$ to $\boldsymbol{\lambda}_K^*$.

For presentatoin simplicity, we define the following notations any $(\boldsymbol{\theta}, t, z) \in \mathbb{R}^2 \times \mathcal{T} \times \mathcal{Z}$:

$$\overline{R}(T, \boldsymbol{X}, Y; \boldsymbol{\theta}, t) := \mathbb{E}\left[R(T, \boldsymbol{X}, Y; \boldsymbol{\theta}, t) | T, \boldsymbol{X}\right],$$

$$\widetilde{\Sigma}_K^{(b)} := -\frac{1}{N}\sum_{i=1}^{N}\xi_i^{(b)}\rho''\left\{(\widetilde{\boldsymbol{\lambda}}^{(b)})^\top u_K(T_i,\boldsymbol{X}_i)\right\}u_K(T_i,\boldsymbol{X}_i)u_K^\top(T_i,\boldsymbol{X}_i),$$

$$\Sigma_K := -\mathbb{E}\left[\rho''\left((\boldsymbol{\lambda}_K^*)^\top u_K(T,\boldsymbol{X})\right)u_K(T,\boldsymbol{X})u_K^\top(T,\boldsymbol{X})\right],$$

$$\widetilde{\Psi}_{K,\boldsymbol{h}}^{(b)}(\boldsymbol{\theta},t,z) := -\int \overline{R}(t',\boldsymbol{x}';\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(t'-t,z'''\left\{(\widetilde{\boldsymbol{\lambda}}^{(b)})^\top u_K(t',\boldsymbol{x}')\right\}u_K(t',\boldsymbol{x}')dF_{X,T}(\boldsymbol{x}',t'),$$

$$\Psi_{K,\boldsymbol{h}}(\boldsymbol{\theta},t,z) := -\int \overline{R}(t',\boldsymbol{x}';\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(t'-t,z'''\left\{(\boldsymbol{\lambda}_K^*)^\top u_K(t',\boldsymbol{x}')\right\}u_K(t',\boldsymbol{x})dF_{X,T}(\boldsymbol{x}',t').$$

Then

$$\widehat{\boldsymbol{\lambda}}_K^{(b)} - \boldsymbol{\lambda}_K^* \tag{C.1}$$

$$=\widetilde{\Sigma}_K^{(b)-1}\left[\frac{1}{N}\sum_{i=1}^{N}\xi_i^{(b)}\rho'\left(\{\boldsymbol{\lambda}_K^*\}^\top u_K(T_i,\boldsymbol{X}_i)\right)u_K(T_i,\boldsymbol{X}_i) - \frac{1}{N(N-1)}\sum_{i=1,i\neq j}^{N}\sum_{j=1}^{N}u_K(T_i,\boldsymbol{X}_j)\right].$$

We make the following decomposition:

$$\frac{1}{\sqrt{Nh_{pro}}}\sum_{i=1}^{N}\Big\{\xi_i^{(b)}\widehat{\pi}_K^{(b)}(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)$$

$$-\mathbb{E}\left[\pi_0(T_i,\boldsymbol{X}_i)R(T_i,\boldsymbol{X}_i,Y_i;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)\right]\Big\}$$

$$=\frac{1}{\sqrt{Nh_{pro}}}\sum_{i=1}^{N}\Big[\xi_i^{(b)}\left\{\widehat{\pi}_K^{(b)}(T_i,\boldsymbol{X}_i)-\pi_K^*(T_i,\boldsymbol{X}_i)\right\}R(T_i,\boldsymbol{X}_i,Y_i;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)$$

$$-\int_{\mathcal{T}}\int_{\mathcal{X}}\left\{\widehat{\pi}_K^{(b)}(t',\boldsymbol{x}')-\pi_K^*(t',\boldsymbol{x}')\right\}\overline{R}(t',\boldsymbol{x}';\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(t'-t,z'-z)dF_{X,T}(\boldsymbol{x}',t')\Big] \tag{C.2}$$

$$+\frac{1}{\sqrt{Nh_{pro}}}\sum_{i=1}^{N}\Big[\xi_i^{(b)}\left\{\pi_K^*(T_i,\boldsymbol{X}_i)-\pi_0(T_i,\boldsymbol{X}_i)\right\}R(T_i,\boldsymbol{X}_i,Y_i;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)$$

$$-\int_{\mathcal{T}}\int_{\mathcal{X}}\overline{R}(t',\boldsymbol{x}';\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(t'-t,z'-z)\left\{\pi_K^*(t',\boldsymbol{x}')-\pi_0(t',\boldsymbol{x}')\right\}dF_{X,T}(\boldsymbol{x}',t')\Big] \tag{C.3}$$

$$+\sqrt{\frac{N}{h_{pro}}}\int_{\mathcal{T}}\int_{\mathcal{X}}\overline{R}(t',\boldsymbol{x}';\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(t'-t,z'-z)\left\{\pi_K^*(t',\boldsymbol{x}')-\pi_0(t',\boldsymbol{x}')\right\}dF_{X,T}(\boldsymbol{x}',t') \tag{C.4}$$

$$+\sqrt{\frac{N}{h_{pro}}}\Big[\int_{\mathcal{T}}\int_{\mathcal{X}}\left\{\widehat{\pi}_K^{(b)}(t',\boldsymbol{x}')-\pi_K^*(t',\boldsymbol{x}')\right\}\overline{R}(t',\boldsymbol{x}';\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(t'-t,z'-z)dF_{X,T}(\boldsymbol{x}',t')$$

$$+\widetilde{\Psi}_{K,\boldsymbol{h}}^{(b)\top}(\boldsymbol{\theta},t,z)(\widehat{\boldsymbol{\lambda}}_K^{(b)}-\boldsymbol{\lambda}_K^*)\Big] \tag{C.5}$$

$$-\sqrt{\frac{N}{h_{pro}}}\widetilde{\Psi}_{K,\boldsymbol{h}}^{(b)\top}(\boldsymbol{\theta},t,z)(\widehat{\boldsymbol{\lambda}}_K^{(b)}-\boldsymbol{\lambda}_K^*) \tag{C.6}$$

$$+\frac{1}{\sqrt{Nh_{pro}}}\sum_{i=1}^{N}\Big\{\xi_i^{(b)}\pi_0(T_i,\boldsymbol{X}_i)\overline{R}(T_i,\boldsymbol{X}_i;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)$$

$$- \mathbb{E}\left[\pi_0(T_i, \boldsymbol{X}_i) R(T_i, \boldsymbol{X}_i, Y_i; \boldsymbol{\theta}, t) \mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z)\right] \Big\} \tag{C.7}$$

$$+ \frac{1}{\sqrt{N h_{pro}}} \sum_{i=1}^{N} \xi_i^{(b)} \pi_0(T_i, \boldsymbol{X}_i) \left\{ R(T_i, \boldsymbol{X}_i, Y_i; \boldsymbol{\theta}, t) - \overline{R}(T_i, \boldsymbol{X}_i; \boldsymbol{\theta}, t) \right\} \mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z).$$

We shall show that the terms (C.2)-(C.5) and (C.6) + (C.7) are all of $o_P\{(\log N)^{-1/2}\}$ uniformly over $(\boldsymbol{\theta}, t, z) \in [-M, M]^2 \times \mathcal{T} \times \mathcal{Z}$. Then, by rearranging the equation, we have the results.

***For term* (C.2):** Recalling that $\widehat{\pi}_K^{(b)}(t, \boldsymbol{x}) = \rho'\{(\widehat{\lambda}_K^{(b)})^\top u_K(t, \boldsymbol{x})\}$, by applying the mean value theorem, we obtain obtain

$$\begin{aligned}
\text{(C.2)} =& \frac{1}{\sqrt{N h_{pro}}} \sum_{i=1}^{N} \Bigg[\xi_i^{(b)} R(T_i, \boldsymbol{X}_i, Y_i; \boldsymbol{\theta}, t) \mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z) \rho'' \left\{ (\widetilde{\boldsymbol{\lambda}}^{(b)})^\top u_K(T_i, \boldsymbol{X}_i) \right\} \\
& \qquad \times \{\widehat{\boldsymbol{\lambda}}_K^{(b)} - \boldsymbol{\lambda}_K^*\}^\top u_K(T_i, \boldsymbol{X}_i) \\
& - \int_{\mathcal{T} \times \mathcal{X}} \overline{R}(t, \boldsymbol{x}; \boldsymbol{\theta}, t_0) \mathcal{K}_{\boldsymbol{h}}(t' - t, z''' \left\{ (\widetilde{\boldsymbol{\lambda}}^{(b)})^\top u_K(T_i, \boldsymbol{X}_i) \right\} \{\widehat{\boldsymbol{\lambda}}_K^{(b)} - \boldsymbol{\lambda}_K^{*(b)}\}^\top u_K(t', \boldsymbol{x}') dF_{X,T}(\boldsymbol{x}', t') \Bigg] \\
=& W_1(\boldsymbol{\theta}, t, z) + W_2(\boldsymbol{\theta}, t, z) + W_3(\boldsymbol{\theta}, t, z),
\end{aligned}$$

where, by applying the mean value theorem again,

$$\begin{aligned}
& W_1(\boldsymbol{\theta}, t, z) \\
& := \frac{1}{\sqrt{N h_{pro}}} \sum_{i=1}^{N} \Bigg[\xi_i^{(b)} R(T_i, \boldsymbol{X}_i, Y_i; \boldsymbol{\theta}, t) \mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z) \rho'' \left(\{\boldsymbol{\lambda}_K^*\}^\top u_K(T_i, \boldsymbol{X}_i) \right) \\
& \qquad \times \{\widehat{\boldsymbol{\lambda}}_K^{(b)} - \boldsymbol{\lambda}_K^*\}^\top u_K(T_i, \boldsymbol{X}_i) \\
& - \int_{\mathcal{T} \times \mathcal{X}} \overline{R}(t, \boldsymbol{x}; \boldsymbol{\theta}, t) \mathcal{K}_{\boldsymbol{h}}(t' - t, z''' \left(\{\boldsymbol{\lambda}_K^{*(b)}\}^\top u_K(t', \boldsymbol{x}') \right) \{\widehat{\boldsymbol{\lambda}}_K^{(b)} - \boldsymbol{\lambda}_K^*\}^\top u_K(t', \boldsymbol{x}') dF_{X,T}(\boldsymbol{x}', t') \Bigg], \\
& W_2(\boldsymbol{\theta}, t, z) := \frac{1}{\sqrt{N h_{pro}}} \sum_{i=1}^{N} \Bigg[\xi_i^{(b)} R(T_i, \boldsymbol{X}_i, Y_i; \boldsymbol{\theta}, t) \mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z) \rho''' \{\xi_3(T_i, \boldsymbol{X}_i)\} \\
& \qquad \times \{\widetilde{\boldsymbol{\lambda}}^{(b)} - \boldsymbol{\lambda}_K^*\}^\top u_K(T_i, \boldsymbol{X}_i) \{\widehat{\boldsymbol{\lambda}}_K^{(b)} - \boldsymbol{\lambda}_K^*\}^\top u_K(T_i, \boldsymbol{X}_i) \Bigg], \\
& W_3(\boldsymbol{\theta}, t, z) := -\sqrt{\frac{N}{h_{pro}}} \int_{\mathcal{T} \times \mathcal{X}} \overline{R}(t', \boldsymbol{x}'; \boldsymbol{\theta}, t) \mathcal{K}_{\boldsymbol{h}}(t' - t, z'''' \{\xi_3(t', \boldsymbol{x}')\} \\
& \qquad \times \{\widetilde{\boldsymbol{\lambda}}^{(b)} - \boldsymbol{\lambda}_K^*\}^\top u_K(t', \boldsymbol{x}') \{\widehat{\boldsymbol{\lambda}}_K^{(b)} - \boldsymbol{\lambda}_K^*\}^\top u_K(t', \boldsymbol{x}') dF_{X,T}(\boldsymbol{x}', t'),
\end{aligned}$$

and $\xi_3(t, \boldsymbol{x})$ lies between $(\widetilde{\boldsymbol{\lambda}}^{(b)})^\top u_K(t, \boldsymbol{x})$ and $\{\boldsymbol{\lambda}_K^*\}^\top u_K(t, \boldsymbol{x})$.

