NONPARAMETRIC INFERENCE ON DOSE-RESPONSE CURVES WITHOUT THE POSITIVITY CONDITION

BY YIKUN ZHANG[1,*], YEN-CHI CHEN[1,†] AND ALEXANDER GIESSING[1,‡]

[1]*Department of Statistics, University of Washington*
[*]*yikun@uw.edu*; [†]*yenchic@uw.edu*; [‡]*giessing@uw.edu*

Existing statistical methods in causal inference often rely on the assumption that every individual has some chance of receiving any treatment level regardless of its associated covariates, which is known as the positivity condition. This assumption could be violated in observational studies with continuous treatments. In this paper, we present a novel integral estimator of the causal effects with continuous treatments (*i.e.*, dose-response curves) without requiring the positivity condition. Our approach involves estimating the derivative function of the treatment effect on each observed data sample and integrating it to the treatment level of interest so as to address the bias resulting from the lack of positivity condition. The validity of our approach relies on an alternative weaker assumption that can be satisfied by additive confounding models. We provide a fast and reliable numerical recipe for computing our estimator in practice and derive its related asymptotic theory. To conduct valid inference on the dose-response curve and its derivative, we propose using the nonparametric bootstrap and establish its consistency. The practical performances of our proposed estimators are validated through simulation studies and an analysis of the effect of air pollution exposure ($\mathrm{PM}_{2.5}$) on cardiovascular mortality rates.

1. Introduction. In observational studies, the causal effect of interest does not always result from a standard binary intervention but rather comes as a consequence of a continuous treatment or exposure. Such a causal effect on the outcome variable $Y \in \mathcal{Y} \subset \mathbb{R}$ from a continuous treatment variable $T \in \mathcal{T} \subset \mathbb{R}$ is known as the (causal) dose-response curve or relationship due to its major application in studying biological effects (Waud, 1975). More precisely, a dose-response curve characterizes the average outcome if all units would have been assigned to a certain treatment level $T = t$. In practice, it is fairly common that an additional set of covariates $\boldsymbol{S} \in \mathcal{S} \subset \mathbb{R}^d$ is collected, which, to some extent, contains all the possible confounding variables that influence both the treatment T and the outcome Y. Under regularity conditions (see Section 2.2), the dose-response curve coincides with the so-called covariate-adjusted regression function $t \mapsto m(t) = \mathbb{E}\left[\mu(t, \boldsymbol{S})\right]$ with $\mu(t, \boldsymbol{s}) = \mathbb{E}\left(Y | T = t, \boldsymbol{S} = \boldsymbol{s}\right)$ and is thus identifiable from the observed data $\left\{(Y_i, T_i, \boldsymbol{S}_i)\right\}_{i=1}^n$ for any $t \in \mathcal{T}$ (Robins et al., 2000; Neugebauer and van der Laan, 2007; Díaz and van der Laan, 2013; Kennedy et al., 2017). Alternatively, the covariate-adjusted regression function $m(t)$ can be viewed as a univariate summary function of the outcome against a continuous covariate T when we average the regression function $\mu(t, \boldsymbol{s})$ over all other covariates $\boldsymbol{S}$ (Takatsu and Westling, 2022).

However, the regularity conditions for identifying and estimating the dose-response curve might not be verifiable or could even be violated in observational studies. In particular, it is commonly assumed that there is a sufficient amount of variability in the treatment assignment within each strata of the covariates, which is captured by the following positivity or overlapping condition.

MSC2020 subject classifications: Primary 62G05; secondary 62D20, 62G20.
Keywords and phrases: Causal inference, dose-response curve, positivity, kernel smoothing.

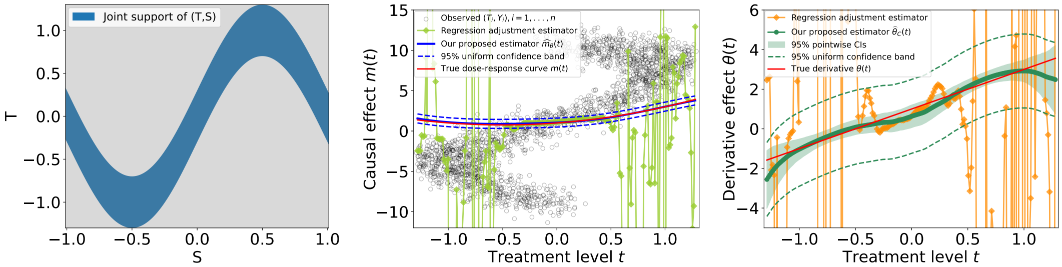

Fig 1: Simulation results under the single confounder model (1). **Left:** The support of the joint distribution of (T, S). **Middle:** The estimated dose-response curves by usual regression adjustment and our proposed estimators overlain with the true one $m(t)$. **Right:** The estimated derivatives of the dose-response curve by usual regression adjustment and our proposed estimators overlain with the true one $\theta(t)$. The middle and right panels also present the 95% confidence intervals and/or uniform confidence bands from our proposed estimators as shaded regions and dashed lines, respectively.

ASSUMPTION A0 (Positivity). The conditional density $p(t|\boldsymbol{s})$ is bounded above and away from zero almost surely for all $t \in \mathcal{T}$ and $\boldsymbol{s} \in \mathcal{S}$.

The positivity condition (A0) may be violated in observational studies for either of the following two reasons: (i) theoretically, it is impossible for some individuals with certain covariate values to receive some levels of the treatment, and (ii) practically, individuals at some levels of the treatment may not be collected in the finite data sample; see Cole and Hernán (2008); Westreich and Cole (2010); Petersen et al. (2012) for related discussions. These problems are particularly pervasive under the context of continuous treatments. For instance, air pollution levels can vary greatly across larger regions but remain relatively consistent within smaller and nearby areas. Therefore, individuals at the same location are typically only get exposed to one level of air pollution, *i.e.*, spatial confounding variables may change at a finer scale than the variation of exposure, thereby violating the positivity condition (Paciorek, 2010; Schnell and Papadogeorgou, 2020; Keller and Szpiro, 2020). As a more concrete example, consider a single confounder model

$$(1) \quad Y = T^2 + T + 1 + 10S + \epsilon, \quad T = \sin(\pi S) + E, \quad \text{and} \quad S \sim \text{Uniform}[-1, 1] \subset \mathbb{R},$$

where $E \sim \text{Uniform}[-0.3, 0.3]$ is an independent treatment variation and $\epsilon \sim \mathcal{N}(0, 1)$ is an exogenous normal noise. The marginal supports of T and S are $\mathcal{T} = [-1.3, 1.3]$ and $\mathcal{S} = [-1, 1]$ respectively, while the joint support of (T, S) only covers a thin band region of the product space $\mathcal{T} \times \mathcal{S}$; see the left panel of Fig 1. The conditional density $p(t|s)$ for any $s \in \mathcal{S}$ is 0 within the gray regions, and the positivity condition (A0) clearly fails. Without (A0), the existing approaches for estimating the dose-response curve $m(t)$ and its derivative $\theta(t) = m'(t)$ can be very unstable at some specific treatment levels; see the usual regression adjustment estimators (defined in (6) and Remark 1 below) in the middle and right panels of Fig 1 for illustrations.

In this paper, we propose a novel integral estimator that can consistently recover the entire dose-response curve $m(t)$ and a localized estimator of its derivative $\theta(t) = m'(t)$ even when the positivity condition (A0) fails to hold in some regions of the product space $\mathcal{T} \times \mathcal{S}$; see the middle and right panels of Fig 1. Our main contributions are summarized as follows:

1. Methodology: After discussing the identification conditions of $m(t)$ as a causal dose-response curve in Section 2, we introduce our integral estimator of $m(t)$ in Section 3.1, which is constructed from a localized estimator of $\theta(t)$ around the observed data and extrapolate to any treatment level $t \in \mathcal{T}$ through the fundamental theorem of calculus. Such an integral estimator can also be efficiently computed in practice via the Riemann sum approximation in Section 3.2 and be reliably inferred through nonparametric bootstrap in Section 3.3.

2. Asymptotic Theory: We establish uniform consistencies of our proposed integral estimator and localized derivative estimator under the context of kernel smoothing methods in Section 4.2. We also prove the validity of nonparametric bootstrap inference in Section 4.3.

3. Experiments: We demonstrate the finite-sample performances of our proposed estimators through simulation studies and a case study of the effect of fine particulate matter ($\mathrm{PM}_{2.5}$) on cardiovascular mortality rates in Section 5. All the code for our experiments is available at `https://github.com/zhangyk8/npDoseResponse/tree/main/Paper_Code`.

1.1. *Other Related Works.* Estimating the dose-response curve $m(t)$ is a technically difficult problem in causal inference due to the fact that $m(t)$ is not pathwise differentiable and cannot be consistently estimated in a $\sqrt{n}$ rate (Chamberlain, 1986; van der Vaart, 1991). Parametrically, Robins et al. (2000) pioneered a marginal structural model to estimate $m(t)$, whose dependence on the correct specification of a parametric model was later relaxed by van der Laan and Robins (2003); Neugebauer and van der Laan (2007) through a projection of $m(t)$ to the parametric model space. Nonparametrically, a regression adjustment approach to estimating $m(t) = \mathbb{E}\left[\mu(t, \boldsymbol{S})\right]$ under a two-stage kernel smoothing estimator was first studied by Newey (1994) and later adapted to the estimation of dose-response curves by Flores (2007). However, the consistency and asymptotic normality of their kernel-based estimators rely on an undersmoothing bandwidth, which is asymptotically smaller than the one that attains the optimal asymptotic bias-variance trade-off and is difficult to select in practice; see Section 5.7 in Wasserman (2006). On the contrary, our proposed integral estimator of $m(t)$ under the context of kernel smoothing is based on the derivative estimation and require neither explicit undersmoothing nor bias correction (Calonico et al., 2018; Takatsu and Westling, 2022), permitting the use of any bandwidth selection method in nonparametric regression.

There are many existing works about the estimation of average derivative effects in the literature; see Härdle and Stoker (1989); Powell et al. (1989); Newey and Stoker (1993); Cattaneo et al. (2010); Hirshberg and Wager (2020) and references therein. Under some regularity conditions, average derivative effects are identical to the so-called incremental treatment effects, which are closely related to the derivatives $\theta(t) = m'(t)$ of dose-response curves; see Proposition 1 in Rothenhäusler and Yu (2019) and Section 6.1 in Hines et al. (2023).