For the term $W_1(\boldsymbol{\theta}, t, z)$, we first note that

$$\begin{aligned}
& \sup_{(\boldsymbol{\theta}, t, z)} |W_1(\boldsymbol{\theta}, t, z)| \\
= & \sup_{(\boldsymbol{\theta}, t, z)} \Bigg| \{\widehat{\boldsymbol{\lambda}}_K^{(b)} - \boldsymbol{\lambda}_K^*\}^\top
\end{aligned}$$

$$
\times \frac{1}{\sqrt{Nh_{pro}}}\sum_{i=1}^{N}\Big[\xi_i^{(b)}R(T_i,\boldsymbol{X}_i,Y_i;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)\rho''\left(\{\boldsymbol{\lambda}_K^*\}^\top u_K(T_i,\boldsymbol{X}_i)\right)u_K(T_i,\boldsymbol{X}_i)
$$

$$
-\int_{\mathcal{T}}\int_{\mathcal{X}}\overline{R}(t',\boldsymbol{x}';\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(t'-t,z'''\left(\{\boldsymbol{\lambda}_K^*\}^\top u_K(t',\boldsymbol{x}')\right)u_K(t',\boldsymbol{x}')dF_{X,T}(\boldsymbol{x}',t')\Big]\Big\|
$$

$$
\leq \|\widehat{\boldsymbol{\lambda}}_K^{(b)}-\boldsymbol{\lambda}_K^*\|
$$

$$
\times\sup_{\boldsymbol{\theta},t,z}\Big\|\frac{1}{\sqrt{Nh_{pro}}}\sum_{i=1}^{N}\Big[\xi_i^{(b)}R(T_i,\boldsymbol{X}_i,Y_i;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)\rho''\left(\{\boldsymbol{\lambda}_K^*\}^\top u_K(T_i,\boldsymbol{X}_i)\right)u_K(T_i,\boldsymbol{X}_i)
$$

$$
-\int_{\mathcal{T}}\int_{\mathcal{X}}\overline{R}(t',\boldsymbol{x}';\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(t'-t,z'''\left(\{\boldsymbol{\lambda}_K^*\}^\top u_K(t',\boldsymbol{x}')\right)u_K(t',\boldsymbol{x}')dF_{X,T}(\boldsymbol{x}',t')\Big]\Big\|.
$$

For each $b=0,1,\ldots,B$ and $\ell=1,\ldots,K$ let

$$
\mathcal{F}_{b,\ell}=\{(\xi^{(b)},t',\boldsymbol{x}'^{(b)}R(t',\boldsymbol{x}',y;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(t'-t,z'-z)
\times\rho''\left[\{\boldsymbol{\lambda}_K^*\}^\top u_K(t',\boldsymbol{x}')\right]u_{K,\ell}(t',\boldsymbol{x}'):(\boldsymbol{\theta},t,z)\in\mathbb{R}^2\times\mathcal{T}\times\mathcal{Z}\},
$$

where $u_{K,\ell}$ is the ℓ-th component of u_K. We apply Corollary 5.1 of Chernozhukov et al. (2014b) to each $\mathcal{F}_{b,\ell}$ to obtain the uniform convergence rate of $W_1(\boldsymbol{\theta},t,z)$. Using Lemma 2 and Lemma 2.6.18 (vi) of Van Der Vaart and Wellner (1996), we have $\mathcal{F}_{b,\ell}$ is a VC-type class, for $b=0,1,\ldots,B$ and $\ell=1,\ldots,K$. Next, we calculate that

$$
\sup_{(\boldsymbol{\theta},t,z)}\mathbb{E}\Big[\left|\xi^{(b)}R(T,\boldsymbol{X},Y;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T-t,Z-z)\rho''\left\{(\boldsymbol{\lambda}_K^*)^\top u_K(T,\boldsymbol{X})\right\}u_{K,\ell}(T,\boldsymbol{X})\right|^2\Big]
$$

$$
\leq 2\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}\left|\rho''\left\{(\boldsymbol{\lambda}_K^*)^\top u_K(t,\boldsymbol{x})\right\}u_{K,\ell}(t,\boldsymbol{x})\right|^2\cdot C\cdot h_{pro}
$$

$$
=\text{const}\times h_{pro}\,.
$$

We then let $\sigma^2=\text{const}\cdot h_{pro}$. By Assumption 10 (i), we know that $\mathcal{F}_{b,\ell}$ admits an envelope function

$$
F_\ell(\xi^{(b)},t,\boldsymbol{x},y):=2C\cdot\sup_{\alpha,\beta\in[-M,M]}|\mathcal{L}'(y-\alpha-\beta\cdot t)|\cdot\left|\rho''\left(\{\boldsymbol{\lambda}_K^*\}^\top u_K(t,\boldsymbol{x})\right)\right||u_{K,\ell}(t,\boldsymbol{x})|
\times\sup_{t'\in\mathcal{T}}|\phi_N(t,\boldsymbol{x},y;t')|
$$

for some sufficiently large $C>0$. Hence, the conditions imposed in Chernozhukov et al. (2014b, Corollary 5.1) are satisfied. Next, we calculate the quantities in Chernozhukov et al. (2014b, Corollary 5.1). In their notations, A and v are both constants and we have

$$
\|F_\ell\|_{P,2}\leq 2C\times\left\|\sup_{\alpha,\beta\in[-M,M]}|\mathcal{L}'(y-\alpha-\beta\cdot t)|\sup_{t'\in\mathcal{T}}|\phi_N(t,\boldsymbol{x},y;t')|\right\|_{P,2}
\times\sup_{(t,x)\in\mathcal{T}\times\mathcal{X}}\left|\rho''\left(\{\boldsymbol{\lambda}_K^*\}^\top u_K(t,\boldsymbol{x})\right)u_{K,\ell}(t,\boldsymbol{x})\right|
$$

$$
=C'\cdot\left\|\sup_{t'\in\mathcal{T}}|\phi_N(t,\boldsymbol{x},y;t')|\right\|_{P,2}=O(N^{c_0})\,,
$$

for some positive contstant C'. Furthermore, for some $q > 2$, under Assumption 10 (ii), we have

$$\|M_\ell\|_2^2 := \mathbb{E}\left\{\max_{1\le i\le N} F_\ell^2(T_i, \boldsymbol{X}_i, Y_i)\right\} \le \left[\mathbb{E}\left\{\max_{1\le i\le N} F_\ell^q(T_i, \boldsymbol{X}_i, Y_i)\right\}\right]^{2/q}$$
$$\le N^{2/q}\left[\mathbb{E}\left\{F_\ell^q(T_i, \boldsymbol{X}_i, Y_i)\right\}\right]^{2/q} \le \text{const} \times N^{2/q+2c_0}\,.$$

Then, by Chernozhukov et al. (2014b, Corollary 5.1), we have the first moment

$$\mathbb{E}\left(\sup_{\boldsymbol{\theta},t,z}\left|\frac{1}{\sqrt{N}}\sum_{i=1}^N\left[\xi_i^{(b)} R(T_i, \boldsymbol{X}_i, Y_i; \boldsymbol{\theta}, t)\mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z)\rho''\left(\{\boldsymbol{\lambda}_K^{*(b)}\}^\top u_K(T_i, \boldsymbol{X}_i)\right)u_{K,\ell}(T_i, \boldsymbol{X}_i)\right.\right.\right.$$
$$\left.\left.\left. - \int_{\mathcal{T}}\int_{\mathcal{X}} \overline{R}(t', \boldsymbol{x}'; \boldsymbol{\theta}, t)\mathcal{K}_{\boldsymbol{h}}(t' - t, z''' \left(\{\boldsymbol{\lambda}_K^{*(b)}\}^\top u_K(t', \boldsymbol{x}')\right)u_{K,\ell}(t', \boldsymbol{x}')dF_{X,T}(\boldsymbol{x}', t')\right]\right|\right)$$
$$= O_P\left\{\sqrt{\sigma^2 \log\left(\frac{\|F_{b,\ell}\|_{P,2}}{\sigma}\right)} + \frac{\|M_{b,\ell}\|_2}{\sqrt{N}}\log\left(\frac{\|F_{b,\ell}\|_{P,2}}{\sigma}\right)\right\}$$
$$= O_P\left(\sqrt{N^{-c_1}\log(N^{c_1+c_0})} + \frac{N^{1/q+c_0}\log(N^{c_1+c_0})}{\sqrt{N}}\right).$$

Note that the second moment is bounded above by $\sigma^2 = O(h_{pro})$. Combining the results above and given that $h_{pro} \sim N^{-c_1}$ and $K \sim N^{c_2}$ for some constant $c_1, c_2 > 0$, we have

$$\sup_{(\boldsymbol{\theta},t,z)} |W_1(\boldsymbol{\theta}, t, z)|$$
$$\le \|\widehat{\boldsymbol{\lambda}}_K - \boldsymbol{\lambda}_K^*\| O_P\left(\sqrt{\log(N^{c_1+c_0})K} + \frac{N^{1/q+c_0}\log(N^{c_1+c_0})\sqrt{K}}{\sqrt{Nh_{pro}}}\right)$$
$$= O_p\left(\sqrt{K/N}\right) O_P\left(\sqrt{\log(N^{c_1+c_0})K} + \frac{N^{1/q+c_0}\log(N^{c_1+c_0})\sqrt{K}}{\sqrt{Nh_{pro}}}\right)$$
$$= o_p\{(\log N)^{-1/2}\}\,,$$

where the last equality holds provided that Assumption 11 holds.

For the term $W_3(\boldsymbol{\theta}, t, z)$, by Lemma 1, we get

$$\sup_{(\boldsymbol{\theta},t,z)} |W_3(\boldsymbol{\theta}, t, z)|$$
$$= \sup_{(\boldsymbol{\theta},t,z)}\left|\sqrt{\frac{N}{h_{pro}}}\int_{\mathcal{T}}\int_{\mathcal{X}} \overline{R}(t', \boldsymbol{x}'; \boldsymbol{\theta}, t)\mathcal{K}_{\boldsymbol{h}}(t' - t, z'''' \{\xi_3(t', \boldsymbol{x}')\}\{\widetilde{\boldsymbol{\lambda}}^{(b)} - \boldsymbol{\lambda}_K^*\}^\top u_K(t', \boldsymbol{x}')\right.$$
$$\left.\times\{\widehat{\boldsymbol{\lambda}}_K^{(b)} - \boldsymbol{\lambda}_K^*\}^\top u_K(t', \boldsymbol{x}')dF_{X,T}(\boldsymbol{x}', t')\right|$$
$$\le \sqrt{N}\cdot \sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\,\mathcal{X}} |\rho'''(\xi_3(t, \boldsymbol{x}))| \sup_{(\boldsymbol{\theta},t,z)}\left\{\frac{1}{h_{pro}}\int_{\mathcal{T}}\int_{\mathcal{X}} |\overline{R}(t', \boldsymbol{x}'; \boldsymbol{\theta}, t)|^2\mathcal{K}_{\boldsymbol{h}}^2(t' - t, z' - z)dF_{X,T}(\boldsymbol{x}', t')\right\}^{1/2}$$
$$\cdot\left\{\int_{\mathcal{T}}\int_{\mathcal{X}}\left|\{\widetilde{\boldsymbol{\lambda}}^{(b)} - \boldsymbol{\lambda}_K^*\}^\top u_K(t, \boldsymbol{x})\right|^2 dF_{X,T}(\boldsymbol{x}, t)\right\}^{1/2}\cdot\left\|\widehat{\boldsymbol{\lambda}}_K^{(b)} - \boldsymbol{\lambda}_K^*\right\|\cdot \sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}\|u_K(t, \boldsymbol{x})\|$$

$$=\sqrt{N}\cdot O_P(1)\cdot O(1)\cdot O_p\left(\sqrt{K/N}\right)\cdot O_P\left(\sqrt{K/N}\right)\cdot\zeta(K)=O_p\left(\zeta(K)K/\sqrt{N}\right). \tag{C.8}$$

For the term $W_2(\boldsymbol{\theta},t,z)$, we have

$$\begin{aligned}
&\sup_{(\boldsymbol{\theta},t,z)}|W_2(\boldsymbol{\theta},t,z)|\\
\leq&2\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}|\rho'''(\xi_3(t,\boldsymbol{x}))|\cdot\sup_{(\boldsymbol{\theta},t,z)}\left\{\frac{1}{Nh_{pro}}\sum_{i=1}^{N}R(T_i,\boldsymbol{X}_i,Y_i;\boldsymbol{\theta},t)^2\mathcal{K}_{\boldsymbol{h}}^2(t'-t,z'-z)\right\}^{1/2}\\
&\times\sqrt{N}\cdot\left\{\frac{1}{N}\sum_{i=1}^{N}\left|\{\widetilde{\boldsymbol{\lambda}}^{(b)}-\boldsymbol{\lambda}_K^*\}^\top u_K(T_i,\boldsymbol{X}_i)\right|^2\right\}^{1/2}\cdot\left\|\widehat{\boldsymbol{\lambda}}_K^{(b)}-\boldsymbol{\lambda}_K^*\right\|\cdot\sup_{(t',\boldsymbol{x}')\in\mathcal{T}\times\mathcal{X}}\|u_K(t',\boldsymbol{x}')\|\\
\leq&O_p(1)\cdot O_p(1)\cdot\sqrt{N}\cdot O_p\left(\sqrt{K/N}\right)\cdot O_p\left(\zeta(K)\sqrt{K/N}\right)=O_p\left(\zeta(K)K/\sqrt{N}\right).
\end{aligned}$$

Combining all results, we get

$$\begin{aligned}
|(\text{C.2})|\leq&\sup_{(\boldsymbol{\theta},t,z)}|W_1(\boldsymbol{\theta},t,z)|+\sup_{(\boldsymbol{\theta},t,z)}|W_2(\boldsymbol{\theta},t,z)|+\sup_{(\boldsymbol{\theta},t,z)}|W_3(\boldsymbol{\theta},t,z)|\\
=&o_p\{(\log N)^{-1/2}\}+O_p\left(\zeta(K)K/\sqrt{N}\right)=o_P\{(\log N)^{-1/2}\},
\end{aligned}$$

where the rate holds uniform over $(\boldsymbol{\theta},t,z)\in[-M,M]^2\times\mathcal{T}\times\mathcal{Z}$.

***For term* (C.3):** We apply again Chernozhukov et al. (2014b, Corollary 5.1) to obtain the uniform convergence rate. For $b=0,1,\ldots,B$, let

$$\begin{aligned}
\mathcal{F}_{b,N}=&\{(\xi^{(b)},t',\boldsymbol{x}'^{(b)}R(t',\boldsymbol{x}',y;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(t'-t,z'-z)\{\pi_K^*(t',\boldsymbol{x}')-\pi_0(t',\boldsymbol{x}')\}\\
&:(\boldsymbol{\theta},t,z)\in[-M,M]^2\times\mathcal{T}\times\mathcal{Z}\}.
\end{aligned}$$

Using Lemma 2 and Lemma 2.6.18 (vi) of Van Der Vaart and Wellner (1996), we have $\mathcal{F}_{b,N}$ is a VC-type class, for $b=0,1,\ldots,B$. Moreover, we have

$$\begin{aligned}
\sigma^2:=&\sup_{(\boldsymbol{\theta},t,z)}\mathbb{E}\left[|\xi^{(b)}R(T,\boldsymbol{X},Y;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T-t,Z-z)\{\pi_K^*(T,\boldsymbol{X})-\pi_0(T,\boldsymbol{X})\}|^2\right]\\
\leq&2\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}|\pi_K^*(t,\boldsymbol{x})-\pi_0(t,\boldsymbol{x})|^2\cdot\sup_{(\boldsymbol{\theta},t,z)}\mathbb{E}\left[|R(T,\boldsymbol{X},Y;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T-t,Z-z)|^2\right]\\
\leq&2\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}|\pi_K^*(t,\boldsymbol{x})-\pi_0(t,\boldsymbol{x})|^2\times C\cdot h_{pro}\\
=&\text{const}\times h_{pro}\cdot\zeta(K)^2\cdot K^{-2\omega}.
\end{aligned}$$

By Assumption 10 (i), we know that $\mathcal{F}$ admits an envelope function

$$F_N(t,\boldsymbol{x},y)=2C\cdot\sup_{a,b\in[-M,M]}|\mathcal{L}'(y-a-b\cdot t)|\cdot\sup_{t'\in\mathcal{T}}|\phi_N(t,\boldsymbol{x},y;t')||\pi_K^*(t,\boldsymbol{x})-\pi_0(t,\boldsymbol{x})|$$

for some sufficiently large $C>0$ and $\mathcal{F}$ is a VC-type class. Hence, the condition of the Corollary 5.1 in Chernozhukov et al. (2014b) is satisfied. Next, we calculate the quantities in Corollary 5.1 of Chernozhukov et al. (2014b). In their notations, A and v are both constants and we have

$$\|F\|_{P,2}\leq\text{const}\times\left\|\sup_{a,b\in[-M,M]}\left|\mathcal{L}'(y-a-b\cdot t)\sup_{t'}\phi_N(t,\boldsymbol{x},y;t')\right|\right\|_{P,2}$$

$$
\begin{aligned}
&\times \sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} |\pi_K^*(t,\boldsymbol{x}) - \pi_0(t,\boldsymbol{x})| \\
=&\text{const} \cdot N^{c_0} \times \zeta(K) K^{-\omega}.
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
\|M\|_2^2 :=& \mathbb{E}\left\{\max_{1\le i\le n} F^2(T_i,\boldsymbol{X}_i,Y_i)\right\} \le \left[\mathbb{E}\left\{\max_{1\le i\le n} F^q(T_i,\boldsymbol{X}_i,Y_i)\right\}\right]^{2/q} \\
\le& N^{2/q}\left[\mathbb{E}\left\{F^q(T_i,\boldsymbol{X}_i,Y_i)\right\}\right]^{2/q} \le \text{const} \times N^{2/q+2c_0} \times \zeta(K)^2 K^{-2\omega}.
\end{aligned}
$$

Then, simliar to the arguments for term $\sup_{(\boldsymbol{\theta},t,z)} |W_1(\boldsymbol{\theta},t,z)|$, we have

$$
\begin{aligned}
&\sup_{(\boldsymbol{\theta},t,z)} \left| \frac{1}{\sqrt{Nh_{pro}}} \sum_{i=1}^N \left\{ \xi_i^{(b)} \left(\pi_K^*(T_i,\boldsymbol{X}_i) - \pi_0(T_i,\boldsymbol{X}_i)\right) R(T_i,\boldsymbol{X}_i,Y_i;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z) \right.\right. \\
&\qquad \left.\left. - \mathbb{E}\left[\overline{R}(T,\boldsymbol{X};\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T-t,Z-z)\left\{\pi_K^*(T,\boldsymbol{X}) - \pi_0(T,\boldsymbol{X})\right\}\right] \right\} \right| \\
&= o_P\{(\log N)^{-1/2}\},
\end{aligned}
$$

where the second last equality holds provided that Assumption 11 holds.
***For term* (C.4):** By Lemma 1, we can deduce that

$$
\begin{aligned}
&\sup_{(\boldsymbol{\theta},t,z)} \left| \sqrt{\frac{N}{h_{pro}}} \cdot \mathbb{E}\left[\overline{R}(T,\boldsymbol{X};\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T-t,Z-z)\left\{\pi_K^*(T,\boldsymbol{X}) - \pi_0(T,\boldsymbol{X})\right\}\right] \right| \\
\le& \sqrt{\frac{N}{h_{pro}}} \cdot \sup_{(\boldsymbol{\theta},t,z)} \mathbb{E}[\{\overline{R}(T,\boldsymbol{X};\boldsymbol{\theta},t,z)\}^2 \mathcal{K}_{\boldsymbol{h}}^2(T-t,Z-z)]^{\frac{1}{2}} \\
&\times \mathbb{E}\left[|\pi_K^*(T,\boldsymbol{X}) - \pi_0(T,\boldsymbol{X})|^2\right]^{\frac{1}{2}} \\
=& O\left(\sqrt{N} K^{-\omega}\right) = o\{(\log N)^{-1/2}\}.
\end{aligned}
$$

***For term* (C.5):** By the mean value theorem, the term (C.5) is exactly equal to zero.
***For term* (C.6):** For the term (C.6), using (C.1), we have

$$
\begin{aligned}
-\text{(C.6)} =& \sqrt{\frac{N}{h_{pro}}} \Psi_{K,\boldsymbol{h}}^\top(\boldsymbol{\theta},t,z)\Sigma_K^{-1} \cdot \left[\frac{1}{N}\sum_{i=1}^N \xi_i^{(b)} \pi_K^*(T_i,\boldsymbol{X}_i) u_K(T_i,\boldsymbol{X}_i) \right. \\
&\qquad \left. - \frac{1}{N(N-1)} \sum_{i=1,i\neq j}^N \sum_{j=1}^N u_K(T_i,\boldsymbol{X}_j)\right] \qquad \text{(C.9)} \\
&+ \sqrt{\frac{N}{h_{pro}}} \left\{\widetilde{\Psi}_{K,\boldsymbol{h}}^{(b)}(\boldsymbol{\theta},t,z) - \Psi_{K,\boldsymbol{h}}(\boldsymbol{\theta},t,z)\right\}^\top \Sigma_K^{-1} \\
&\quad \cdot \left[\frac{1}{N}\sum_{i=1}^N \xi_i^{(b)} \pi_K^*(T_i,\boldsymbol{X}_i) u_K(T_i,\boldsymbol{X}_i) - \frac{1}{N(N-1)} \sum_{i=1,i\neq j}^N \sum_{j=1}^N u_K(T_i,\boldsymbol{X}_j)\right] \qquad \text{(C.10)}
\end{aligned}
$$

$$
+\sqrt{\frac{N}{h_{pro}}}\widetilde{\Psi}_{K,\boldsymbol{h}}^{(b)\top}(\boldsymbol{\theta},t,z)\widetilde{\Sigma}_K^{(b)-1}\left\{\Sigma_K-\widetilde{\Sigma}_K^{(b)}\right\}\Sigma_K^{-1}\cdot\left[\frac{1}{N}\sum_{i=1}^N\xi_i^{(b)}\pi_K^*(T_i,\boldsymbol{X}_i)u_K(T_i,\boldsymbol{X}_i)\right.
$$

$$
\left.-\frac{1}{N(N-1)}\sum_{i=1,i\neq j}^N\sum_{j=1}^N u_K(T_i,\boldsymbol{X}_j)\right]. \tag{C.11}
$$

Consider the term (C.10) first. Let

$$
E_N:=\frac{1}{N}\sum_{i=1}^N\xi_i^{(b)}\pi_K^*(T_i,\boldsymbol{X}_i)u_K(T_i,\boldsymbol{X}_i)-\frac{1}{N(N-1)}\sum_{i=1,i\neq j}^N\sum_{j=1}^N u_K(T_i,\boldsymbol{X}_j),
$$

and $\mathbb{E}[E_N]=0$. We have $\|E_N\|=O_P(\sqrt{K/N})$ from Chebyshev's inequality. Note that

$$
\begin{aligned}
|(\text{C.10})|^2 \leq & \sup_{(\boldsymbol{\theta},t,z)}\left|\sqrt{\frac{N}{h_{pro}}}\{\widetilde{\Psi}_{K,\boldsymbol{h}}^{(b)}(\boldsymbol{\theta},t,z)-\Psi_{K,\boldsymbol{h}}(\boldsymbol{\theta},t,z)\}^\top\Sigma_K^{-1}\cdot E_N\right|^2\\
\leq & \frac{N}{h_{pro}}\cdot\sup_{(\boldsymbol{\theta},t,z)}\|\widetilde{\Psi}_{K,\boldsymbol{h}}^{(b)}(\boldsymbol{\theta},t,z)-\Psi_{K,\boldsymbol{h}}(\boldsymbol{\theta},t,z)\|^2\cdot\left\|\Sigma_K^{-1}\cdot E_N\right\|^2\\
\leq & N\cdot\lambda_{\min}^{-2}(\Sigma_K)\cdot\sup_{(\boldsymbol{\theta},t,z)}\frac{1}{h_{pro}}\|\widetilde{\Psi}_{K,\boldsymbol{h}}(\boldsymbol{\theta},t,z)-\Psi_{K,\boldsymbol{h}}(\boldsymbol{\theta},t,z)\|^2\cdot\|E_N\|^2\\
= & N\cdot O(1)\cdot O_P\left(h_{pro}\cdot\zeta(K)^2\cdot\frac{K}{N}\right)\cdot O_P\left(\frac{K}{N}\right)\\
= & O_P\left(h_{pro}\cdot\zeta(K)^2\frac{K^2}{N}\right)=o_P\{(\log N)^{-1}\},
\end{aligned}
$$

where the third equality holds because

$$
\begin{aligned}
&\sup_{(\boldsymbol{\theta},t,z)}\|\widetilde{\Psi}_{K,\boldsymbol{h}}^{(b)}(\boldsymbol{\theta},t,z)-\Psi_{K,\boldsymbol{h}}(\boldsymbol{\theta},t,z)\|^2\\
=&\sup_{(\boldsymbol{\theta},t,z)}\left\|\int_{\mathcal{T}}\int_{\mathcal{X}}\overline{R}(t',\boldsymbol{x}';\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(t'-t,z'-z)\right.\\
&\qquad\left.\times\left(\rho''\left\{\widetilde{\boldsymbol{\lambda}}^{(b)\top}u_K(t',\boldsymbol{x}')\right\}-\rho''\left\{(\boldsymbol{\lambda}_K^*)^\top u_K(T,\boldsymbol{X})\right\}\right)u_K(t',\boldsymbol{x}')dF_{X,T}(\boldsymbol{x}',t')\right\|^2\\
=&O(1)\cdot\sup_{(\boldsymbol{\theta},t,z)}\int_{\mathcal{T}}\int_{\mathcal{X}}|\overline{R}(t',\boldsymbol{x}';\boldsymbol{\theta},t)|^2\cdot|\mathcal{K}_{\boldsymbol{h}}(t'-t,z'^2\\
&\qquad\times\left|\rho''\left\{\widetilde{\boldsymbol{\lambda}}^{(b)\top}u_K(t',\boldsymbol{x}')\right\}-\rho''\left\{(\boldsymbol{\lambda}_K^*)^\top u_K(t',\boldsymbol{x}')\right\}\right|^2dF_{X,T}(\boldsymbol{x}',t')\quad\text{(Assumption 7)}\\
=&O(1)\cdot\sup_{(\boldsymbol{\theta},t,z)}\int_{\mathcal{T}}\int_{\mathcal{X}}|\overline{R}(t',\boldsymbol{x}';\boldsymbol{\theta},t)|^2\cdot|\mathcal{K}_{\boldsymbol{h}}(t'-t,z'^2dF_{X,T}(\boldsymbol{x}',t')\\
&\times\sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}}\left|\rho''\left\{\widetilde{\boldsymbol{\lambda}}^{(b)\top}u_K(t,\boldsymbol{x})\right\}-\rho''\left\{(\boldsymbol{\lambda}_K^*)^\top u_K(t,\boldsymbol{x})\right\}\right|^2\\
=&O(1)\cdot\sup_{(\boldsymbol{\theta},t,z)}\int_{\mathcal{T}}\int_{\mathcal{X}}|\overline{R}(t',\boldsymbol{x}';\boldsymbol{\theta},t)|^2\cdot|\mathcal{K}_{\boldsymbol{h}}(t'-t,z'^2dF_{X,T}(\boldsymbol{x}',t')
\end{aligned}
$$

$$
\begin{aligned}
&\times \sup_{(t,\boldsymbol{x})\in\mathcal{T}\times\mathcal{X}} \left(|\rho'''\{\xi(t,\boldsymbol{x})\}|^2 \left| (\widetilde{\boldsymbol{\lambda}}^{(b)} - \boldsymbol{\lambda}_K^*)^\top u_K(t,\boldsymbol{x}) \right|^2 \right) \quad \text{(Mean value theorem)} \\
=& O_P\left(h_{pro} \cdot \zeta(K)^2 \cdot \frac{K}{N} \right).
\end{aligned}
$$