All the aforementioned methods for estimating $m(t)$ and $\theta(t)$ assume the positivity condition (A0). Under the context of binary treatments, there are previous researches studying the dependence of common estimators on (A0) from the perspectives of convergence rates (Khan and Tamer, 2010; D'Amour et al., 2021) and empirical performances (Busso et al., 2014; Léger et al., 2022). To address the violation of positivity, some approaches considered trimming the extreme values of (estimated) propensity scores (*i.e.*, discrete versions of the conditional density $p(t|\boldsymbol{s})$; Dehejia and Wahba 1999; Crump et al. 2009; Yang and Ding 2018), while others proposed robust estimation methods against the violation (Rothe, 2017; Ma and Wang, 2020). In addition, Kennedy (2019) considered switching the causal estimand to the incremental treatment effect for discrete treatments, which can be nonparametrically estimated without positivity. Rothenhäusler and Yu (2019) later studied the same estimand for continuous treatments. To the best of our knowledge, none of existing works address the nonparametric inference on $m(t)$ and $\theta(t)$ without assuming (A0).

1.2. *Notations.* Throughout the paper, we consider a real-valued outcome variable $Y \in \mathcal{Y} \subset \mathbb{R}$, univariate continuous treatment $T \in \mathcal{T} \subset \mathbb{R}$, and a vector of covariates $\boldsymbol{S} =$

$(S_1, ..., S_d) \in \mathcal{S} \subset \mathbb{R}^d$ with a fixed dimension d. We write $Y \perp\!\!\!\perp X$ when the random variables Y, X are independent. The data sample consists of independent and identically distributed (i.i.d.) observations $\boldsymbol{U}_i = (Y_i, T_i, \boldsymbol{S}_i), i = 1, ..., n$ with the common distribution P and Lebesgue density $p(y, t, \boldsymbol{s}) = p(y|t, \boldsymbol{s}) \cdot p(t|\boldsymbol{s}) \cdot p_S(\boldsymbol{s})$. Here, $p_T(t)$ and $p_S(\boldsymbol{s})$ are the marginal densities of T and $\boldsymbol{S}$, respectively, and $p(t|\boldsymbol{s}) = \frac{\partial}{\partial t}\mathrm{P}\left(T \leq t | \boldsymbol{S} = \boldsymbol{s}\right)$ is the conditional density of T given covariates $\boldsymbol{S} = \boldsymbol{s}$. We also denote the joint density of $(T, \boldsymbol{S})$ by $p(t, \boldsymbol{s}) = p(t|\boldsymbol{s}) \cdot p_S(\boldsymbol{s}) = p(\boldsymbol{s}|t) \cdot p_T(t)$. Let $\mathbb{P}_n$ be the empirical distribution of the observed data so that $\mathbb{P}_n g = \int g(\boldsymbol{u})\, d\mathbb{P}_n(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^n g(\boldsymbol{U}_i)$ for some measurable function g. At the population level, we denote $\mathrm{P}g = \mathbb{E}\left[g(\boldsymbol{U})\right] = \int g(\boldsymbol{u})\, d\mathrm{P}(\boldsymbol{u})$ for P-integrable function g. In addition, we define the empirical process evaluated at a P-integrable function g as $\mathbb{G}_n(g) = \sqrt{n}\left(\mathbb{P}_n - \mathrm{P}\right) g = \frac{1}{\sqrt{n}}\sum_{i=1}^n \left[g(\boldsymbol{U}_i) - \mathbb{E}\left(g(\boldsymbol{U}_i)\right)\right]$. Finally, we use $\mathbb{1}_A$ to denote the indicator function of a set A.

We use the big-O notation $h_n = O(g_n)$ if $|h_n|$ is upper bounded by a positive constant multiple of $g_n > 0$ when n is sufficiently large. In contrast, $h_n = o(g_n)$ when $\lim_{n\to\infty} \frac{|h_n|}{g_n} = 0$. For random variables, the notation $o_P(1)$ is short for a sequence of random variables converging to zero in probability, while the expression $O_P(1)$ denotes the sequence that is bounded in probability. We also use the notation $a_n \lesssim b_n$ or $b_n \gtrsim a_n$ when there exists an absolute constant $A > 0$ such that $a_n \leq A b_n$ when n is large. If $a_n \gtrsim b_n$ and $a_n \lesssim b_n$, then a_n, b_n are asymptotically equal and it is denoted by $a_n \asymp b_n$.

2. Model Setups and Identification Conditions. We assume that $\boldsymbol{U}_i = (Y_i, T_i, \boldsymbol{S}_i), i = 1, ..., n$ are i.i.d. observations generated from the following model:

$$Y = \mu(T, \boldsymbol{S}) + \epsilon \quad \text{and} \quad T = f(\boldsymbol{S}) + E, \tag{2}$$

where $E \in \mathbb{R}$ is the treatment variation with $\mathbb{E}(E) = 0$, $E \perp\!\!\!\perp \boldsymbol{S}$, and $\epsilon \in \mathbb{R}$ is an exogenous noise variable with $\epsilon \perp\!\!\!\perp E, \epsilon \perp\!\!\!\perp \boldsymbol{S}$ and $\mathbb{E}(\epsilon) = 0, \mathrm{Var}(\epsilon) = \sigma^2 > 0, \mathbb{E}(\epsilon^4) < \infty$. The function f determines how the covariates (or confounding variables) $\boldsymbol{S}$ influence the treatment T. Under model (2), the covariate-adjusted regression function (or the dose-response curve under identification conditions in Section 2.2) is given by

$$t \mapsto m(t) = \mathbb{E}\left[\mathbb{E}(Y|T = t, \boldsymbol{S})\right] = \mathbb{E}\left[\mu(t, \boldsymbol{S})\right],$$

and its derivative is written as $\theta(t) = m'(t) = \frac{d}{dt}\mathbb{E}\left[\mu(t, \boldsymbol{S})\right]$. Notice that the derivative function $\theta(t)$ can be viewed as the continuous version of the average treatment effect in causal inference. In the presence of confounding variables $\boldsymbol{S}$, $m(t)$ cannot be identified by regressing Y only with respect to T, because $m(t) = \mathbb{E}\left[\mu(t, \boldsymbol{S})\right] \neq \mathbb{E}\left(Y|T = t\right) = \mathbb{E}\left[\mu(T, \boldsymbol{S})|T = t\right]$.

2.1. *Motivating Example: Additive Confounding Model.* One important exemplification of model (2) that we will frequently refer to in this paper is the following additive confounding model:

$$Y = \bar{m}(T) + \eta(\boldsymbol{S}) + \epsilon \quad \text{and} \quad T = f(\boldsymbol{S}) + E, \tag{3}$$

where $\bar{m}(t)$ is the primary treatment effect of interest and $\eta(\boldsymbol{s})$ is the random effect with $\mathbb{E}\left[\eta(\boldsymbol{S})\right] = 0$. Such an additive form is a common working model in spatial confounding problems (Paciorek, 2010; Schnell and Papadogeorgou, 2020) and also known as the geoadditive structural equation model (Kammann and Wand, 2003; Thaden and Kneib, 2018; Wiecha and Reich, 2024), where $\boldsymbol{S} \in \mathbb{R}^d$ are the spatial locations (usually with $d = 2$) or other spatially correlated covariates that affect both the treatment T and the outcome Y. We summarize key properties of the additive confounding model (3) in the following proposition.

PROPOSITION 1 (Properties of the additive confounding model (3)). *Let* $\theta_M(t) = \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,\boldsymbol{S})\right]$ *and* $\theta_C(t) = \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,\boldsymbol{S})\middle|T=t\right]$, *where* $\mu(t,\boldsymbol{s}) = \mathbb{E}\left[Y|T=t,\boldsymbol{S}=\boldsymbol{s}\right]$. *Under the additive confounding model* (3) *with* $\mathbb{E}\left[\eta(\boldsymbol{S})\right]=0$, *the following results hold for all* $t\in\mathcal{T}$:

(a) $m(t)=\bar{m}(t)$.
(b) $m(t)\neq\mathbb{E}(Y|T=t)=\bar{m}(t)+\mathbb{E}\left[\eta(\boldsymbol{S})|T=t\right]$ *when* $\mathbb{E}\left[\eta(\boldsymbol{S})|T=t\right]\neq 0$.
(c) $\theta(t)=\theta_M(t)=\theta_C(t)$.
(d) $\theta(t)\neq\frac{d}{dt}\mathbb{E}\left[\mu(t,\boldsymbol{S})|T=t\right]=\theta(t)+\frac{d}{dt}\mathbb{E}\left[\eta(\boldsymbol{S})|T=t\right]$ *when* $\frac{d}{dt}\mathbb{E}\left[\eta(\boldsymbol{S})|T=t\right]\neq 0$.
(e) $\mathbb{E}\left[\mu(T,\boldsymbol{S})\right]=\mathbb{E}\left[m(T)\right]$ *even when* $\mathbb{E}\left[\eta(\boldsymbol{S})\right]\neq 0$.

The above results hold even if the treatment variation $E=0$ *almost surely.*

The proof of Proposition 1 is in Section B.1. Note that in Proposition 1(e), the first expectation is taken with respect to the joint distribution of $(T,\boldsymbol{S})$ while the second expectation is taken with respect to the marginal distribution of T.

2.2. *Identification Conditions.* While $m(t)$ can be defined through model (2), we need some additional assumptions in order to express $m(t)$ as a causal effect of T on Y and identify it from the observed data $\{(Y_i,T_i,\boldsymbol{S}_i)\}_{i=1}^n$. Following the potential outcome framework (Rubin, 1974), we let $Y(t,\boldsymbol{s})$ denote the potential outcome that would have been observed under treatment level $T=t$ and covariate vector $\boldsymbol{S}=\boldsymbol{s}$. We summarize the required assumptions as follows.

ASSUMPTION A1 (Identification conditions for $m(t)$).

(a) (Treatment effect and consistency) $Y(t,\boldsymbol{s})=Y(t)=Y$ for any $t\in\mathcal{T}$ and $\boldsymbol{s}\in\mathcal{S}$.
(b) (Ignorability or unconfoundingness) $Y(t)\perp\!\!\!\perp T\,\big|\,\boldsymbol{S}$ for all $t\in\mathcal{T}$.
(c) (Treatment variation) The treatment variation E has nonzero variance, *i.e.*, $\mathrm{Var}(E)>0$.