Consider the term (C.11), we have

$$
\begin{aligned}
|(\text{C.11})|^2 \leq & \sup_{(\boldsymbol{\theta},t,z)} \left| \sqrt{\frac{N}{h_{pro}}} \widetilde{\Psi}_{K,\boldsymbol{h}}^{(b)\top}(\boldsymbol{\theta},t,z) \widetilde{\Sigma}_K^{(b)-1} \left\{ \Sigma_K - \widetilde{\Sigma}_K^{(b)} \right\} \Sigma_K^{-1} E_N \right|^2 \\
\leq & \frac{N}{h_{pro}} \cdot \sup_{(\boldsymbol{\theta},t,z)} \|\widetilde{\Psi}_{K,\boldsymbol{h}}^{(b)}(\boldsymbol{\theta},t,z)\|^2 \cdot \left\| \widetilde{\Sigma}_K^{(b)-1} \left\{ \Sigma_K - \widetilde{\Sigma}_K^{(b)} \right\} \Sigma_K^{-1} \right\|^2 \cdot \|E_N\|^2 \\
=& \frac{N}{h_{pro}} \cdot O_P(h_{pro}) \cdot O_P\left(\zeta^4(K) \cdot \frac{K}{N} \right) \cdot O_P\left(\frac{K}{N} \right) \\
=& O_P\left(\zeta^4(K) \cdot \frac{K^2}{N} \right) = o_P\{(\log N)^{-1}\},
\end{aligned}
$$

where the first equality holds because

$$
\begin{aligned}
& \left\| \widetilde{\Sigma}_K^{(b)-1} \left\{ \Sigma_K - \widetilde{\Sigma}_K^{(b)} \right\} \Sigma_K^{-1} \right\|^2 \\
=& \operatorname{tr}\left(\widetilde{\Sigma}_K^{(b)-2} \left\{ \Sigma_K - \widetilde{\Sigma}_K^{(b)} \right\} \Sigma_K^{-2} \left\{ \Sigma_K - \widetilde{\Sigma}_K^{(b)} \right\} \right) \\
\leq & \lambda_{\min}^{-2}\left(\widetilde{\Sigma}_K^{(b)} \right) \cdot \operatorname{tr}\left(\left\{ \Sigma_K - \widetilde{\Sigma}_K^{(b)} \right\} \Sigma_K^{-2} \left\{ \Sigma_K - \widetilde{\Sigma}_K^{(b)} \right\} \right) \\
\leq & \lambda_{\min}^{-2}\left(\widetilde{\Sigma}_K^{(b)} \right) \lambda_{\min}^{-2}\left(\Sigma_K \right) \cdot \operatorname{tr}\left(\left\{ \Sigma_K - \widetilde{\Sigma}_K^{(b)} \right\} \left\{ \Sigma_K - \widetilde{\Sigma}_K^{(b)} \right\} \right) \\
=& O_P(1) \cdot \|\Sigma_K - \widetilde{\Sigma}_K^{(b)}\|^2,
\end{aligned}
$$

and

$$
\begin{aligned}
& \|\Sigma_K - \widetilde{\Sigma}_K^{(b)}\|^2 \\
\leq & \left\| \frac{1}{N} \sum_{i=1}^N \xi_i^{(b)} \rho''\left((\boldsymbol{\lambda}_K^*)^\top u_K(T_i,\boldsymbol{X}_i) \right) u_K(T_i,\boldsymbol{X}_i) u_K^\top(T_i,\boldsymbol{X}_i) \right. \\
& \quad \left. - \mathbb{E}\left[\rho''\left((\boldsymbol{\lambda}_K^*)^\top u_K(T,\boldsymbol{X}) \right) u_K(T,\boldsymbol{X}) u_K^\top(T,\boldsymbol{X}) \right] \right\|^2 \quad \text{(mean value theorem)} \\
& + \left\| \frac{1}{N} \sum_{i=1}^N \xi_i^{(b)} \rho'''\left(\widetilde{\boldsymbol{\lambda}}^{(b)\top} u_K(T_i,\boldsymbol{X}_i) \right) u_K(T_i,\boldsymbol{X}_i) u_K^\top(T_i,\boldsymbol{X}_i) \{\widetilde{\boldsymbol{\lambda}}^{(b)} - \boldsymbol{\lambda}_K^*\}^\top u_K(T_i,\boldsymbol{X}_i) \right\|^2 \\
=& O_P\left\{ \zeta^2(K) \cdot \frac{K}{N} \right\} + O_P\left(\frac{K}{N} \right) \cdot O\{\zeta(K)^4\} = O_P\left(\frac{K}{N} \right) \cdot O\{\zeta(K)^4\}.
\end{aligned}
$$

For the term (C.9), note that $\Psi_{K,\boldsymbol{h}}^\top(\boldsymbol{\theta},t,z)\Sigma_K^{-1}u_K(t',\boldsymbol{x}')$ is the weighted L^2-projection of $\overline{R}(t',\boldsymbol{x}';\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(t'-t,z'-z)$ on the space linearly spanned by $u_K(t',\boldsymbol{x}')$ with weighted measure being $-\rho''\{(\boldsymbol{\lambda}_K^*)^\top u_K(t',\boldsymbol{x}')\}dF_{X,T}(t',\boldsymbol{x}')$, i.e.

$$
\begin{aligned}
&\Sigma_K^{-1}\Psi_{K,\boldsymbol{h}}(\boldsymbol{\theta},t,z)\\
=&\operatorname*{argmin}_{\gamma}\mathbb{E}\left([-\rho''\{(\boldsymbol{\beta}_K^*)^\top u_K(T,\boldsymbol{X})\}]\left\{\overline{R}(T,\boldsymbol{X};\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T-t,Z-z)-\gamma^\top u_K(T,\boldsymbol{X})\right\}^2\right),
\end{aligned}
$$

where the projection error is of $O(K^{-\alpha})$ for some $\alpha>0$. Combining this and Lemma 1, we have

$$
\begin{aligned}
(\text{C.9})=&\sqrt{\frac{N}{h_{pro}}}\Psi_{K,\boldsymbol{h}}^\top(\boldsymbol{\theta},t,z)\Sigma_K^{-1}\cdot\left[\frac{1}{N}\sum_{i=1}^N\xi_i^{(b)}\pi_K^*(T_i,\boldsymbol{X}_i)u_K(T_i,\boldsymbol{X}_i)\right.\\
&\left.-\frac{1}{N(N-1)}\sum_{i=1,i\neq j}^N\sum_{j=1}^N u_K(T_i,\boldsymbol{X}_j)\right]\\
=&\sqrt{\frac{N}{h_{pro}}}\cdot\left[\frac{1}{N}\sum_{i=1}^N\xi_i^{(b)}\pi_0(T_i,\boldsymbol{X}_i)\overline{R}(T_i,\boldsymbol{X}_i;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)\right.\\
&\left.-\frac{1}{N(N-1)}\sum_{i=1,i\neq j}^N\sum_{j=1}^N\overline{R}(T_i,\boldsymbol{X}_j;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_j-z)\right]+o_P\{(\log N)^{-1/2}\}\\
=&\sqrt{\frac{N}{h_{pro}}}\cdot\left[\frac{1}{N}\sum_{i=1}^N\xi_i^{(b)}\pi_0(T_i,\boldsymbol{X}_i)\overline{R}(T_i,\boldsymbol{X}_i;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)\right.\\
&\left.-\mathbb{E}\left\{\overline{R}(T_1,\boldsymbol{X}_2;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T_1-t,Z_2-z)\right\}\right]\\
&+o_P\{(\log N)^{-1/2}\}\\
=&\frac{1}{\sqrt{Nh_{pro}}}\sum_{i=1}^N\left[\xi_i^{(b)}\pi_0(T_i,\boldsymbol{X}_i)\overline{R}(T_i,\boldsymbol{X}_i;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)\right.\\
&\left.-\mathbb{E}\left\{\pi_0(T_i,\boldsymbol{X}_i)\overline{R}(T_i,\boldsymbol{X}_i;\boldsymbol{\theta},t)\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)\right\}\right]+o_P\{(\log N)^{-1/2}\},
\end{aligned}
$$

where the third equality holds by using the asymptotic properties of U-statistic. Thus, we obtain that (C.6)+(C.7) is of $o_P\{(\log N)^{-1/2}\}$.

C.3 Proof of Lemma 4

For simplicity, we only proof the case $p=1$, which is also the case we will use. For general $p\geq 1$, we refer the proof of convexity lemma in Pollard (1991) for a detailed statement.

Fix $\epsilon>0$. Since convexity implies continuity, there exists $\delta>0$ such that $\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}|A(\theta_1,t,z)-A(\theta_2,t,z)|\leq\epsilon$ when $|\theta_1-\theta_2|\leq\delta$. Then we partition K into a union of interval with length δ and denote the finite set of endpoints as $\mathcal{V}$. Then by the condition, we know that

$$
\sup_{\theta\in\mathcal{V}}\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}|A_n(\theta,t,z)-A(\theta,t,z)|=o_P(1).
$$

For each θ in K, there exists $\theta_1, \theta_2 \in \mathcal{V}$ such that $\theta \in [\theta_1, \theta_2]$. Then we can write θ as $\theta = \alpha\theta_1 + (1-\alpha)\theta_2$ for some $\alpha \in [0,1]$. Then the convexity of A_n gives

$$\begin{aligned}A_n(\theta,t,z) \leq& \alpha A_n(\theta_1,t,z) + (1-\alpha)A_n(\theta_2,t,z) \\ \leq& \alpha A(\theta_1,t,z) + (1-\alpha)A(\theta_2,t,z) + \sup_{\theta\in\mathcal{V}} \sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}} |A_n(\theta,t,z) - A(\theta,t,z)| \\ \leq& A(\theta,t,z) + \epsilon + \sup_{\theta\in\mathcal{V}} \sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}} |A_n(\theta,t,z) - A(\theta,t,z)|.\end{aligned}$$

Suppose that $\theta_3 = \theta_1 - \delta \in \mathcal{V}$. Then θ_1 can be written as $\theta_1 = \beta\theta + (1-\beta)\theta_3$ with $\beta = \delta/(\delta + \theta - \theta_1) \geq 1/2$. Then by the convexity of A_n, we have

$$\begin{aligned}\beta A_n(\theta,t,z) \geq& A_n(\theta_1,t,z) - (1-\beta)A_n(\theta_3,t,z) \\ \geq& A(\theta_1,t,z) - (1-\beta)A(\theta_3,t,z) - 2\sup_{\theta\in\mathcal{V}} \sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}} |A_n(\theta,t,z) - A(\theta,t,z)| \\ \geq& A(\theta,t,z) - \epsilon - (1-\beta)\{A(\theta,t,z) + \epsilon\} - 2\sup_{\theta\in\mathcal{V}} \sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}} |A_n(\theta,t,z) - A(\theta,t,z)|\} \\ \geq& \beta A(\theta,t,z) - 2\epsilon - 2\sup_{\theta\in\mathcal{V}} \sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}} |A_n(\theta,t,z) - A(\theta,t,z)|\end{aligned}$$

Above all, we have

$$\begin{aligned}-4\epsilon - 4\sup_{\theta\in\mathcal{V}} \sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}} |A_n(\theta,t,z) - A(\theta,t,z)| \leq& A_n(\theta,t,z) - A(\theta,t,z) \\ \leq& \epsilon + \sup_{\theta\in\mathcal{V}} \sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}} |A_n(\theta,t,z) - A(\theta,t,z)|\end{aligned}$$

which means

$$\sup_{\theta\in K}\sup_{t,z} |A_n(\theta,t,z) - A(\theta,t,z)| \leq 4\epsilon + 4\sup_{\theta\in\mathcal{V}} \sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}} |A_n(\theta,t,z) - A(\theta,t,z)|$$

The implies the desired result holds.

D Proof of Theorem 2 and 4

In this section, we proof Theorem 2 and 4. Since the loss function $\mathcal{L}(\cdot)$ may not be smooth (e.g. $\mathcal{L}(v) = v\{\tau - \mathbb{1}(v \le 0)\}$ in quantile regression), the Delta method for deriving the large sample property is not applicable in our case. To circumvent this problem, we apply the *nearness of arg mins* argument in Hjort and Pollard (2011).

Let $\widehat{g}^{(0)}(t,z) := \widehat{g}(t,z)$ and $\widehat{\partial_t g}^{(0)}(t,z) := \widehat{\partial_t g}(t,z)$,

$$\widehat{\boldsymbol{\theta}}^{(b)}(t,z) := \sqrt{Nh_{pro}/\log N} \cdot \left(\widehat{g}^{(b)}(t,z) - g_0(t,z), h_0 \cdot \{\widehat{\partial_t g}^{(b)}(t,z) - \partial_t g_0(t,z)\} \right)^\top,$$

$$\boldsymbol{\theta}_0(t,z) := \sqrt{Nh_{pro}/\log N} \cdot (g_0(t,z), h_0 \cdot \partial_t g_0(t,z))^\top,\ \boldsymbol{S}_i = (1, (T_i - t)/h_0)^\top,$$

and

$$L_{N,\boldsymbol{h}}(\boldsymbol{\theta}, t, z; \widehat{\pi}_K^{(b)}) := \sum_{i=1}^N \xi_i^{(b)} \widehat{\pi}_K^{(b)}(T_i, \boldsymbol{X}_i) \mathcal{L}\left(Y_i - \frac{(\boldsymbol{\theta} + \boldsymbol{\theta}_0(t,z))^\top \boldsymbol{S}_i}{\sqrt{Nh_{pro}/\log N}} \right) \mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z),$$

for $b = 0, 1, \dots, B$. By definition, for every $(t,z) \in \mathcal{T} \times \mathcal{Z}$ and $b = 0, 1, \dots, B$,

$$\widehat{\boldsymbol{\theta}}^{(b)}(t,z) = \operatorname*{argmin}_{\boldsymbol{\theta}} L_{N,\boldsymbol{h}}(\boldsymbol{\theta}, t, z; \widehat{\pi}_K^{(b)}).$$

Note that $\xi^{(b)}$ is independent of $(\boldsymbol{X}, T, Y)$. Hence we give the proof of two theorems simultaneously. Note that

$$\begin{aligned}
\widehat{\boldsymbol{\theta}}^{(b)}(t,z) &= \operatorname*{argmin}_{\boldsymbol{\theta}} \left\{ L_{N,\boldsymbol{h}}(\theta, t, z; \widehat{\pi}_K^{(b)}) - L_{N,\boldsymbol{h}}(\boldsymbol{0}, t, z; \widehat{\pi}_K^{(b)}) \right\} \\
&= \operatorname*{argmin}_{\theta} \sum_{i=1}^N \xi_i^{(b)} \widehat{\pi}_K^{(b)}(T_i, \boldsymbol{X}_i) \Bigg\{ \mathcal{L}\left(Y_i - \frac{(\boldsymbol{\theta} + \boldsymbol{\theta}_0(t,z))^\top \boldsymbol{S}_i}{\sqrt{Nh_{pro}/\log N}} \right) \\
&\qquad - \mathcal{L}\left(Y_i - \frac{\boldsymbol{\theta}_0^\top(t,z) \boldsymbol{S}_i}{\sqrt{Nh_{pro}/\log N}} \right) \Bigg\} \mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z) \\
&= \operatorname*{argmin}_{\boldsymbol{\theta}} \Bigg[- \frac{1}{\sqrt{Nh_{pro}/\log N}} \sum_{i=1}^N \xi_i^{(b)} \widehat{\pi}_K^{(b)}(T_i, \boldsymbol{X}_i) \mathcal{L}'\left(Y_i - \frac{\boldsymbol{\theta}_0^\top(t,z) \boldsymbol{S}_i}{\sqrt{Nh_{pro}/\log N}} \right) \boldsymbol{\theta}^\top \boldsymbol{S}_i \cdot \mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z) \\
&\quad + \sum_{i=1}^N \xi_i^{(b)} \widehat{\pi}_K^{(b)}(T_i, \boldsymbol{X}_i) \Bigg\{ \mathcal{L}\left(Y_i - \frac{(\boldsymbol{\theta} + \boldsymbol{\theta}_0(t,z))^\top \boldsymbol{S}_i}{\sqrt{Nh_{pro}/\log N}} \right) - \mathcal{L}\left(Y_i - \frac{\boldsymbol{\theta}_0^\top(t,z) \boldsymbol{S}_i}{\sqrt{Nh_{pro}/\log N}} \right) \\
&\qquad + \mathcal{L}'\left(Y_i - \frac{\boldsymbol{\theta}_0^\top(t,z) \boldsymbol{S}_i}{\sqrt{Nh_{pro}/\log N}} \right) \frac{\boldsymbol{\theta}^\top \boldsymbol{S}_i}{\sqrt{Nh_{pro}/\log N}} \Bigg\} \mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z) \Bigg].
\end{aligned}$$

Define the following functions:

$$D(\boldsymbol{S}_i, Y_i, \boldsymbol{\theta}, t, z) := \mathcal{L}\left(Y_i - \frac{(\boldsymbol{\theta} + \boldsymbol{\theta}_0(t,z))^\top \boldsymbol{S}_i}{\sqrt{Nh_{pro}/\log N}} \right) - \mathcal{L}\left(Y_i - \frac{\boldsymbol{\theta}_0^\top(t,z) \boldsymbol{S}_i}{\sqrt{Nh_{pro}/\log N}} \right)$$

$$
\begin{aligned}
&+\mathcal{L}'\left(Y_i-\frac{\boldsymbol{\theta}_0^\top(t,z)\boldsymbol{S}_i}{\sqrt{Nh_{pro}/\log N}}\right)\frac{\boldsymbol{\theta}^\top\boldsymbol{S}_i}{\sqrt{Nh_{pro}/\log N}},\\
H_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z,\widehat{\pi}_K^{(b)}) :=& L_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z;\widehat{\pi}_K^{(b)})-L_{N,\boldsymbol{h}}(\mathbf{0},t,z;\widehat{\pi}_K^{(b)})\\
=&-\frac{1}{\sqrt{Nh_{pro}/\log N}}\sum_{i=1}^N\xi_i^{(b)}\widehat{\pi}_K^{(b)}(T_i,\boldsymbol{X}_i)\mathcal{L}'\left(Y_i-\frac{\boldsymbol{\theta}_0^\top(t,z)\boldsymbol{S}_i}{\sqrt{Nh_{pro}/\log N}}\right)\\
&\times\boldsymbol{\theta}^\top\boldsymbol{S}_i\cdot\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)\\
&+\sum_{i=1}^N\xi_i^{(b)}\widehat{\pi}_K^{(b)}(T_i,\boldsymbol{X}_i)D(\boldsymbol{S}_i,Y_i,\boldsymbol{\theta},t,z)\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z).
\end{aligned}
$$

Further, we define the following quadratic function

$$
\widetilde{H}_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z)=-\sqrt{\log N}\cdot\boldsymbol{\theta}^\top U_N+\frac{\log N}{2}\boldsymbol{\theta}^\top V\boldsymbol{\theta},
$$

where $U_N=(U_{N,1},U_{N,2})^\top$ with

$$
\begin{aligned}
U_{N,1}=&\frac{1}{\sqrt{Nh_{pro}}}\sum_{i=1}^N\xi_i^{(b)}\Big(\pi_0(T_i,\boldsymbol{X}_i)\mathcal{L}'\{Y_i-g_0(t,z)-\partial_t g_0(t,z)(T_i-t)\}\\
&-\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)\mathcal{L}'\{Y_i-g_0(t,z)-\partial_t g_0(t,z)(T_i-t)\}|T_i,\boldsymbol{X}_i]\Big)K_{\boldsymbol{h}}(T_i-t,Z_i-z),\\
U_{N,2}=&\frac{1}{\sqrt{Nh_{pro}}h_0}\sum_{i=1}^N\xi_i^{(b)}\Big(\pi_0(T_i,\boldsymbol{X}_i)\mathcal{L}'\{Y_i-g_0(t,z)-\partial_t g_0(t,z)(T_i-t)\}\\
&-\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)\mathcal{L}'\{Y_i-g_0(t,z)-\partial_t g_0(t,z)(T_i-t)\}|T_i,\boldsymbol{X}_i]\Big)(T_i-t)K_{\boldsymbol{h}}(T_i-t,Z_i-z),\\
V:=&\partial_a^2\widetilde{\phi}(g_0(t,z)|t,z)f_{T,Z}(t,z)\begin{pmatrix}1&0\\0&\kappa_{21}\end{pmatrix},
\end{aligned}
$$

where $\widetilde{\phi}(a|t,z)$ is defined before Assumption 9. Note that $\widetilde{H}_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z)$ does not depend on $\widehat{\pi}_K^{(b)}$, and its minimizer is defined by

$$
\widetilde{\boldsymbol{\theta}}(t,z):=\underset{\boldsymbol{\theta}}{\operatorname{argmin}}\,\widetilde{H}_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z)=\frac{V^{-1}U_N}{\sqrt{\log N}}.
$$

We complete the proof via the following steps: for $b=0,1,\ldots,B$,

-
- Step I: for every fixed $\boldsymbol{\theta}$, show that $\xi_{N,\boldsymbol{h}}(\boldsymbol{\theta},\widehat{\pi}_K^{(b)}):=\widetilde{H}_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z)-H_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z,\widehat{\pi}_K^{(b)})=o_P(1)$ uniformly over $(t,z)\in\mathcal{T}\times\mathcal{Z}$;
- Step II: show that $\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}\|\widehat{\boldsymbol{\theta}}^{(b)}(t,z)-\widetilde{\boldsymbol{\theta}}(t,z)\|=o_P\{(\log N)^{-1}\}$;

- Step III: obtain the following results:

$$\sqrt{Nh_{pro}}\{\widehat{g}^{(b)}(t,z)-g_0(t,z)\}=\frac{1}{\sqrt{Nh_{pro}}}\sum_{i=1}^{N}\xi_i^{(b)}\psi_{0,\boldsymbol{h}}(T_i,\boldsymbol{X}_i,Y_i;t,z)+o_P\left(\frac{1}{\sqrt{\log N}}\right),$$

and

$$\sqrt{Nh_{pro}}h_0\left\{\widehat{\partial_t g}^{(b)}(t,z)-\partial_t g_0(t,z)\right\}=\frac{1}{\sqrt{Nh_{pro}}h_0}\sum_{i=1}^{N}\xi_i^{(b)}\psi_{1,\boldsymbol{h}}(T_i,\boldsymbol{X}_i,Y_i;t,z)+o_P\left(\frac{1}{\sqrt{\log N}}\right),$$

where all the $o_P(\cdot)$ holds uniformly over $(t,z)\in\mathcal{T}\times\mathcal{Z}$, $\psi_{0,\boldsymbol{h}}$ and $\psi_{1,\boldsymbol{h}}$ are defined in Theorem 2.

- Step IV: taking the differences:

$$\sqrt{Nh_{pro}}\{\widehat{g}^{(b)}(t,z)-g_0(t,z)\}-\sqrt{Nh_{pro}}\{\widehat{g}^{(0)}(t,z)-g_0(t,z)\}$$

and

$$\sqrt{Nh_{pro}}h_0\left\{\widehat{\partial_t g}^{(b)}(t,z)-\partial_t g_0(t,z)\right\}-\sqrt{Nh_{pro}}h_0\left\{\widehat{\partial_t g}^{(0)}(t,z)-\partial_t g_0(t,z)\right\},$$

for $b=1,\ldots,B$, gives the desired result in Theorem 4.

We start Step I by showing that $\widetilde{H}_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z)$ is a quadratic approximation to the objective function $H_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z,\widehat{\pi}_K^{(b)})$. Note that $\boldsymbol{\theta}=(\alpha,\beta)^\top$, we have

$$\begin{aligned}
&\left|\widetilde{H}_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z)-H_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z,\widehat{\pi}_K^{(b)})\right|\\
\leq&\left|\frac{1}{\sqrt{Nh_{pro}/\log N}}\sum_{i=1}^{N}\xi_i^{(b)}\widehat{\pi}_K^{(b)}(T_i,\boldsymbol{X}_i)\mathcal{L}'\left(Y_i-\frac{\boldsymbol{\theta}_0^\top(t,z)\boldsymbol{S}_i}{\sqrt{Nh_{pro}/\log N}}\right)\right.\\
&\qquad\left.\times\boldsymbol{\theta}^\top\boldsymbol{S}_i\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)-\sqrt{\log N}\boldsymbol{\theta}^\top U_N\right|\\
+&\left|\frac{\log N}{2}\boldsymbol{\theta}^\top V\boldsymbol{\theta}-\sum_{i=1}^{N}\xi_i^b\widehat{\pi}_K^{(b)}(T_i,\boldsymbol{X}_i)D(\boldsymbol{S}_i,Y_i,\boldsymbol{\theta},t,z)\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)\right|\\
\leq&|\alpha|\cdot\sqrt{\log N}\left|\frac{1}{\sqrt{Nh_{pro}}}\sum_{i=1}^{N}\xi_i^{(b)}\left[\widehat{\pi}_K^{(b)}(T_i,\boldsymbol{X}_i)\mathcal{L}'\{Y_i-g_0(t,z)-\partial_t g_0(t,z)(T_i-t)\}\right.\right.\\
&\qquad-\pi_0(T_i,\boldsymbol{X}_i)\mathcal{L}'\{Y_i-g_0(t,z)-\partial_t g_0(t,z)(T_i-t)\}\\
&\qquad\left.\left.+\mathbb{E}[\pi_0(T_i,\boldsymbol{X}_i)\mathcal{L}'\{Y_i-g_0(t,z)-\partial_t g_0(t,z)(T_i-t)\}|T_i,\boldsymbol{X}_i]\right]\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)\right|\\
&+|\beta|\cdot\sqrt{\log N}\left|\frac{1}{\sqrt{Nh_{pro}}}\sum_{i=1}^{N}\xi_i^{(b)}\left(\widehat{\pi}_K^{(b)}(T_i,\boldsymbol{X}_i)\mathcal{L}'\{Y_i-g_0(t,z)-\partial_t g_0(t,z)(T_i-t)\}\right.\right.
\end{aligned}$$

$$- \pi_0(T_i, \boldsymbol{X}_i)\mathcal{L}'(Y_i - g_0(t,z) - \partial_t g_0(t,z)(T_i - t))$$

$$\left. + \mathbb{E}\{\pi_0(T_i, \boldsymbol{X}_i)\mathcal{L}'(Y_i - g_0(t,z) - \partial_t g_0(t,z)(T_i - t))|T_i, \boldsymbol{X}_i\}\right) \frac{T_i - t}{h_0}\mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z)\Bigg|$$

$$+ \left| \frac{\log N}{2}\boldsymbol{\theta}^\top V \boldsymbol{\theta} - \sum_{i=1}^N \xi_i^{(b)} \widehat{\pi}_K^{(b)}(T_i, \boldsymbol{X}_i) D(\boldsymbol{S}_i, Y_i, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z) \right|$$

$$= |\alpha| \cdot \sqrt{\log N}\, |\xi_{1,N}| + |\beta| \cdot \sqrt{\log N}\, |\xi_{2,N}| + |\xi_{3,N}|\,,$$

where the definitions of ξ_{1N}, ξ_{2N} and ξ_{3N} are obvious. Hence, Step I holds if we can prove that for every fixed $\boldsymbol{\theta} \in [-M, M]^2$, the following result holds:

$$\xi_{1,N} = o_P(\{\log N\}^{-1/2}),\ \xi_{2,N} = o_P(\{\log N\}^{-1/2}),\ \xi_{3,N} = o_P(1). \tag{D.1}$$