Assumption A1(a) is a continuous version of the stable unit treatment value assumption (SUTVA; Page 19 of Cox 1958 and Rubin 1980). The first equality ensures that the potential outcome $Y(t)$ under treatment level $T=t$ is well-defined whatever value the covariate vector $\boldsymbol{S}$ takes, while the second equality guarantees that the treatment level of any subject does not affect the potential outcomes of others (*i.e.*, no interference) and there are no different versions of treatments. Given that both the treatment T and the outcome Y are continuous, the essential consistency condition that we need is the identity of conditional distributions $\mathrm{P}\left(Y(T)\leq y|T=t\right)=\mathrm{P}\left(Y(t)\leq|T=t\right)$ for any $t\in\mathcal{T}$ and $y\in\mathcal{Y}$; see Gill and Robins (2001) for details. The ignorability condition (Assumption A1(b)) was first generalized to continuous treatments by Hirano and Imbens (2004), stating that the potential outcome variable is independent of the treatment level within any specific strata of covariates. In the context of spatial confounding, the ignorability condition holds when the spatial locations are taken into account (Gilbert et al., 2023). It also implies that the mean potential outcome under $T=t$ remains the same across all treatment levels when we condition on $\boldsymbol{S}=\boldsymbol{s}$. Finally, the treatment variation condition (Assumption A1(c)) is crucial for identifying the conditional mean outcome (or regression) function $\mu(t,\boldsymbol{s})$ on a non-degenerate region of $\mathcal{T}\times\mathcal{S}$. When $\mathrm{Var}(E)=0$, $\mu(t,\boldsymbol{s})$ can only be identified on the lower dimensional surface $\{(t,\boldsymbol{s})\in\mathcal{T}\times\mathcal{S}:t=f(\boldsymbol{s})\}$, which is also the support of $(T,\boldsymbol{S})$. We derive nonparametric bounds on $m(t)=\mathbb{E}\left[\mu(t,\boldsymbol{S})\right]$ and its derivative $\theta(t)=\frac{d}{dt}\mathbb{E}\left[\mu(t,\boldsymbol{S})\right]$ in Section A. We also demonstrate how $\mathrm{Var}(E)=0$ can lead to ambiguous definitions of the associated dose-response curves in Example 1 below.

EXAMPLE 1 (Necessity of $\mathrm{Var}(E) > 0$). Suppose that $\mathrm{Var}(E) = 0$ and $T = f(\boldsymbol{S}) + E = S_1$ almost surely, where S_1 is the first component of $\boldsymbol{S} \in \mathcal{S} \subset \mathbb{R}^d$. We further assume that $\mathbb{E}(S_1) = 0$. Now, consider two equivalent conditional mean outcome functions

$$\mu_1(T, \boldsymbol{S}) \equiv T + 2S_1 \quad \text{and} \quad \mu_2(T, \boldsymbol{S}) \equiv 2T + S_1,$$

both of which are equal to $3S_1$ and agree on the support $\{(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S} : t = f(\boldsymbol{s}) = s_1\}$. However, these two conditional mean outcome functions lead to two distinct treatment effects:

$$m_1(t) = \mathbb{E}[\mu_1(t, \boldsymbol{S})] = t \quad \text{and} \quad m_2(t) = \mathbb{E}[\mu_2(t, \boldsymbol{S})] = 2t,$$

whose derivatives are different as well.

Under Assumption A1, the dose-response curve $t \mapsto \mathbb{E}[Y(t)]$ is identical to the standard covariate-adjusted regression function $m(t) = \mathbb{E}[\mu(t, \boldsymbol{S})] = \int_{\mathcal{S}} \mu(t, \boldsymbol{s}) \cdot p_S(\boldsymbol{s})\, d\boldsymbol{s}$ and can thus be estimated from the observed data $\{(Y_i, T_i, \boldsymbol{S}_i)\}_{i=1}^n$. To estimate the derivative $\theta(t) = m'(t) = \frac{d}{dt}\mathbb{E}[\mu(t, \boldsymbol{S})]$ of the dose-response curve from the observed data sample, we impose an additional assumption on the conditional mean outcome function $\mu(t, \boldsymbol{s}) = \mathbb{E}(Y|T = t, \boldsymbol{S} = \boldsymbol{s})$ that enables us to bypass the positivity condition (A0).

ASSUMPTION A2 (Identification condition for $\theta(t)$). The function $\mu(t, \boldsymbol{s})$ is continuously differentiable with respect to t for any $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$ and the following two equalities hold true:

$$\theta(t) = \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t, \boldsymbol{S})\right] = \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t, \boldsymbol{S}) \Big| T = t\right] \tag{4}$$

and

$$\mathbb{E}[\mu(T, \boldsymbol{S})] = \mathbb{E}[m(T)]. \tag{5}$$

The first equality in (4) is a mild condition and can be satisfied under various settings. In particular, it holds when $\left|\frac{\partial}{\partial t}\mu(t, \boldsymbol{S})\right|$ is bounded by an integrable function $\bar{\mu}(\boldsymbol{S})$ with respect to the distribution of $\boldsymbol{S}$; see Theorem 1.1 and Example 1.8 in Shao (2003). The second equality in (4) as well as the equation (5) of Assumption A2 are stricter but still valid under the additive confounding model in Section 2.1; recall Proposition 1(c,e). Under Assumption A2, we can express $\theta(t)$ in three different but also equivalent ways

$$\theta(t) = \underbrace{\frac{d}{dt}\mathbb{E}[\mu(t, \boldsymbol{S})]}_{m'(t)} = \underbrace{\mathbb{E}\left[\frac{\partial}{\partial t}\mu(t, \boldsymbol{S})\right]}_{\theta_M(t)} = \underbrace{\mathbb{E}\left[\frac{\partial}{\partial t}\mu(t, \boldsymbol{S}) \Big| T = t\right]}_{\theta_C(t)}.$$

Estimating the derivative effect $\theta(t)$ through the form of $\theta_C(t)$ is our key technique to bypass the particularly strong positivity condition (A0) under the continuous treatment setting. After $\theta(t)$ is consistently recovered, we refer the estimation back to the dose-response curve $m(t)$ by (5) and the fundamental theorem of calculus; see Section 3.1 for details.

3. Nonparametric Inference Without the Positivity Condition. Under the positivity condition (A0), together with Assumption A1, the dose-response curve $m(t) = \mathbb{E}[Y(t)]$ can be identified through the covariate-adjusted regression function $\mathbb{E}[\mu(t, \boldsymbol{S})]$ from the observed data $\{(Y_i, T_i, \boldsymbol{S}_i)\}_{i=1}^n$, suggesting the following regression adjustment (RA) or G-computation (Robins, 1986; Gill and Robins, 2001) estimator as:

$$\widehat{m}_{RA}(t) = \frac{1}{n}\sum_{i=1}^n \widehat{\mu}(t, \boldsymbol{S}_i), \tag{6}$$

where $\widehat{\mu}(t,\boldsymbol{s})$ is any consistent estimator of the conditional mean outcome function $\mu(t,\boldsymbol{s})$. However, when the positivity condition (A0) does not hold for some region in $\mathcal{T}\times\mathcal{S}$, the above estimator (6) will be unstable and even inconsistent. This is because without (A0), the joint density $p(t,\boldsymbol{S}_i)=p(t|\boldsymbol{S}_i)\cdot p_S(\boldsymbol{S}_i)$ can be close to 0 for some $i=1,...,n$, and $\mu(t,\boldsymbol{s})$ cannot be consistently estimated for those query points $(t,\boldsymbol{S}_i), i=1,...,n$. Other existing methods for estimating $m(t)$, such as the inverse probability of treatment weighted (IPTW) or its augmented variants, also relies on the validity of (A0) for their consistency and empirical behaviors (Díaz and van der Laan, 2013; Kennedy et al., 2017; Huber et al., 2020; Colangelo and Lee, 2020). The same issue incurred by the failure of (A0) also applies to the estimation of $\theta(t)=m'(t)$.

In this section, we introduce a novel integral estimator to resolve the inconsistency issues of existing estimators in recovering the dose-response curve $m(t)$ and its derivative $\theta(t)$ without the positivity condition (A0). On one hand, the estimation of $\theta(t)$ is based on the conditional expectation $\theta_C(t)\equiv\mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,\boldsymbol{S})\Big|T=t\right]$ in (4), which only relies on the consistent estimation of $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ in the high density region of $p(\boldsymbol{s}|t)$. On the other hand, our proposed estimator of $m(t)$ generalizes the regression adjustment estimator (6) but remains consistent and numerically stable even when the conditional density $p(t|\boldsymbol{s})$ is zero for some values of $\boldsymbol{s}\in\mathcal{S}$ thanks to its integral formulation. We also provide a fast algorithm for computing our proposed estimator of $m(t)$ in practice and delineate the bootstrap inference procedures for both estimators of $\theta(t)$ and $m(t)$.

3.1. *Proposed Integral Estimator of* $m(t)$. Given the observed data $\{(Y_i,T_i,\boldsymbol{S}_i)\}_{i=1}^n$, our proposal for removing the reliance of the positivity condition (A0) from (6) is based on three critical insights.