Taking $\phi_N(T, \boldsymbol{X}, Y; t)$ in the definition of $R(T, \boldsymbol{X}, Y; \boldsymbol{\theta}, t)$ in Lemma 3 to be 1 or $(T - t)/h_0$, and using Assumption 11 that $\left(\sum_{s=0}^{d_z} h_s^2\right)^2 \prec (N h_{pro})^{-1}$, we have $\xi_{1,N} = o_P(\{\log N\}^{-1/2})$ and $\xi_{2,N} = o_P(\{\log N\}^{-1/2})$. We next prove $\xi_{3,N} = o_P(1)$. Note that

$$|\xi_{3,N}| \leq \left| \sum_{i=1}^N \xi_i^{(b)} \left\{ \widehat{\pi}_K^{(b)}(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i) \right\} D(\boldsymbol{S}_i, Y_i, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z) \right| \tag{D.2}$$

$$+ \Bigg| \sum_{i=1}^N \xi_i^{(b)} \pi_0(T_i, \boldsymbol{X}_i) D(\boldsymbol{S}_i, Y_i, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z)$$

$$- N \cdot \mathbb{E}\left[\xi_i^{(b)} \pi_0(T, \boldsymbol{X}) R(\boldsymbol{S}, Y, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T - t, Z - z) \right] \Bigg| \tag{D.3}$$

$$+ \left| N \cdot \mathbb{E}\left[\xi_i^{(b)} \pi_0(T, \boldsymbol{X}) R(\boldsymbol{S}, Y, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T - t, Z - z) \right] - \frac{\log N}{2}\boldsymbol{\theta}^\top V \boldsymbol{\theta} \right|. \tag{D.4}$$

For (D.2), we first note that a differentiable function f is convex on an interval if and only if its graph lies above all its tangents: $f(x) \geq f(y) + f'(y)(x - y)$ for all x, y on the interval. Thus, the convexity of $\mathcal{L}(\cdot)$ implies $D(\boldsymbol{S}, Y, \boldsymbol{\theta}, t, z) \geq 0$ almost surely. Then, using the fact that $\xi^{(b)} \geq 0$ and Lemma 1, we have

$$\text{(D.2)} = \left| \sum_{i=1}^N \xi_i^{(b)} \left\{ \widehat{\pi}_K^{(b)}(T_i, \boldsymbol{X}_i) - \pi_0(T_i, \boldsymbol{X}_i) \right\} D(\boldsymbol{S}_i, Y_i, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z) \right|$$

$$\leq \sup_{(t, \boldsymbol{x}) \in \mathcal{T} \times \mathcal{X}} |\widehat{\pi}_K^{(b)}(t, \boldsymbol{x}) - \pi_0(t, \boldsymbol{x})| \cdot \sum_{i=1}^N \xi_i^{(b)} D(\boldsymbol{S}_i, Y_i, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z)$$

$$= o_P\{(\log N)^{-1}\} \cdot \sum_{i=1}^N \xi_i^{(b)} D(\boldsymbol{S}_i, Y_i, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z)\,. \tag{D.5}$$

We decompose (D.5) as follows:

$$\text{(D.5)} = o_P\{(\log N)^{-1}\} \cdot \sum_{i=1}^N \Big[\xi_i^{(b)} D(\boldsymbol{S}_i, Y_i, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z)$$

$$-\mathbb{E}\{D(\boldsymbol{S}_i, Y_i, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z)\}\Big] \tag{D.6}$$

$$+o_P\{(\log N)^{-1}\} \cdot N \cdot \mathbb{E}\{D(\boldsymbol{S}_i, Y_i, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T_i - t, Z_i - z)\}\,. \tag{D.7}$$

To show that (D.5) is of $o_p(1)$ uniformly in $(t, z) \in \mathcal{T} \times \mathcal{Z}$, we will calculate the uniform rate of the first moment of $D(\boldsymbol{S}, Y, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T - t, Z - z)$ for (D.7), and apply the empirical process theories on (D.6), which will require the rate of the second moment of $D(\boldsymbol{S}, Y, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T - t, Z - z)$.

Thereby, we next compute the first and second moments of $D(\boldsymbol{S}, Y, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T-t, Z-z)$. Recalling the definition of ϕ before Assumption 9, the first moment of $D(\boldsymbol{S}, Y, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T - t, Z - z)$ is given by

$$\begin{aligned}
&\mathbb{E}\left[D(\boldsymbol{S}, Y, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T - t, Z - z)\right] = \mathbb{E}\left[\mathbb{E}\{D(\boldsymbol{S}, Y, \boldsymbol{\theta}, t, z)|T, Z\}\mathcal{K}_{\boldsymbol{h}}(T - t, Z - z)\right] \\
=&\mathbb{E}\left[\left\{\phi\left(\frac{(\boldsymbol{\theta} + \boldsymbol{\theta}_0(t, z))^\top \boldsymbol{S}}{\sqrt{Nh_{pro}\,/\log N}}\middle| T, Z\right) - \phi\left(\frac{\boldsymbol{\theta}_0^\top(t, z)\boldsymbol{S}}{\sqrt{Nh_{pro}\,/\log N}}\middle| T, Z\right)\right.\right. \\
&\left.\left.\quad - \partial_a\phi\left(\frac{\boldsymbol{\theta}_0^\top(t, z)\boldsymbol{S}}{\sqrt{Nh_{pro}\,/\log N}}\middle| T, Z\right) \cdot \frac{\boldsymbol{\theta}^\top \boldsymbol{S}}{\sqrt{Nh_{pro}/\log N}}\right\} \cdot \mathcal{K}_{\boldsymbol{h}}(T - t, Z - z)\right] \\
=&\frac{\log N}{2N} \cdot \mathbb{E}\left[\partial_a^2\phi\left(\frac{\boldsymbol{\theta}_0^\top(t, z)\boldsymbol{S}}{\sqrt{Nh_{pro}/\log N}}\middle| T, Z\right) \cdot \{\boldsymbol{\theta}^\top \boldsymbol{S}\}^2 \cdot \frac{1}{h_{pro}}\mathcal{K}_{\boldsymbol{h}}(T - t, Z - z)\right] \cdot \{1 + o(1)\} = O\left(\frac{\log N}{N}\right),
\end{aligned}$$

where the $o(\cdot)$ and $O(\cdot)$ are uniformly in $(t, z) \in \mathcal{T} \times \mathcal{Z}$, the second equality holds by the Taylor's theorem and the smoothness of $\phi(\cdot|T, Z)$ and the last equality follows from the fact that for any integer $\ell = 1, 2$ and $0 \le j \le 4$,

$$\begin{aligned}
&\mathbb{E}\left[\left\{\partial_a^2\phi\left(\frac{\boldsymbol{\theta}_0^\top(t, z)\boldsymbol{S}_i}{\sqrt{Nh_{pro}\,/\log N}}\middle| T, Z\right)\right\}^\ell \left(\frac{T - t}{h_0}\right)^j \frac{1}{h_{pro}}\mathcal{K}_{\boldsymbol{h}}(T - t, Z - z)\right] \qquad \text{(D.8)} \\
=&\mathbb{E}\left[[\partial_a^2\phi\{g_0(t, z) + \partial_t g_0(t, z)(T - t)|T, Z\}]^\ell \left(\frac{T - t}{h_0}\right)^j \frac{1}{h_{pro}}\mathcal{K}_{\boldsymbol{h}}(T - t, Z - z)\right] \\
=&\int [\partial_a^2\phi\,\{g_0(t, z) + \partial_t g_0(t, z)(\widetilde{t} - t)|\widetilde{t}, \widetilde{\boldsymbol{z}}\}]^\ell f_{T,Z}(\widetilde{t}, \widetilde{\boldsymbol{z}}) \left(\frac{\widetilde{t} - t}{h_0}\right)^j \frac{1}{h_{pro}}\mathcal{K}_{\boldsymbol{h}}(\widetilde{t} - t, \widetilde{\boldsymbol{z}} - z)d\widetilde{t}d\widetilde{\boldsymbol{z}} \\
=&\int [\partial_a^2\phi\,\{g_0(t, z) + h_0\partial_t g_0(t, z)\widetilde{t}|t + h_0\widetilde{t}, z + h_{\boldsymbol{Z}} \cdot \widetilde{\boldsymbol{z}}\}]^\ell f_{T,Z}(t + h\widetilde{t}, z + h_{\boldsymbol{Z}} \cdot \widetilde{\boldsymbol{z}})\widetilde{t}^j\mathcal{K}(\widetilde{t})\mathcal{K}(\widetilde{z}_1)...\mathcal{K}(\widetilde{z}_{d_Z})d\widetilde{t}d\widetilde{\boldsymbol{z}} \\
=&[\partial_a^2\phi\{g_0(t, z)|t, z\}]^\ell f_{T,Z}(t, z) \int \widetilde{t}^j\mathcal{K}(\widetilde{t})\mathcal{K}(\widetilde{z}_1)...\mathcal{K}(\widetilde{z}_{d_Z})d\widetilde{t}d\widetilde{\boldsymbol{z}} \cdot \left\{1 + O\left(\sum_{s=0}^{d_Z} h_s^2\right)\right\} \\
=&[\partial_a^2\phi\{g_0(t, z)|t, z\}]^\ell f_{T,Z}(t, z)\kappa_{j1} + o\{(\log N)^{-1}\}\,,
\end{aligned}$$

where the $o(\cdot)$ is uniformly in $(t, z) \in \mathcal{T} \times \mathcal{Z}$. Similarly, the second moment of $D(\boldsymbol{S}, Y, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T - t, Z - z)$ is given by

$$\begin{aligned}
&\mathbb{E}\left[D(\boldsymbol{S}, Y, \boldsymbol{\theta}, t, z)^2\mathcal{K}_{\boldsymbol{h}}^2(T - t, Z - z)\right] \qquad \text{(D.9)} \\
=&\mathbb{E}\left(\mathbb{E}\left[D(\boldsymbol{S}, Y, \boldsymbol{\theta}, t, z)^2|T, Z\right]\mathcal{K}_{\boldsymbol{h}}^2(T - t, Z - z)\right) \\
=&\frac{(\log N)^2}{(2N)^2} \cdot \mathbb{E}\left[\left\{\partial_a^2\phi\left(\frac{\boldsymbol{\theta}_0^\top(t, z)\boldsymbol{S}}{\sqrt{Nh_{pro}/\log N}}\middle| T, Z\right)\right\}^2 \cdot \{\boldsymbol{\theta}^\top \boldsymbol{S}\}^4 \cdot \frac{1}{h_{pro}^2}\mathcal{K}_{\boldsymbol{h}}^2(T - t, Z - z)\right] \cdot \{1 + o(1)\}
\end{aligned}$$

$$=O\left(\frac{\log^2 N}{N^2 h_{pro}}\right),$$

where the $O(\cdot)$ is uniformly in $(t,z) \in \mathcal{T} \times \mathcal{Z}$. With the result of the first moment of $D(\boldsymbol{S}, Y, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T-t, Z-z)$, we have

$$\text{(D.7)} = O\left(N \cdot \frac{\log N}{N}\right) \cdot o_P(\{\log N\}^{-1}) = o_P(1)$$

uniformly over $(t,z) \in \mathcal{T} \times \mathcal{Z}$.