- **Insight 1: Consistent estimation of** $\mu(t,\boldsymbol{s})$ **and** $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ **at each** $(T_i,\boldsymbol{S}_i)$**.** Given that the observed data $(T_i,\boldsymbol{S}_i), i=1,...,n$ generally appear in a high density region of $p(t,\boldsymbol{s})$, the conditional mean outcome function $\mu(t,\boldsymbol{s})$ can be well-estimated at each observation $(T_i,\boldsymbol{S}_i)$. As a result, one can expect that the partial derivative $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ can be consistently estimated at each observation $(T_i,\boldsymbol{S}_i)$ as well.
- **Insight 2: Consistent estimation of** $\theta(t)$ **from a localized form** $\theta_C(t)$**.** The equations (4) in Assumption A2 lead to two possible approaches to estimating the derivative effect $\theta(t)=\frac{d}{dt}\mathbb{E}\left[\mu(t,\boldsymbol{S})\right]$. One approach is via $\theta_M(t)=\mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,\boldsymbol{S})\right]$, while the alternative is based on $\theta_C(t)=\mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,\boldsymbol{S})\Big|T=t\right]$. Estimating $\theta(t)$ by $\theta_C(t)$ is preferable because $\theta_M(t)$ relies on a consistent estimator of $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ at each pair $(t,\boldsymbol{S}_i)$, which is not possible when the positivity condition (A0) fails to hold at some $(t,\boldsymbol{S}_i)$ for $i=1,...,n$. In contrast, the expression $\theta_C(t)$ only requires the estimator of $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ to be accurate at the covariate vector $\boldsymbol{s}$ with a high conditional density value $p(\boldsymbol{s}|t)$.
- **Insight 3: Integral relation between** $\theta(t)$ **and** $m(t)$**.** For any $t\in\mathcal{T}$, the fundamental theorem of calculus reveals that
$$m(t)=m(T)+\int_{\tilde{t}=T}^{\tilde{t}=t}m'(\tilde{t})\,d\tilde{t}=m(T)+\int_{\tilde{t}=T}^{\tilde{t}=t}\theta(\tilde{t})\,d\tilde{t}.$$
Under Assumption A2, we take the expectation on both sides of the above equality to obtain that
$$\begin{aligned}m(t)&=\mathbb{E}\left[m(T)+\int_{\tilde{t}=T}^{\tilde{t}=t}\theta(\tilde{t})\,d\tilde{t}\right]=\mathbb{E}\left[\mu(T,\boldsymbol{S})\right]+\mathbb{E}\left[\int_{\tilde{t}=T}^{\tilde{t}=t}\theta_C(\tilde{t})\,d\tilde{t}\right]\\&=\mathbb{E}(Y)+\mathbb{E}\left\{\int_{\tilde{t}=T}^{\tilde{t}=t}\mathbb{E}\left[\frac{\partial}{\partial t}\mu(\tilde{t},\boldsymbol{S})\Big|T=\tilde{t}\right]d\tilde{t}\right\}.\end{aligned}\tag{7}$$

This expression suggests that as long as we have consistent estimators of $\mu(t,\boldsymbol{s})$ at $(T,\boldsymbol{S})$ and $\theta(t)$ near T, we can then use the integration to extrapolate the estimation to $\mu(t,\boldsymbol{S})$ and $m(t)$ for any $t\in\mathcal{T}$ even when $p(t,\boldsymbol{S})=0$.

According to the integral relation (7), we propose an *integral estimator* of the dose-response curve $m(t)$ as:

$$\widehat{m}_\theta(t)=\frac{1}{n}\sum_{i=1}^{n}\left[Y_i+\int_{\widetilde{t}=T_i}^{\widetilde{t}=t}\widehat{\theta}_C(\widetilde{t})\,d\widetilde{t}\right], \tag{8}$$

where $\widehat{\theta}_C(t)$ is a consistent estimator of $\theta_C(t)=\mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,\boldsymbol{S})\Big|T=t\right]=\int\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})\,d\mathrm{P}(\boldsymbol{s}|t)$. To construct an estimator of $\theta_C(t)$, we need to estimate two nuisance functions: (i) the partial derivative $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ of the conditional mean outcome function $\mu(t,\boldsymbol{s})$ and (ii) the conditional cumulative distribution function (CDF) $P(\boldsymbol{s}|t)$. In this paper, we leverage the following two kernel smoothing methods for estimating these two nuisance functions and leave other possibilities as a future direction.

3.1.1. *Local Polynomial Regression Estimator of* $\frac{\partial}{\partial t}\mu(t,\mathrm{s})$. We consider estimating $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ by the local polynomial regression (Fan and Gijbels, 1996) because of its robustness around the boundary of support (also known as the automatic kernel carpentry in Hastie and Loader 1993).

Let $K_T:\mathbb{R}\to[0,\infty), K_S:\mathbb{R}^d\to[0,\infty)$ be two symmetric kernel functions and $h,b>0$ be their corresponding smoothing bandwidth parameters. Some commonly used univariate kernel functions include the Epanechnikov kernel $K(u)=\frac{3}{4}\left(1-u^2\right)\cdot\mathbb{1}_{\{|u|\le 1\}}$ and Gaussian kernel $K(u)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)$. For the multivariate kernel function, one often resorts to the product kernel technique as $K_S(\boldsymbol{u})=\prod_{i=1}^{d}K(u_i)$ for $\boldsymbol{u}\in\mathbb{R}^d$. To estimate $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ from the observed data $\{(Y_i,T_i,\boldsymbol{S}_i)\}_{i=1}^n$, we fit a partial local polynomial regression of order q ($q\ge 1$) with monomials of $\{(T_i-t)\}_{i=1}^n$ as the polynomial basis in treatment variable T and the local linear function in covariate vector $\boldsymbol{S}$ (Ruppert and Wand, 1994). Specifically, we let $\boldsymbol{X}_i(t,\boldsymbol{s})=\left(1,(T_i-t),...,(T_i-t)^q,(S_{i,1}-s_1),...,(S_{i,d}-s_d)\right)^T\in\mathbb{R}^{q+1+d}$ for $i=1,...,n$ and consider

$$\begin{aligned}
&\left(\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}),\widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\right)^T\\
&=\operatorname*{arg\,min}_{(\boldsymbol{\beta},\boldsymbol{\alpha})^T\in\mathbb{R}^{q+1}\times\mathbb{R}^d}\sum_{i=1}^{n}\left[Y_i-\sum_{j=0}^{q}\beta_j(T_i-t)^q-\sum_{\ell=1}^{d}\alpha_\ell(S_{i,\ell}-s_\ell)\right]^2K_T\left(\frac{T_i-t}{h}\right)K_S\left(\frac{\boldsymbol{S}_i-\boldsymbol{s}}{b}\right)\\
&=\operatorname*{arg\,min}_{(\boldsymbol{\beta},\boldsymbol{\alpha})^T\in\mathbb{R}^{q+1}\times\mathbb{R}^d}\sum_{i=1}^{n}\left(Y_i-\boldsymbol{X}_i(t,\boldsymbol{s})^T\begin{pmatrix}\boldsymbol{\beta}\\ \boldsymbol{\alpha}\end{pmatrix}\right)^2K_T\left(\frac{T_i-t}{h}\right)K_S\left(\frac{\boldsymbol{S}_i-\boldsymbol{s}}{b}\right).
\end{aligned}\tag{9}$$

For simplicity, we use the same bandwidth parameter b for each coordinate in K_S here. One can generalize the above method and related theoretical results to a general bandwidth matrix in K_S with little effort. Let $\boldsymbol{X}(t,\boldsymbol{s})\in\mathbb{R}^{n\times(q+1+d)}$ be a matrix with the j-th row as $\boldsymbol{X}_j(t,\boldsymbol{s})^T$ and $\boldsymbol{Y}=(Y_1,...,Y_n)^T\in\mathbb{R}^n$. We also define a diagonal weight matrix as:

$$\boldsymbol{W}(t,\boldsymbol{s})=\mathtt{Diag}\left(K_T\left(\frac{T_1-t}{h}\right)K_S\left(\frac{\boldsymbol{S}_1-\boldsymbol{s}}{b}\right),...,K_T\left(\frac{T_n-t}{h}\right)K_S\left(\frac{\boldsymbol{S}_n-\boldsymbol{s}}{b}\right)\right)\in\mathbb{R}^{n\times n}.$$

Then, (9) has a closed-form solution from a weighted least square problem as:

$$\left(\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}),\widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\right)^T=\left[\boldsymbol{X}^T(t,\boldsymbol{s})\boldsymbol{W}(t,\boldsymbol{s})\boldsymbol{X}(t,\boldsymbol{s})\right]^{-1}\boldsymbol{X}(t,\boldsymbol{s})^T\boldsymbol{W}(t,\boldsymbol{s})\boldsymbol{Y}. \tag{10}$$

Finally, the second component $\widehat{\beta}_2(t,\boldsymbol{s})$ of the fitted coefficient $\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s})$ provides a natural estimator of $\beta_2(t,\boldsymbol{s}) \equiv \frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$. In practice, we recommend to choose q to be an even number when estimating the first-order (partial) derivative $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ via (9), because there is an increment to the asymptotic variance of $\widehat{\beta}_2(t,\boldsymbol{s})$ when q changes from an even number to the consecutive odd number. Additionally, fitting (partial) local polynomial regressions of higher orders often give rise to a possible reduction of bias but also a substantial increase of the variability; see Chapter 3.3 in Fan and Gijbels (1996). Therefore, we mainly focus on the (partial) local quadratic regression $q=2$ when constructing our derivative estimator $\widehat{\beta}_2(t,\boldsymbol{s})$ in the subsequent analysis.

3.1.2. *Kernel-Based Conditional CDF Estimator of* $\mathrm{P}(\mathbf{s}|t)$. We consider estimating $\mathrm{P}(\boldsymbol{s}|t)$ through a Nadaraya-Watson conditional CDF estimator (Hall et al., 1999) defined as:

$$\widehat{P}_{\hbar}(\boldsymbol{s}|t) = \frac{\sum_{i=1}^n \mathbb{1}_{\{\boldsymbol{S}_i \leq \boldsymbol{s}\}} \cdot \bar{K}_T\left(\frac{T_i - t}{\hbar}\right)}{\sum_{j=1}^n \bar{K}_T\left(\frac{T_j-t}{\hbar}\right)}, \tag{11}$$

where $\bar{K}_T : \mathbb{R} \to [0,\infty)$ is again a kernel function and $\hbar > 0$ is the smoothing bandwidth parameter that needs not be the same as the bandwidth parameter $h > 0$ for estimating $\beta_2(t,\boldsymbol{s}) = \frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ by the local polynomial regression. Practically, there are several strategies for choosing the bandwidth $\hbar > 0$ in (11) as described by Bashtannyk and Hyndman (2001); Holmes et al. (2012).