We are now ready to apply the empirical process theories on (D.6). Define the function class

$$\mathcal{F}_{DK} = \{(\xi^{(b)}, y, t, z) \mapsto \xi^{(b)} D(\boldsymbol{s}, y, \boldsymbol{\theta}, \widetilde{t}, \widetilde{z})\mathcal{K}_{\boldsymbol{h}}(t-\widetilde{t}, z-\widetilde{z}) : (\widetilde{t}, \widetilde{z}) \in \mathcal{T} \times \mathcal{Z}\}.$$

By the definition of $D(\boldsymbol{s}, y, \boldsymbol{\theta}, \widetilde{t}, \widetilde{z})$, for sufficiently large N, using the mean value theorem, we have

$$\begin{aligned}
&\sup_{(\widetilde{t},\widetilde{z}) \in \mathcal{T}\times\mathcal{Z}} |\xi^{(b)} \cdot D(\boldsymbol{s}, y, \boldsymbol{\theta}, \widetilde{t}, \widetilde{z})\mathcal{K}_{\boldsymbol{h}}(t-\widetilde{t}, z-\widetilde{z})| \\
=&2 \sup_{(\widetilde{t},\widetilde{z}) \in \mathcal{T}\times\mathcal{Z}} \Bigg| \mathcal{L}'\left(y - g_0(\widetilde{t}, \widetilde{z}) - \partial_t g_0(\widetilde{t}, \widetilde{z})(t-\widetilde{t}) - \epsilon(\widetilde{t}, \widetilde{z}) \frac{\boldsymbol{\theta}^\top \boldsymbol{s}}{\sqrt{N h_{pro}/\log N}}\right) \\
&\qquad - \mathcal{L}'\left(y - g_0(\widetilde{t}, \widetilde{z}) - \partial_t g_0(\widetilde{t}, \widetilde{z})(t-\widetilde{t})\right)\Bigg| \cdot \frac{|\boldsymbol{\theta}^\top \boldsymbol{s}|}{\sqrt{N h_{pro}/\log N}} \mathcal{K}_{\boldsymbol{h}}(t-\widetilde{t}, z-\widetilde{z}) \\
\leq&4 \cdot \sup_{\widetilde{a},\widetilde{b} \in [-M,M]} |\mathcal{L}'(y - \widetilde{a} - \widetilde{b}t)| \cdot \frac{|\boldsymbol{\theta}^\top \boldsymbol{s}|}{\sqrt{N h_{pro}/\log N}} \mathcal{K}_{\boldsymbol{h}}(t-\widetilde{t}, z-\widetilde{z}) := F_{DK}(\boldsymbol{s}, y, \boldsymbol{\theta}),
\end{aligned}$$

where M is defined in Assumption 10(i) and $\epsilon(t,z) \in [0,1]$ comes from the mean value theorem. By Proposition 3.6.12 in Giné and Nickl (2021), we know that

$$\sup_Q \log \mathcal{N}\left(\left\{\mathcal{K}\left(\frac{\cdot - u}{h}\right) : u \in \mathbb{R}\right\}, \|\cdot\|_{Q,2}, \tau\right) \lesssim \log(1/\tau).$$

Noting that the VC type property is "stable"under summation, product and even Lipschlitz-type transformation of VC-type function class (see e.g. Chapter 2 in Van Der Vaart and Wellner, 1996). Then, denoting $F_1(s,y) = \sup_{\widetilde{a},\widetilde{b} \in [-M,M]} |\mathcal{L}(y - \widetilde{a} - \widetilde{b}t)|$ and $F_2(s,y) = \sup_{\widetilde{a},\widetilde{b} \in [-M,M]} |\mathcal{L}'(y - \widetilde{a} - \widetilde{b}t)|$, by Assumption 10(i), we have that

$$\sup_Q \log \mathcal{N}\left(\left\{\mathcal{L}\left(Y - \frac{(\boldsymbol{\theta} + \boldsymbol{\theta}(t,z))^\top \boldsymbol{S}}{\sqrt{N h_{pro}/\log N}}\right) : (t,z) \in \mathcal{T} \times \mathcal{Z}\right\}, \|\cdot\|_{Q,2}, \tau\|F_1\|_{Q,2}\right) \lesssim \log(1/\tau)$$

and

$$\sup_Q \log \mathcal{N}\left(\left\{\mathcal{L}'\left(Y - \frac{(\boldsymbol{\theta} + \boldsymbol{\theta}(t,z))^\top \boldsymbol{S}}{\sqrt{N h_{pro}/\log N}}\right) : (t,z) \in \mathcal{T} \times \mathcal{Z}\right\}, \|\cdot\|_{Q,2}, \tau\|F_2\|_{Q,2}\right) \lesssim \log(1/\tau).$$

Since the VC type property is stable under summation and product, by the facts above, we can conclude that

$$\sup_Q \log \mathcal{N}\left(\mathcal{F}_{DK}, \|\cdot\|_{Q,2}, \tau\|F_{DK}\|_{Q,2}\right) \lesssim \log(1/\tau).$$

Then, to apply Chernozhukov et al. (2014b, Corollary 5.1), we calculate some quantities in their notations. By the result of the second moment of $D(\boldsymbol{S}, Y, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T-t, Z-z)$ in (D.9) and the fact that $\xi^{(b)}$ is independent of the data and with a second moment of 1 ($b=0$) or 2 ($b=1,\ldots,B$), we have

$$\begin{aligned}
&\sigma^2 := \sup_{(\tilde{t},\tilde{z})} \mathbb{E}[|\xi^{(b)} D(\boldsymbol{S}, Y, \boldsymbol{\theta}, \tilde{t}, \tilde{z})\mathcal{K}_{\boldsymbol{h}}(T-\tilde{t}, Z-\tilde{z})|^2] \lesssim \frac{\log^2 N}{N^2 h_{pro}},\\
&\|F_{DK}\|_{P,2}^2 \lesssim \frac{\log N}{N} \quad \text{(From (D.8))}\\
&\|M\|_2^2 := \mathbb{E}\left\{\max_{1\le i\le n} F_{DK}^2(T_i, \boldsymbol{X}_i, Y_i)\right\} \le \left[\mathbb{E}\left\{\max_{1\le i\le n} F_{DK}^q(T_i, \boldsymbol{X}_i, Y_i)\right\}\right]^{2/q}\\
&\qquad\qquad \le N^{2/q}\left[\mathbb{E}\left\{F_{DK}^q(T_i, \boldsymbol{X}_i, Y_i)\right\}\right]^{2/q} \le \text{const} \times N^{2/q}/\left(N h_{pro} h_0^2/\log N\right).
\end{aligned}$$

Then, denoting the controlling term of (D.6) by $\mathbb{G}_N$, by Chernozhukov et al. (2014b, Corollary 5.1), we have

$$\|\mathbb{G}_N\|_{\mathcal{F}_{DK}} = O_P\left\{\left(\sqrt{\frac{\log^2 N}{N h_{pro}}} + \frac{N^{1/q}\sqrt{\log N}}{\sqrt{N h_{pro}} h_0}\right)\cdot(\log N)\right\}.$$

Then, we can conclude that

$$\text{(D.6)} = o_P(1)$$

holds uniformly over $(t,z) \in \mathcal{T}\times\mathcal{Z}$. Combining (D.6) and (D.7), we have that (D.2) is of $o_P(1)$ uniformly over $(t,z)\in\mathcal{T}\times\mathcal{Z}$.

For (D.3), repeating the arguments using in controlling term (D.6), we have

$$\text{(D.3)} = O_P\left(\sqrt{\frac{\log^2 N}{N h_{pro}}} + \frac{N^{1/q}\sqrt{\log N}}{\sqrt{N h_{pro}} h_0}\right) = o_P(1).$$

For (D.4), note that

$$\begin{aligned}
&N\cdot\mathbb{E}\left[\pi_0(T,\boldsymbol{X}) D(\boldsymbol{S}, Y, \boldsymbol{\theta}, t, z)\mathcal{K}_{\boldsymbol{h}}(T-t, Z-z)\right]\\
=&N\cdot\mathbb{E}\left[\mathbb{E}\left[\pi_0(T,\boldsymbol{X}) D(\boldsymbol{S}, Y, \boldsymbol{\theta}, t, z)|T,Z\right]\mathcal{K}_{\boldsymbol{h}}(T-t, Z-z)\right]\\
=&N\cdot\mathbb{E}\Bigg\{\Bigg[\tilde{\phi}\left(\frac{(\boldsymbol{\theta}+\boldsymbol{\theta}_0(t,z))^\top \boldsymbol{S}}{\sqrt{N h_{pro}/\log N}}\middle| T,Z\right) - \tilde{\phi}\left(\frac{\boldsymbol{\theta}_0^\top(t,z) S}{\sqrt{N h_{pro}/\log N}}\middle| T,Z\right)\\
&\qquad - \partial_a\tilde{\phi}\left(\frac{\boldsymbol{\theta}_0^\top(t,z)\boldsymbol{S}}{\sqrt{N h_{pro}/\log N}}\middle| T,Z\right)\cdot\frac{\boldsymbol{\theta}^\top(t,z)\boldsymbol{S}}{\sqrt{N h_{pro}/\log N}}\Bigg]\cdot\mathcal{K}_{\boldsymbol{h}}(T-t, Z-z)\Bigg\}
\end{aligned}$$

$$=\frac{N}{2}\cdot\mathbb{E}\left\{\partial_a^2\widetilde{\phi}\left(\frac{\boldsymbol{\theta}_0^\top(t,z)\boldsymbol{S}}{\sqrt{Nh_{pro}/\log N}}\middle|T,Z\right)\frac{(\boldsymbol{\theta}^\top S)^2}{Nh_{pro}/\log N}\mathcal{K}_{\boldsymbol{h}}(T-t,Z-z)\right\}\cdot\{1+o(1)\}$$

$$=\frac{\log N}{2}\cdot\boldsymbol{\theta}^\top\mathbb{E}\left[\partial_a^2\widetilde{\phi}\left(\frac{\boldsymbol{\theta}_0^\top(t,z)\boldsymbol{S}}{\sqrt{Nh_{pro}}}\middle|T,Z\right)\boldsymbol{S}\boldsymbol{S}^\top\frac{1}{h_{pro}}\mathcal{K}_{\boldsymbol{h}}(T-t,Z-z)\right]\boldsymbol{\theta}\cdot\{1+o(1)\}. \quad \text{(D.10)}$$

Then, using (D.8), we have

$$\text{(D.10)}=\frac{\log N}{2}\cdot\boldsymbol{\theta}^\top V\boldsymbol{\theta}+o(1)$$

which implies (D.4) $=o(1)$ uniformly over $(t,z)\in\mathcal{T}\times\mathcal{Z}$. So for every fixed θ,

$$\xi_{N,\boldsymbol{h}}(\boldsymbol{\theta},\widehat{\pi}_K^{(b)})=\widetilde{H}_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z)-H_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z,\widehat{\pi}_K^{(b)})=o_P(1), \quad \text{(D.11)}$$

uniformly over $(t,z)\in\mathcal{T}\times\mathcal{Z}$ and the proof of Step I is completed.

To prove Step II, we first show that $\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}|\widetilde{\boldsymbol{\theta}}(t,z)|$ is asymptotically stochastic bounded. By definition of $\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}\widetilde{\boldsymbol{\theta}}(t,z)$, it's sufficient to prove that

$$\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}|U_{N,1}|=O_P\left(\sqrt{\log N}\right)\text{ and }\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}|U_{N,2}|=O_P\left(\sqrt{\log N}\right).$$

To show that, we apply Chernozhukov et al. (2014b, Corollary 5.1) again. We denote

$$\mathcal{F}_{U1}=\Big\{(\xi^{(b)},Y,T,\boldsymbol{X})\mapsto\frac{1}{\sqrt{h_{pro}}}\Big[\xi^{(b)}\pi_0(T,\boldsymbol{X})\mathcal{L}'(Y-g_0(t,z)-\partial_t g_0(t,z)(T-t))$$
$$-\mathbb{E}\{\pi_0(T,\boldsymbol{X})\mathcal{L}'(Y-g_0(t,z)-\partial_t g_0(t,z)(T-t))|T,\boldsymbol{X}\}\Big]K_{\boldsymbol{h}}(T-t,Z-z):(t,z)\in\mathcal{T}\times\mathcal{Z}\Big\}$$

By Assumption 10, we know that $\mathcal{F}_{U1}$ is a VC-type class. By applying Chernozhukov et al. (2014b, Corollary 5.1), we can conclude that

$$\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}|U_{N,1}|=O\left(\sqrt{\log N}\right)+O\left(\sqrt{\log N}\cdot\frac{N^{1/q}}{\sqrt{Nh_{pro}}}\right)=O\left(\sqrt{\log N}\right)$$

By similar argument, we can conclude that

$$\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}|U_{N,2}|=O\left(\sqrt{\log N}\right)$$

Then, by Markov's inequality, there exists constant $C>0$ such that

$$\mathbb{P}\left(\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}|\widetilde{\boldsymbol{\theta}}(t,z)|\leq C\cdot(1,1)^\top\right)\to 1$$

In the following, we consider the event set $\mathcal{E}_N$, defined as

$$\mathcal{E}_N:=\left\{\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}|\widetilde{\boldsymbol{\theta}}(t,z)|\leq C\cdot(1,1)^\top\right\}$$

We finish step II by showing (D.12) and (D.13) below, for every $\varepsilon > 0$,

$$\begin{aligned}
&\mathbb{P}\left(\left\{\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}} \log N \cdot \|\widehat{\boldsymbol{\theta}}^{(b)}(t,z) - \widetilde{\boldsymbol{\theta}}(t,z)\| \geq \varepsilon\right\} \bigcap \mathcal{E}_N\right) \\
\leq &\mathbb{P}\left(\sup_{\|\boldsymbol{\theta}-\boldsymbol{\theta}_0\|\leq\varepsilon} \sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}} |\xi_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z,\widehat{\pi}_K)| \geq \inf_{(t,z)\in\mathcal{T}\times\mathcal{Z}} \frac{1}{4}\cdot \partial_a^2\widetilde{\phi}(g_0(t,z))f_{T,Z}(t,z)\cdot(\kappa_{21}\wedge 1)\cdot\varepsilon^2\right) \qquad \text{(D.12)} \\
= &o(1). \qquad \text{(D.13)}
\end{aligned}$$