Combining the partial derivative estimator $\widehat{\beta}_2(t,\boldsymbol{s})$ in (10) with the conditional CDF estimator $\widehat{P}_{\hbar}(\boldsymbol{s}|t)$ in (11), we deduce the final localized estimator of $\theta_C(t) = \int \frac{\partial}{\partial t}\mu(t,\boldsymbol{s})\, d\mathrm{P}(\boldsymbol{s}|t)$ as:

$$\widehat{\theta}_C(t) = \int \widehat{\beta}_2(t,\boldsymbol{s})\, d\widehat{P}_{\hbar}(\boldsymbol{s}|t) = \frac{\sum_{i=1}^n \widehat{\beta}_2(t,\boldsymbol{S}_i)\cdot \bar{K}_T\left(\frac{T_i-t}{\hbar}\right)}{\sum_{j=1}^n \bar{K}_T\left(\frac{T_j-t}{\hbar}\right)}. \tag{12}$$

In essence, $\widehat{\theta}_C(t)$ is a regression adjustment estimator with two nuisance functions: $\beta_2(t,\boldsymbol{s}) = \frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ and $P(\boldsymbol{s}|t)$. Here, we have showcased how to use the (partial) local polynomial regression (9) to estimate $\beta_2(t,\boldsymbol{s})$ and Nadaraya-Watson conditional CDF estimator (11) to estimate $P(\boldsymbol{s}|t)$.

REMARK 1 (Regression adjustment estimator of $\theta(t)$). Under the additive confounding model (3), one can directly estimate $\theta(t)$ via $\widehat{\beta}_2(t,\boldsymbol{s})$ or the regression adjustment estimator $\widehat{\theta}_{RA}(t) = \frac{1}{n}\sum_{i=1}^n \widehat{\beta}_2(t,\boldsymbol{S}_i)$, because by Proposition 1,

$$\theta(t) = \frac{d}{dt}\mathbb{E}\left[\mu(t,\boldsymbol{S})\right] = \frac{d}{dt}\mathbb{E}\left[\bar{m}(t) + \eta(\boldsymbol{S})\right] = \bar{m}'(t) = \frac{\partial}{\partial t}\mu(t,\boldsymbol{s}).$$

The estimator $\widehat{\beta}_2(t,\boldsymbol{s})$ is suboptimal, because it only uses the value from a single location. Furthermore, $\widehat{\theta}_{RA}(t)$ will not be a stable and consistent estimator of $\theta(t)$ when the positivity condition fails to hold at those $(t,\boldsymbol{S}_i) \in \mathcal{T}\times\mathcal{S}$; see Fig 1 for an illustration. In contrast, what we propose in (12) relies on the conditional CDF estimator and remains valid and consistent even without the positivity condition (A0).

REMARK 2 (Linear smoother). It is worth noting that our integral estimator (8) under the kernel-based estimator (12) is that it is a linear smoother. Let $\boldsymbol{e}_2 = (0,1,0,...,0)^T \in \mathbb{R}^{q+1+d}$.

Then, with (10) and $\widehat{\beta}_2(t,\boldsymbol{s}) = \boldsymbol{e}_2^T\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s})$, we know that $\widehat{\theta}(t)$ in (12) can also be written as:

$$\widehat{\theta}_C(t) = \left\{\int \boldsymbol{e}_2^T \left[\boldsymbol{X}^T(t,\boldsymbol{s})\boldsymbol{W}(t,\boldsymbol{s})\boldsymbol{X}(t,\boldsymbol{s})\right]^{-1} \boldsymbol{X}(t,\boldsymbol{s})^T\boldsymbol{W}(t,\boldsymbol{s})\, d\widehat{P}_{\hbar}(\boldsymbol{s}|t)\right\}\boldsymbol{Y}.$$

It implies that

$$\begin{aligned}
&\widehat{m}_\theta(t)\\
&= \frac{1}{n}\sum_{i=1}^n Y_i + \left\{\frac{1}{n}\sum_{i=1}^n \int_{\widetilde{t}=T_i}^{\widetilde{t}=t}\int \boldsymbol{e}_2^T \left[\boldsymbol{X}^T(t,\boldsymbol{s})\boldsymbol{W}(t,\boldsymbol{s})\boldsymbol{X}(t,\boldsymbol{s})\right]^{-1} \boldsymbol{X}(t,\boldsymbol{s})^T\boldsymbol{W}(t,\boldsymbol{s})\, d\widehat{P}_{\hbar}(\boldsymbol{s}|t)\, d\widetilde{t}\right\}\boldsymbol{Y}\\
&\equiv \sum_{i=1}^n l_i(t)Y_i.
\end{aligned}$$

As a consequence, we can also utilize the theory of linear smoothers to derive its effective degrees of freedom and fine-tune the smoothing bandwidth parameters (Buja et al., 1989; Wasserman, 2006).

3.2. *A Fast Computing Algorithm for the Proposed Integral Estimator.* Our proposed estimator $\widehat{m}_\theta(t) = \frac{1}{n}\sum_{i=1}^n\left[Y_i + \int_{T_i}^t \widehat{\theta}_C(\widetilde{t})\, d\widetilde{t}\right]$ of the dose-response curve $m(t)$ involves an integral that may be analytically difficult to compute in practice. Here, we propose a fast computing algorithm that can numerically approximate $\widehat{m}_\theta(t)$ with an error of the order at most $O_P\left(\frac{1}{n}\right)$. The key idea is to approximate the integral via a Riemann sum, evaluate $\widehat{m}_\theta(t)$ only at the data sample $T_1,...,T_n$, and then use the linear interpolation to obtain the value at any arbitrary $t \in \mathcal{T}$.

Let $T_{(1)} \leq \cdots \leq T_{(n)}$ be the order statistics of $T_1,...,T_n$ and $\Delta_j = T_{(j+1)} - T_{(j)}$ for $j = 1,...,n-1$ be their consecutive differences. The integral estimator $\widehat{m}_\theta(t)$ in (8) can be rewritten as:

$$\widehat{m}_\theta(t) = \frac{1}{n}\sum_{i=1}^n Y_i + \frac{1}{n}\sum_{i=1}^n \int_{\widetilde{t}=T_{(i)}}^{\widetilde{t}=t} \widehat{\theta}_C(\widetilde{t})\, d\widetilde{t}.$$

To compute $\widehat{m}_\theta(T_{(j)})$ evaluated at the j-th order statistic, we consider approximating the second integral term $\frac{1}{n}\sum_{i=1}^n \int_{T_{(i)}}^{T_{(j)}} \widehat{\theta}_C(\widetilde{t})\, d\widetilde{t}$ as follows.

- When $i < j$, we have the following Riemann sum approximation as:

$$\int_{\widetilde{t}=T_{(i)}}^{\widetilde{t}=T_{(j)}} \widehat{\theta}_C(\widetilde{t})d\widetilde{t} \approx \sum_{\ell=i}^{j-1}\widehat{\theta}_C(T_{(\ell)})\Delta_\ell.$$

- When $i > j$, we use another Riemann sum approximation as:

$$\int_{\widetilde{t}=T_{(i)}}^{\widetilde{t}=T_{(j)}} \widehat{\theta}_C(\widetilde{t})d\widetilde{t} \approx -\sum_{\ell=j}^{i-1}\widehat{\theta}_C(T_{(\ell+1)})\Delta_\ell.$$

Plugging the above results into $\frac{1}{n}\sum_{i=1}^n \int_{T_{(i)}}^{T_{(j)}} \widehat{\theta}_C(\widetilde{t})\, d\widetilde{t}$, we obtain that for $1 < j < n$,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n \int_{\widetilde{t}=T_{(i)}}^{\widetilde{t}=T_{(j)}} \widehat{\theta}_C(\widetilde{t})d\widetilde{t} &\approx \frac{1}{n}\left[\sum_{i=1}^{j-1}\sum_{\ell=i}^{j-1}\widehat{\theta}_C(T_{(\ell)})\Delta_\ell - \sum_{i=j}^{n}\sum_{\ell=j}^{i-1}\widehat{\theta}_C(T_{(\ell+1)})\Delta_\ell\right]\\
&\overset{\text{(i)}}{=} \frac{1}{n}\left[\sum_{\ell=1}^{j-1}\ell\cdot\widehat{\theta}_C(T_{(\ell)})\Delta_\ell - \sum_{\ell=j}^{n-1}(n-\ell)\cdot\widehat{\theta}_C(T_{(\ell+1)})\Delta_\ell\right]
\end{aligned}$$

$$= \frac{1}{n}\sum_{i=1}^{n-1}\Delta_i\left[i\cdot\widehat{\theta}_C(T_{(i)})\mathbb{1}_{\{i<j\}} - (n-i)\cdot\widehat{\theta}_C(T_{(i+1)})\mathbb{1}_{\{i\geq j\}}\right],$$

where the equality (i) follows from switching the orders of summations. Similarly, when $j=1$ or $j=n$, we also have that

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n}\int_{\widetilde{t}=T_{(i)}}^{\widetilde{t}=T_{(j)}}\widehat{\theta}_C(\widetilde{t})d\widetilde{t} &\approx \begin{cases}-\frac{1}{n}\sum_{i=2}^{n}\sum_{\ell=1}^{i-1}\widehat{\theta}_C(T_{(\ell+1)})\Delta_\ell & \text{when } j=1,\\ \frac{1}{n}\sum_{i=1}^{n-1}\sum_{\ell=i}^{n-1}\widehat{\theta}_C(T_{(\ell)})\Delta_\ell & \text{when } j=n,\end{cases}\\ &= \begin{cases}-\frac{1}{n}\sum_{\ell=1}^{n-1}(n-\ell)\cdot\widehat{\theta}_C(T_{(\ell+1)})\Delta_\ell & \text{when } j=1,\\ \frac{1}{n}\sum_{\ell=1}^{n-1}\ell\cdot\widehat{\theta}_C(T_{(\ell)})\Delta_\ell & \text{when } j=n,\end{cases}\\ &= \frac{1}{n}\sum_{i=1}^{n-1}\Delta_i\left[i\cdot\widehat{\theta}_C(T_{(i)})\mathbb{1}_{\{i<j\}} - (n-i)\cdot\widehat{\theta}_C(T_{(i+1)})\mathbb{1}_{\{i\geq j\}}\right].\end{aligned}$$

Therefore, we propose the following approximation for $\widehat{m}_\theta(T_{(j)})$ as:

$$\widehat{m}_\theta(T_{(j)}) \approx \frac{1}{n}\sum_{i=1}^{n}Y_i + \frac{1}{n}\sum_{i=1}^{n-1}\Delta_i\left[i\cdot\widehat{\theta}_C(T_{(i)})\mathbb{1}_{\{i<j\}} - (n-i)\cdot\widehat{\theta}_C(T_{(i+1)})\mathbb{1}_{\{i\geq j\}}\right]. \tag{13}$$

Finally, to evaluate $\widehat{m}_\theta(t)$ at any arbitrary t, we conduct the linear interpolation between $\widehat{m}_\theta(T_{(j)})$ and $\widehat{m}_\theta(T_{(j+1)})$ on the interval $t\in\left[T_{(j)},T_{(j+1)}\right]$. One biggest advantage of using the approximation formula (13) is that we only need to compute the derivative estimator $\widehat{\theta}_C(t)$ at the order statistics $T_{(1)},...,T_{(n)}$. When $\widehat{m}_\theta(t)$ and its derivative $\widehat{\theta}_C(t)$ are Lipschitz and the marginal density $p_T(t)$ is uniformly bounded away from 0 on the region of interest, this approximation formula has at most $O_P\left(\frac{1}{n}\right)$ error, which is asymptotically negligible compared to the dominating estimation error of $\widehat{m}_\theta(t)$; see Theorem 4 for details.