For (D.12), note that $L_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z;\widehat{\pi}_K) = \sum_{i=1}^N \widehat{\pi}_K(T_i,W_i,Z_i)\mathcal{L}(Y_i-\alpha-\beta(T_i-t))\mathcal{K}_{\boldsymbol{h}}(T_i-t,Z_i-z)$ is convex in $\boldsymbol{\theta}=(\alpha,\beta)$ for every N. As a result, $H_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z,\widehat{\pi}_K^{(b)})$ is convex in $\boldsymbol{\theta}$ for every N. In light of the convexity of $H_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z,\widehat{\pi}_K^{(b)})$, for any fixed $d>\epsilon/\log N$ and any $(t,z)\in\mathcal{T}\times\mathcal{Z}$, we have that for any $\boldsymbol{\theta}(t,z)$ such that $\|\boldsymbol{\theta}(t,z)-\widetilde{\boldsymbol{\theta}}(t,z)\|=d$,

$$\begin{aligned}
&\left(1-\frac{\varepsilon}{d\log N}\right)\cdot H_{N,\boldsymbol{h}}\{\widetilde{\boldsymbol{\theta}}(t,z),t,z,\widehat{\pi}_K\} + \frac{\varepsilon}{d\log N}\cdot H_{N,\boldsymbol{h}}\{\boldsymbol{\theta}(t,z),t,z,\widehat{\pi}_K^{(b)}\} \\
&\quad\geq H_{N,\boldsymbol{h}}\left[\widetilde{\boldsymbol{\theta}}(t,z)-\frac{\varepsilon}{d\log N}\{\widetilde{\boldsymbol{\theta}}(t,z)-\boldsymbol{\theta}(t,z)\},t,z,\widehat{\pi}_K^{(b)}\right].
\end{aligned}$$

Then under $\mathcal{E}_N$, we have that

$$\begin{aligned}
&\frac{\varepsilon}{d\log N}\left[H_{N,\boldsymbol{h}}\{\boldsymbol{\theta}(t,z),t,z,\widehat{\pi}_K^{(b)}\} - H_{N,\boldsymbol{h}}\{\widetilde{\boldsymbol{\theta}}(t,z),t,z,\widehat{\pi}_K^{(b)}\}\right] \\
\geq &\widetilde{H}_{N,\boldsymbol{h}}\left[\widetilde{\boldsymbol{\theta}}(t,z)-\frac{\varepsilon}{d\log N}\{\widetilde{\boldsymbol{\theta}}(t,z)-\boldsymbol{\theta}(t,z)\},t,z\right] - \xi_{N,\boldsymbol{h}}\left[\widetilde{\boldsymbol{\theta}}(t,z)-\frac{\varepsilon}{d\log N}\{\widetilde{\boldsymbol{\theta}}(t,z)-\boldsymbol{\theta}(t,z)\},t,z,\widehat{\pi}_K^{(b)}\right] \\
&\quad-\left[\widetilde{H}_{N,\boldsymbol{h}}\{\widetilde{\boldsymbol{\theta}}(t,z),t,z\} - \xi_{N,\boldsymbol{h}}\{\widetilde{\boldsymbol{\theta}}(t,z),t,z,\widehat{\pi}_K^{(b)}\}\right] \\
\geq &-\sup_{t,z}|\xi_{N,\boldsymbol{h}}[\widetilde{\boldsymbol{\theta}}(t,z)-\frac{\varepsilon}{d\log N}\{\widetilde{\boldsymbol{\theta}}(t,z)-\boldsymbol{\theta}(t,z)\},t,z,\widehat{\pi}_K^{(b)}]| - \sup_{t,z}|\xi_{N,\boldsymbol{h}}\{\widetilde{\boldsymbol{\theta}}(t,z),t,z,\widehat{\pi}_K^{(b)}\}| \\
&+\inf_{t,z}\inf_{\|\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}(t,z)\|=\varepsilon/\log N}\left[\widetilde{H}_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z)-\widetilde{H}_{N,\boldsymbol{h}}\{\widetilde{\boldsymbol{\theta}}(t,z),t,z\}\right] \\
&\geq -2\sup_{\boldsymbol{\theta}:|\boldsymbol{\theta}|\leq(C+\varepsilon/\log N)\cdot(1,1)^\top}\sup_{t,z}\left|\xi_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z,\widehat{\pi}_K^{(b)})\right| + \inf_{t,z}\inf_{\|\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}(t,z)\|=\varepsilon/\log N}\left[\widetilde{H}_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z)-\widetilde{H}_{N,\boldsymbol{h}}\{\widetilde{\boldsymbol{\theta}}(t,z),t,z\}\right],
\end{aligned}$$

where the second inequality holds because $\|\boldsymbol{\theta}(t,z)-\widetilde{\boldsymbol{\theta}}(t,z)\|=d$. Recall that $\widetilde{\boldsymbol{\theta}}(t,z)$ is the minimizer of the quadratic function $\widetilde{H}_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z)$, expanding $\widetilde{H}_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z)$ around $\widetilde{\boldsymbol{\theta}}$, we have

$$\inf_{\|\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}(t,z)\|=\varepsilon/\log N}\left[\widetilde{H}_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z)-\widetilde{H}_{N,\boldsymbol{h}}\{\widetilde{\boldsymbol{\theta}}(t,z),t,z\}\right] = \frac{1}{2}\partial_a^2\widetilde{\phi}(g_0(t,z))f_{T,Z}(t,z)\cdot(\kappa_{21}\wedge 1)\cdot\varepsilon^2$$

Therefore, if

$$\sup_{\boldsymbol{\theta}:|\boldsymbol{\theta}|\leq(C+\varepsilon/\log N)\cdot(1,1)^\top}\sup_{(t,z)\in\mathcal{T}\times\mathcal{Z}}|\xi_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z,\widehat{\pi}_K^{(b)})| \leq \inf_{(t,z)\in\mathcal{T}\times\mathcal{Z}}\frac{1}{4}\partial_a^2\widetilde{\phi}(g_0(t,z))f_{T,Z}(t,z)\cdot(\kappa_{21}\wedge 1)\cdot\varepsilon^2$$

then for any (t,z) and $\boldsymbol{\theta}$ such that $\|\boldsymbol{\theta}-\widetilde{\boldsymbol{\theta}}(t,z)\|>\varepsilon/\log N$, we have

$$H_{N,\boldsymbol{h}}(\boldsymbol{\theta},t,z,\widehat{\pi}_K^{(b)}) > H_{N,\boldsymbol{h}}(\widetilde{\boldsymbol{\theta}}(t,z),t,z,\widehat{\pi}_K^{(b)}),$$

which implies that $\|\widehat{\boldsymbol{\theta}}^{(b)}(t,z) - \widetilde{\boldsymbol{\theta}}(t,z)\| \leq \varepsilon/\log N$, because $\widehat{\boldsymbol{\theta}}^{(b)}(t,z)$ is the minimizer of $H_{N,\boldsymbol{h}}(\boldsymbol{\theta}, t, z, \widehat{\pi}_K^{(b)})$ so that $H_{N,\boldsymbol{h}}(\widetilde{\boldsymbol{\theta}}(t,z), t, z, \widehat{\pi}_K^{(b)}) \geq H_{N,\boldsymbol{h}}(\widehat{\boldsymbol{\theta}}^{(b)}(t,z), t, z, \widehat{\pi}_K^{(b)})$. We thus have (D.12).

For (D.13), considering sufficiently large N, $\{\boldsymbol{\theta} : |\boldsymbol{\theta}| \leq (C + \varepsilon/\log N) \cdot (1,1)^\top\}$ is compact in $\mathbb{R}^2$. Then, by Lemma 4 and the result of Step I, we have $\sup_{\boldsymbol{\theta}:|\boldsymbol{\theta}|\leq(C+\varepsilon/\log N)\cdot(1,1)^\top} \sup_{t,z} |\xi_{N,\boldsymbol{h}}(\boldsymbol{\theta}, \widehat{\pi}_K^{(b)})| = o_P(1)$, which gives that for every $\varepsilon > 0$,

$$\mathbb{P}\left(\sup_{\boldsymbol{\theta}:|\boldsymbol{\theta}|\leq(C+\varepsilon/\log N)\cdot(1,1)^\top} \sup_{t,z} |\xi_{N,\boldsymbol{h}}(\boldsymbol{\theta}, t, z, \widehat{\pi}_K^{(b)})| \geq \inf_{t,z} \frac{1}{4} \cdot \partial_a^2 \widetilde{\phi}(g_0(t,z)) f_{T,Z}(t,z) \cdot (\kappa_{21} \wedge 1) \cdot \varepsilon^2\right) = o(1),$$

as $\inf_{t,z} \partial_a^2 \widetilde{\phi}(g_0(t,z)) f_{T,Z}(t,z) > 0$ by Assumption 10(v). Hence, (D.13) holds and the proof of Step II is done.

For Step III, it's can be concluded directly from the result of Step II by noting the definition of $\widetilde{\boldsymbol{\theta}}(t,z)$.

E Proof of Theorem 5

Theorem 5 is a direct consequence of Corollary 3.1 in Chernozhukov et al. (2014a). In order to apply Corollary in Chernozhukov et al. (2014a), we need to verify their Conditions H1–H4. Our Theorem 4 have already established that the original and multiplier bootstrap estimators can be approximated by local empirical processes with a kernel function and the approximation error are $o_P(\{\log(N)\}^{-1/2})$ uniformly over $(t, z) \in \mathcal{T} \times \mathcal{Z}$. Then, following Proposition 3.2 and Remark 3.2 in Chernozhukov et al. (2014b), the approximation errors are asymptotically negligible. Focusing on the local empirical process part, Conditions H1–H4 can be verified by Theorem 3.2 in Chernozhukov et al. (2014a). Specifically, Condition VC in Chernozhukov et al. (2014a) holds where, in their notation, a_n and v_n are constants, $b_n = h_{pro}^{-1/2} h_0^{-1}$, $K_n = \log(n)$, σ_n^2 is bounded and

$$\log^4(N)/N h_{pro} h_0^2 = o(N^{-c})$$

for some constant $c > 0$ as we assume.

References

Abrevaya, J., Y.-C. Hsu, and R. P. Lieli (2015). Estimating conditional average treatment effects. *Journal of Business & Economic Statistics 33*(4), 485–505.

Ai, C., O. Linton, K. Motegi, and Z. Zhang (2019). Supplemental material for 'a unified framework for efficient estimation of general treatment models'. Technical report, University of Florida.

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 (2022). Estimation and inference for the counterfactual distribution and quantile functions in continuous treatment models. *Journal of Econometrics 228*(1), 39–61.

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

Chen, X., D. Pouzo, and J. L. Powell (2019). Penalized sieve gel for weighted average derivatives of nonparametric quantile iv regressions. *Journal of Econometrics 213*(1), 30–53.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014a). Anti-concentration and honest, adaptive confidence bands. *The Annals of Statistics 42*(5), 1787–1818.

Chernozhukov, V., D. Chetverikov, and K. Kato (2014b). Gaussian approximation of suprema of empirical processes. *The Annals of Statistics 42*(4), 1564–1597.

Chernozhukov, V. and C. Hansen (2005). An iv model of quantile treatment effects. *Econometrica 73*(1), 245–261.

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.

Deville, J. and C. Särndal (1992). Calibration estimators in survey sampling. *J. Am. Statist. Ass. 87*(418), 376–382.

Fan, J., T.-C. Hu, and Y. K. Truong (1994). Robust non-parametric function estimation. *Scandinavian journal of statistics*, 433–446.

Fan, Q., Y.-C. Hsu, R. P. Lieli, and Y. Zhang (2022). Estimation of conditional average treatment effects with high-dimensional data. *Journal of Business & Economic Statistics 40*(1), 313–327.

Florens, J. P., J. J. Heckman, C. Meghir, and E. Vytlacil (2008). Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica 76*(5), 1191–1206.

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.

Giessing, A. and J. Wang (2021). Debiased inference on heterogeneous quantile treatment effects with regression rank-scores. *arXiv e-prints*, arXiv–2102.

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

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

Heckman, J. J., H. Ichimura, and P. Todd (1998). Matching as an econometric evaluation estimator. *The Review of Economic Studies 65*(2), 261–294.

Heckman, J. J. and E. Vytlacil (2005). Structural equations, treatment effects, and econometric policy evaluation. *Econometrica 73*(3), 669–738.

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.

Hjort, N. L. and D. Pollard (2011). Asymptotics for minimisers of convex processes. *arXiv preprint arXiv:1107.3806*.

Huang, W., O. Linton, and Z. Zhang (2022). A unified framework for specification tests of continuous treatment effect models. *Journal of Business & Economic Statistics 40*(4), 1817–1830.

Huang, W. and Z. Zhang (2022). Nonparametric estimation of the continuous treatment effect with measurement error. *arXiv preprint arXiv:2211.04642*.

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., R. Spady, and P. Johnson (1998). Information theoretic approaches to inference in moment condition models. *Econometrica 66*(2), 333–357.

Kennedy, E. H. (2020). Optimal doubly robust estimation of heterogeneous causal effects. *arXiv preprint arXiv:2004.14497*.

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.

Lejeune, M. G. and P. Sarda (1988). Quantile regression: a nonparametric approach. *Computational Statistics & Data Analysis 6*(3), 229–239.

Newey, W. K. (1997, July). Convergence rates and asymptotic normality for series estimators. *Journal of Econometrics 79*(1), 147–168.

Nie, X. and S. Wager (2021). Quasi-oracle estimation of heterogeneous treatment effects. *Biometrika 108*(2), 299–319.

Owen, A. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika 75*(2), 237–249.

Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory 7*(2), 186–199.

Robins, J. M., M. A. Hernán, and B. Brumback (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology 11*(5), 550–560.

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.

Wager, S. and S. Athey (2018). Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association 113*(523), 1228–1242.

Xu, R., J. Gao, T. Oka, and Y.-J. Whang (2022). Estimation of heterogeneous treatment effects using quantile regression with interactive fixed effects. *arXiv preprint arXiv:2208.03632*.

Yiu, S. and L. Su (2018). Covariate association eliminating weights: A unified weighting framework for causal effect estimation. *Biometrika 105*(3), 709–722.