3.3. *Bootstrap Inference.* Since it is complicated to derive consistent estimators of the (asymptotic) variances of our integral estimator (8) and localized derivative estimator (12), we consider conducting inference on $m(t)$ and $\theta(t)$ through the empirical bootstrap method (Efron, 1979) as follows. Other bootstrap methods, including residual bootstrap (Freedman, 1981) and wild bootstrap (Wu, 1986), also work under some modified conditions.

1. Compute the integral estimator $\widehat{m}_\theta(t)$ and localized derivative estimator $\widehat{\theta}_C(t)$ on the original data $\{(Y_i,T_i,\boldsymbol{S}_i)\}_{i=1}^n$.
2. Generate B bootstrap samples $\left\{\left(Y_i^{*(b)},T_i^{*(b)},\boldsymbol{S}_i^{*(b)}\right)\right\}_{i=1}^n, b=1,...,B$ by sampling with replacement from the original data and compute the integral estimator $\widehat{m}_\theta^{*(b)}(t)$ and localized derivative estimator $\widehat{\theta}_C^{*(b)}(t)$ on each bootstrapped sample for $b=1,...,B$.
3. Let $\alpha\in(0,1)$ be a pre-specified significance level.
   - For a pointwise inference at $t_0\in\mathcal{T}$, we calculate the $1-\alpha$ quantiles $\zeta_{1-\alpha}^*(t_0)$ and $\bar{\zeta}_{1-\alpha}^*(t_0)$ of $\{D_1(t_0),...,D_B(t_0)\}$ and $\{\bar{D}_1(t_0),...,\bar{D}_B(t_0)\}$ respectively, where $D_b(t_0)=\left|\widehat{m}_\theta^{*(b)}(t_0)-\widehat{m}_\theta(t_0)\right|$ and $\bar{D}_b(t_0)=\left|\widehat{\theta}_C^{*(b)}(t_0)-\widehat{\theta}_C(t_0)\right|$ for $b=1,...,B$.
   - For an uniform inference on the entire dose-response curve $m(t)$ and its derivative $\theta(t)$, we compute the $1-\alpha$ quantiles $\xi_{1-\alpha}^*$ and $\bar{\xi}_{1-\alpha}^*$ of $\{D_{\sup,1},...,D_{\sup,B}\}$ and $\{\bar{D}_{\sup,1},...,\bar{D}_{\sup,B}\}$ respectively, where $D_{\sup,b}=\sup_{t\in\mathcal{T}}\left|\widehat{m}_\theta^{*(b)}(t)-\widehat{m}_\theta(t)\right|$ and $\bar{D}_{\sup,b}=\sup_{t\in\mathcal{T}}\left|\widehat{\theta}_C^{*(b)}(t)-\widehat{\theta}_C(t)\right|$ for $b=1,...,B$.

4. Define the $1-\alpha$ confidence intervals for $m(t_0)$ and $\theta(t_0)$ as:

$$\left[\widehat{m}_\theta(t_0) - \zeta_{1-\alpha}^*(t_0),\, \widehat{m}_\theta(t_0) + \zeta_{1-\alpha}^*(t_0)\right] \text{ and } \left[\widehat{\theta}_C(t_0) - \bar{\zeta}_{1-\alpha}^*(t_0),\, \widehat{\theta}_C(t_0) + \bar{\zeta}_{1-\alpha}^*(t_0)\right]$$

respectively, as well as the simultaneous $1-\alpha$ confidence bands as:

$$\left[\widehat{m}_\theta(t) - \xi_{1-\alpha}^*,\, \widehat{m}_\theta(t) + \xi_{1-\alpha}^*\right] \quad \text{and} \quad \left[\widehat{\theta}_C(t) - \bar{\xi}_{1-\alpha}^*,\, \widehat{\theta}_C(t) + \bar{\xi}_{1-\alpha}^*\right]$$

for every $t \in \mathcal{T}$.

In Section 4.3, we will establish the consistency of the above bootstrap inference procedures.

4. Asymptotic Theory. In this section, we study the consistency results of our integral estimator (8) and localized derivative estimator (12) proposed in Section 3.1 and the validity of bootstrap inference described in Section 3.3.

4.1. *Notations and Assumptions.* We introduce the regularity conditions under the general confounding model (2) for our subsequent theoretical analysis. Let $\mathcal{E} \subset \mathcal{T} \times \mathcal{S}$ be the support of the joint density $p(t, \boldsymbol{s})$, $\mathcal{E}^\circ$ be the interior of $\mathcal{E}$, and $\partial\mathcal{E}$ be the boundary of $\mathcal{E}$.

ASSUMPTION A3 (Differentiability of the conditional mean outcome function). For any $(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$, the conditional mean outcome function $\mu(t, \boldsymbol{s})$ is at least $(q+1)$ times continuously differentiable with respect to t and at least four times continuously differentiable with respect to $\boldsymbol{s}$, where q is the order of (partial) local polynomial regression in (9). Furthermore, $\mu(t, \boldsymbol{s})$ and all of its partial derivatives are uniformly bounded on $\mathcal{T} \times \mathcal{S}$.

ASSUMPTION A4 (Differentiability of the joint density). The joint density $p(t, \boldsymbol{s})$ is bounded and at least twice continuously differentiable with bounded partial derivatives up to the second order on $\mathcal{E}^\circ$. All these partial derivatives of $p(t, \boldsymbol{s})$ are continuous up to the boundary $\partial\mathcal{E}$. Furthermore, $\mathcal{E}$ is compact and $p(t, \boldsymbol{s})$ is uniformly bounded away from 0 on $\mathcal{E}$. Finally, the marginal density $p_T(t)$ of T is non-degenerate, *i.e.*, its support $\mathcal{T}$ has a nonempty interior.

Assumptions A3 and A4 are commonly assumed differentiability conditions in the literature of local polynomial regression (Ruppert and Wand, 1994; Fan and Gijbels, 1996), which can be slightly relaxed via the Hölder continuity condition. It ensures that the bias term of $\widehat{\beta}_2(t, \boldsymbol{s})$ from the local polynomial regression (10) is at least of the standard order $O(h^q) + O(b^2)$; see Lemma 2 below. Notice that the projection $\mathrm{proj}_T(\mathcal{E})$ of the joint density support $\mathcal{E}$ onto the domain of T coincides with the marginal support $\mathcal{T}$. Hence, $\mathcal{T}$ will be compact as well under Assumption A4.

To control the boundary effects of the local polynomial regression, we impose the following conditions on $p(t, \boldsymbol{s})$, $\mu(t, \boldsymbol{s})$, and the geometric structure of $\mathcal{E}$ near the boundary $\partial\mathcal{E}$.

ASSUMPTION A5 (Boundary conditions).

(a) There exists some constants $r_1, r_2 \in (0, 1)$ such that for any $(t, \boldsymbol{s}) \in \mathcal{E}$ and all $\delta \in (0, r_1]$, there is a point $(t', \boldsymbol{s}') \in \mathcal{E}$ satisfying

$$\mathcal{B}\left((t', \boldsymbol{s}'),\, r_2\delta\right) \subset \mathcal{B}\left((t, \boldsymbol{s}),\, \delta\right) \cap \mathcal{E},$$

where $\mathcal{B}((t, \boldsymbol{s}), r) = \left\{(t_1, \boldsymbol{s}_1) \in \mathbb{R}^{d+1} : \left|\left|(t_1 - t, \boldsymbol{s}_1 - \boldsymbol{s})\right|\right|_2 \leq r\right\}$ with $||\cdot||_2$ being the standard Euclidean norm.

(b) For any $(t,\boldsymbol{s}) \in \partial\mathcal{E}$, the boundary of $\mathcal{E}$, it satisfies that $\frac{\partial}{\partial t}p(t,\boldsymbol{s}) = \frac{\partial}{\partial s_j}p(t,\boldsymbol{s}) = 0$ and $\frac{\partial^2}{\partial s_j^2}\mu(t,\boldsymbol{s}) = 0$ for all $j = 1,...,d$.

(c) For any $\delta > 0$, the Lebesgue measure of the set $\partial\mathcal{E} \oplus \delta$ satisfies $|\partial\mathcal{E} \oplus \delta| \leq A_1 \cdot \delta$ for some absolute constant $A_1 > 0$, where $\partial\mathcal{E} \oplus \delta = \left\{\boldsymbol{z} \in \mathbb{R}^{d+1} : \inf_{\boldsymbol{x}\in\partial\mathcal{E}} ||\boldsymbol{z} - \boldsymbol{x}||_2 \leq \delta\right\}$.

Assumption A5(a) is adopted from the boundary condition (Assumption X) in Fan and Guerre (2015), whose primitive and stronger version also appeared as Assumption (A4) in Ruppert and Wand (1994). This is a relatively mild condition that holds when the boundary $\partial\mathcal{E}$ is smooth or only contains non-smooth vertices from some regular structures, such as $d+1$ dimensional cubes and convex cones. Indeed, Assumption A5(a) is valid as long as any ball centered at a point in $\mathcal{E}$ near a vertex of $\partial\mathcal{E}$ has its radius shrunk linearly when approaching the vertex. More examples and discussions about Assumption A5(b) can be found in Ruppert and Wand (1994); Fan and Guerre (2015). The main purpose of imposing this support condition is to ensure that there are enough observations near the boundary points so that the rate of convergence for our local polynomial estimator $\widehat{\beta}_2(t,\boldsymbol{s})$ remains the same order at the boundary points as the interior points of $\mathcal{E}$. Assumption A5(b) regularizes the slope of $p(t,\boldsymbol{s})$ and the curvature of $\mu(t,\boldsymbol{s})$ at the boundary point $(t,\boldsymbol{s}) \in \partial\mathcal{E}$. Specifically, at the boundary $\partial\mathcal{E}$, the joint density $p(t,\boldsymbol{s})$ needs to be flat, while $\mu(t,\boldsymbol{s})$ should embrace zero curvatures. Similar to Assumption A4, the partial derivatives $\frac{\partial}{\partial t}p(t,\boldsymbol{s}), \frac{\partial}{\partial s_j}p(t,\boldsymbol{s})$ are defined by computing them at a nearby interior point $(t',\boldsymbol{s}')$ and taking the limit $(t',\boldsymbol{s}') \to (t,\boldsymbol{s}) \in \partial\mathcal{E}$. This condition is another key requirement for the bias term from $\widehat{\beta}_2(t,\boldsymbol{s})$ in the local polynomial regression (10) to remain the same rate of convergence at the boundary points as the interior points of $\mathcal{E}$. Finally, Assumption A5(c) also regularizes the boundary $\partial\mathcal{E}$ so that it will not have any fractal or other peculiar structures leading to an infinite perimeter of $\partial\mathcal{E}$.

To establish the (uniform) consistency of our local polynomial regression estimator (10) and Nadaraya-Watson conditional CDF estimator (11), we rely on the following regularity conditions on the kernel functions.

ASSUMPTION A6 (Regular kernel and VC-type conditions).

(a) The functions $K_T : \mathbb{R} \to [0,\infty)$ and $K_S : \mathbb{R}^d \to [0,\infty)$ are compactly supported and Lispchitz continuous kernels such that $\int_{\mathbb{R}} K_T(t)\,dt = \int_{\mathbb{R}^d} K_S(\boldsymbol{s})\,d\boldsymbol{s} = 1$, $K_T(t) = K_T(-t)$, and K_S is radially symmetric with $\int \boldsymbol{s} \cdot K_S(\boldsymbol{s})d\boldsymbol{s} = \boldsymbol{0}$. In addition, for all $j = 1,2,...,$ and $\ell = 1,...,d$, it holds that

$$\kappa_j^{(T)} := \int_{\mathbb{R}} u^j K_T(u)\,du < \infty, \quad \nu_j^{(T)} := \int_{\mathbb{R}} u^j K_T^2(u)\,du < \infty,$$

$$\kappa_{j,\ell}^{(S)} := \int_{\mathbb{R}^d} u_\ell^j K_S(\boldsymbol{u})\,d\boldsymbol{u} < \infty, \quad \text{and} \quad \nu_{j,k}^{(S)} := \int_{\mathbb{R}^d} u_\ell^j K_S^2(\boldsymbol{u})\,d\boldsymbol{u} < \infty.$$

Finally, both K_T and K_S are second-order kernels, *i.e.*, $\kappa_2^{(T)} > 0$ and $\kappa_{2,\ell}^{(S)} > 0$ for all $\ell = 1,...,d$.

(b) Let $\mathcal{K}_{q,d} = \left\{(y,\boldsymbol{z}) \mapsto \left(\frac{y-t}{h}\right)^\ell \left(\frac{z_i - s_i}{b}\right)^{k_1} \left(\frac{z_j - s_j}{b}\right)^{k_2} K_T\left(\frac{y-t}{h}\right) K_S\left(\frac{\boldsymbol{z}-\boldsymbol{s}}{b}\right) : (t,\boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}; i,j = 1,...,d; \ell = 0,...,2q; k_1,k_2 = 0,1; h,b > 0\right\}$. It holds that $\mathcal{K}_{q,d}$ is a bounded VC-type class of measurable functions on $\mathbb{R}^{d+1}$.

(c) The function $\bar{K}_T : \mathbb{R} \to [0,\infty)$ is a second-order, Lipschitz continuous, and symmetric kernel with a compact support, *i.e.*, $\int_{\mathbb{R}} \bar{K}_T(t)\,dt = 1$, $\bar{K}_T(t) = \bar{K}_T(-t)$, and $\int_{\mathbb{R}} t^2 \bar{K}_T(t)\,dt \in (0,\infty)$.

(d) Let $\bar{\mathcal{K}} = \left\{y \mapsto \bar{K}_T\left(\frac{y-t}{\hbar}\right) : t \in \mathcal{T}, \hbar > 0\right\}$. It holds that $\bar{\mathcal{K}}$ is a bounded VC-type class of measurable functions on $\mathbb{R}$.

Assumption A6(a,c) are indeed not the regularity conditions but rather properties of those commonly used kernel functions. Assumption A6(b,d) are critical conditions for the uniform consistency of kernel-based function estimators (Giné and Guillou, 2002; Einmahl and Mason, 2005) and can be satisfied by a wide range of kernel functions, including Gaussian and Epanechnikov kernels. For example, it is satisfied when the kernel is a composite function between a polynomial in $d \geq 1$ variables and a real-valued function of bounded variation; see Lemma 22 in Nolan and Pollard (1987). Given that the kernel functions are bounded, we can always take constant envelope functions for $\mathcal{K}_{q,d}$ and $\bar{\mathcal{K}}$ in Assumption A6(b,d).

Finally, we introduce some notations that will be used in proofs of the consistency results of $\widehat{\beta}_2(t,\boldsymbol{s})$ from the local polynomial regression (10) (Lemmas 2 and 3) and the asymptotic linearity results of our integral estimator and localized derivative estimator (Lemma 5). Under the notations in Assumption A6(a), we define a matrix

(14)

$$\boldsymbol{M}_q = \begin{pmatrix} \left(\kappa_{i+j-2}^{(T)}\right)_{1\leq i,j\leq q+1} & \boldsymbol{0} \\ \boldsymbol{0} & \left(\kappa_{2,i-q-1}^{(S)} \mathbb{1}_{\{i=j\}}\right)_{q+1<i,j\leq q+1+d} \end{pmatrix} \in \mathbb{R}^{(q+1+d)\times(q+1+d)}.$$

Notice that $\boldsymbol{M}_q$ only depends on the kernel functions K_T, K_S in the local polynomial regression. For any $(t,\boldsymbol{s}) \in \mathcal{T}\times\mathcal{S}$, we also define the functions $\boldsymbol{\Psi}_{t,\boldsymbol{s}}, \boldsymbol{\psi}_{t,\boldsymbol{s}} : \mathbb{R}\times\mathbb{R}\times\mathbb{R}^d \to \mathbb{R}^{q+1+d}$ as:

(15)

$$\begin{aligned}\boldsymbol{\Psi}_{t,\boldsymbol{s}}(y,z,\boldsymbol{v}) &= \begin{bmatrix} \left(y\cdot\left(\frac{z-t}{h}\right)^{j-1} K_T\left(\frac{z-t}{h}\right) K_S\left(\frac{\boldsymbol{v}-\boldsymbol{s}}{b}\right)\right)_{1\leq j\leq q+1} \\ \left(y\cdot\left(\frac{v_{j-q-1}-s_{j-q-1}}{b}\right) K_T\left(\frac{z-t}{h}\right) K_S\left(\frac{\boldsymbol{v}-\boldsymbol{s}}{b}\right)\right)_{q+1<j\leq q+1+d} \end{bmatrix} \\ &= y\cdot\begin{bmatrix} \left(\left(\frac{z-t}{h}\right)^{j-1} K_T\left(\frac{z-t}{h}\right) K_S\left(\frac{\boldsymbol{v}-\boldsymbol{s}}{b}\right)\right)_{1\leq j\leq q+1} \\ \left(\left(\frac{v_{j-q-1}-s_{j-q-1}}{b}\right) K_T\left(\frac{z-t}{h}\right) K_S\left(\frac{\boldsymbol{v}-\boldsymbol{s}}{b}\right)\right)_{q+1<j\leq q+1+d} \end{bmatrix} \equiv y\cdot\boldsymbol{\psi}_{t,\boldsymbol{s}}(z,\boldsymbol{v}).\end{aligned}$$

4.2. *Consistency of the Integral Estimator.* Before establishing the consistency of our proposed integral estimator (8) and localized derivative estimator (12), we first discuss the pointwise and uniform consistency results of $\widehat{\beta}_2(t,\boldsymbol{s})$ from the local polynomial regression (10) as building blocks.

LEMMA 2 (Pointwise convergence of $\widehat{\beta}_2(t,\boldsymbol{s})$). *Suppose that Assumptions A3, A4, A5, and A6(a) hold under the confounding model (2). Let $\widehat{\beta}_2(t,\boldsymbol{s})$ be the second element of $\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}) \in \mathbb{R}^{q+1}$ defined by (10) and $\beta_2(t,\boldsymbol{s}) = \frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$. Then, for any fixed $(t,\boldsymbol{s}) \in \mathcal{E}$ and integer $q > 0$, as $h, b, \frac{\max\{h,b\}^4}{h} \to 0$, and $nh^3b^d \to \infty$, we have that*

$$\begin{aligned}&\widehat{\beta}_2(t,\boldsymbol{s}) - \beta_2(t,\boldsymbol{s}) \\ &= \begin{cases} O\left(h^{q+1} + b^2 + \frac{b^4}{h}\right) + O_P\left(\sqrt{\frac{1}{nh^3b^d}}\right) & \text{if } q \text{ is odd and } (t,\boldsymbol{s}) \in \mathcal{E}^\circ, \\ O\left(h^{q} + b^2 + \frac{b^4}{h}\right) + O_P\left(\sqrt{\frac{1}{nh^3b^d}}\right) & \text{if } q \text{ is even and } (t,\boldsymbol{s}) \in \mathcal{E}^\circ, \\ O\left(h^{q} + \frac{\max\{h,b\}^4}{h}\right) + O_P\left(\sqrt{\frac{1}{nh^3b^d}}\right) & \text{if } q \text{ is an integer and } (t,\boldsymbol{s}) \in \partial\mathcal{E}. \end{cases}\end{aligned}$$

The complete statement of the asymptotic expressions for the conditional variances and bias of $\widehat{\beta}_2(t,\boldsymbol{s})$ and the proof of Lemma 2 are given in Section B.2. Notice that our (conditional) rates of convergence of $\widehat{\beta}_2(t,\boldsymbol{s})$ in Lemma 2 align with the standard results in the literature of local polynomial regression (Ruppert and Wand, 1994; Fan and Gijbels, 1996; Lu, 1996). The extra bias rate $O\left(\frac{b^4}{h}\right)$ or $O\left(\frac{\max\{h,b\}^4}{h}\right)$ comes from a higher order term in the Tyler's expansion and will be asymptotically negligible when $h \asymp b$ as $n \to \infty$; see (26) and (28) in Section B.2 for an example. More importantly, when we utilize the (partial) local quadratic regression (*i.e.*, $q=2$) to obtain $\widehat{\beta}_2(t,\boldsymbol{s})$, the order of $h \asymp b$ that optimally trades off the bias and variance in Lemma 2 will be $O\left(n^{-\frac{1}{d+7}}\right)$, which also leads to the optimal rate for derivative estimation established in Stone (1980, 1982). Conventional bandwidth selection methods generally will not yield h, b with this optimal order. However, the validity of bootstrap inference in Section 4.3 requires bandwidths h, b to be of a smaller order $O\left(n^{-\frac{1}{d+5}}\right)$ so that the asymptotic bias is negligible compared to the stochastic variation (Bjerve et al., 1985; Hall, 1992); see Lemma 5 below. Since $O\left(n^{-\frac{1}{d+5}}\right)$ is an optimal bandwidth order for estimating the regression function $\mu(t,\boldsymbol{s})$, we can then apply any off-the-shelf bandwidth selection method of multivariate local polynomial regression (Wand and Jones, 1994; Yang and Tschernig, 1999; Li and Racine, 2004) to obtain our bandwidth parameters $h \asymp b$; see also Remark 4 below.

We also strengthen the pointwise rate of convergence to the uniform one as follows.

LEMMA 3 (Uniform convergence of $\widehat{\beta}_2(t,\boldsymbol{s})$). *Let $q>0$. Suppose that Assumptions A3, A4, A5, and A6(a,b) hold under the confounding model (2). Let $\widehat{\beta}_2(t,\boldsymbol{s})$ be the second element of $\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}) \in \mathbb{R}^{q+1}$ and $\beta_2(t,\boldsymbol{s}) = \frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$. Then, as $h, b, \frac{\max\{h,b\}^4}{h} \to 0$ and $\frac{nh^3b^d}{|\log(hb^d)|}, \frac{|\log(hb^d)|}{\log\log n} \to \infty$, we have that*

$$\sup_{(t,\boldsymbol{s})\in\mathcal{E}} \left|\widehat{\beta}_2(t,\boldsymbol{s}) - \beta_2(t,\boldsymbol{s})\right| = O\left(h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right) + O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}}\right).$$

The proof of Lemma 3 can be found in Section B.3. At a high level, the uniform rate of convergence for the bias term $\mathbb{E}\left[\widehat{\beta}_2(t,\boldsymbol{s})\right] - \beta_2(t,\boldsymbol{s})$ remains the same as in Lemma 2 under our regularity conditions. To handle the stochastic variation term $\widehat{\beta}_2(t,\boldsymbol{s}) - \mathbb{E}\left[\widehat{\beta}_2(t,\boldsymbol{s})\right]$, we approximate it (up to a scaled factor $\sqrt{nh^3b^d}$) by an empirical process so that the upper bound for its uniform rate of convergence follows from the results in Einmahl and Mason (2005). This uniform rate of convergence in Lemma 3 is not only useful for deriving the asymptotic behaviors of our integral estimator (8) and localized derivative estimator (12) in Theorem 4 below but can also facilitate our study of the bootstrap consistency in Section 4.3.

With Lemma 3, we now present the uniform consistency results for our integral estimator (8) and localized derivative estimator (12).

THEOREM 4 (Convergence of $\widehat{\theta}_C(t)$ and $\widehat{m}_\theta(t)$). *Let $q>0$ and $\mathcal{T}' \subset \mathcal{T}$ be a compact set so that $p_T(t)$ is uniformly bounded away from 0 within $\mathcal{T}'$. Suppose that Assumptions A1, A2, A3, A4, A5, and A6 hold. Then, as $h, b, \hbar, \frac{\max\{h,b\}^4}{h} \to 0$ and $\frac{nh^3b^d}{|\log(hb^d)|}, \frac{|\log(hb^d)|}{\log\log n}, \frac{n\hbar}{|\log\hbar|}, \frac{|\log\hbar|}{\log\log n} \to \infty$, we know that*

$$\sup_{t\in\mathcal{T}'} \left|\widehat{\theta}_C(t) - \theta_C(t)\right| = O\left(h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right) + O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}} + \hbar^2 + \sqrt{\frac{|\log\hbar|}{n\hbar}}\right)$$

and

$$\sup_{t\in\mathcal{T}'}|\widehat{m}_\theta(t)-m(t)| = O_P\left(\frac{1}{\sqrt{n}}\right) + O\left(h^q+b^2+\frac{\max\{b,h\}^4}{h}\right) + O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}}+\hslash^2+\sqrt{\frac{|\log\hslash|}{n\hslash}}\right).$$

The pointwise rates of convergence for $\widehat{\theta}_C(t)$ and $\widehat{m}_\theta(t)$ will remain the same as stated in Theorem 4 due to the randomness of observations near the boundary of density support $\mathcal{E}$; see the proof of Theorem 4 in Section B.4 for detailed arguments. We restrict the uniform consistency results in Theorem 4 to a compact set $\mathcal{T}'\subset\mathcal{T}$ in order to avoid the density decay of $p_T(t)$ near the boundary of its support $\mathcal{T}$. If $p_T(t)$ is uniformly bounded away from 0 in its support, we can take $\mathcal{T}'=\mathcal{T}$. The (uniform) rate of convergence of $\widehat{m}_\theta(t)$ consists of two parts. The first part $O_P\left(\frac{1}{\sqrt{n}}\right)$ comes from the simple sample average of $Y_i, i=1,...,n$ and is asymptotically negligible. The second dominant part is due to the integral component

$$\widehat{\Delta}_{h,b}(t) = \frac{1}{n}\sum_{i=1}^n \int_{\tilde{t}=T_i}^{\tilde{t}=t}\widehat{\theta}_C(\tilde{t})d\tilde{t}, \tag{16}$$

whose rate of convergence is determined by $\widehat{\theta}_C(t)$. Furthermore, the (uniform) rate of convergence of $\widehat{\theta}_C(t)$ comprises the rate $O\left(h^q+b^2+\frac{\max\{b,h\}^4}{h}\right)+O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}}\right)$ for estimating the partial derivative $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ and the rate $O(\hslash^2)+O_P\left(\sqrt{\frac{|\log\hslash|}{n\hslash}}\right)$ for estimating the conditional CDF $\mathrm{P}(\boldsymbol{s}|t)$.

4.3. *Validity of the Bootstrap Inference.* Before proving the consistency of our bootstrap procedure in Section 3.3, we first derive the asymptotic linearity of our integral estimator (8) and localized derivative estimator (12) as intermediate results.

LEMMA 5 (Asymptotic linearity). *Let $q\geq 2$ in the local polynomial regression for estimating $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ and $\mathcal{T}'\subset\mathcal{T}$ be a compact set so that $p_T(t)$ is uniformly bounded away from 0 within $\mathcal{T}'$. Suppose that Assumptions A1, A2, A3, A4, A5, and A6 hold. Then, if $h\asymp b\asymp n^{-\frac{1}{\gamma}}$ and $\hslash\asymp n^{-\frac{1}{\varpi}}$ for some $\gamma\geq\varpi>0$ such that $\frac{nh^{d+5}}{\log n}\to c_1$ and $\frac{n\hslash^5}{\log n}\to c_2$ for some finite number $c_1,c_2\geq 0$ and $\frac{\log n}{n\hslash^2}, \frac{h^{d+3}\log n}{\hslash}, \frac{h^{d+3}}{\hslash^2}\to 0$ as $n\to\infty$, then for any $t\in\mathcal{T}'$, we have that*

$$\sqrt{nh^3b^d}\left[\widehat{\theta}_C(t)-\theta_C(t)\right]=\mathbb{G}_n\bar{\varphi}_t+o_P(1) \quad \text{and} \quad \sqrt{nh^3b^d}\left[\widehat{m}_\theta(t)-m(t)\right]=\mathbb{G}_n\varphi_t+o_P(1),$$

where $\bar{\varphi}_t(Y,T,\boldsymbol{S})=\mathbb{E}_{(T_{i_3},\boldsymbol{S}_{i_3})}\left[\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{S}_{i_3}}(Y,T,\boldsymbol{S})}{\sqrt{hb^d}\cdot p(t,\boldsymbol{S}_{i_3})\cdot p_T(t)}\cdot\frac{1}{\hslash}\bar{K}_T\left(\frac{t-T_{i_3}}{\hslash}\right)\right]$ *and*

$$\begin{aligned}\varphi_t(Y,T,\boldsymbol{S}) &= \mathbb{E}_{T_{i_2}}\left[\int_{T_{i_2}}^t \bar{\varphi}_{\tilde{t}}(Y,T,\boldsymbol{S})\,d\tilde{t}\right]\\ &= \mathbb{E}_{T_{i_2}}\left\{\int_{T_{i_2}}^t \mathbb{E}_{(T_{i_3},\boldsymbol{S}_{i_3})}\left[\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\tilde{t},\boldsymbol{S}_{i_3}}(Y,T,\boldsymbol{S})}{\sqrt{hb^d}\cdot p(\tilde{t},\boldsymbol{S}_{i_3})\cdot p_T(\tilde{t})}\cdot\frac{1}{\hslash}\bar{K}_T\left(\frac{\tilde{t}-T_{i_3}}{\hslash}\right)\right]d\tilde{t}\right\}.\end{aligned} \tag{17}$$

Furthermore, we have the following uniform results as:

$$\left|\sqrt{nh^3b^d}\sup_{t\in\mathcal{T}'}\left|\widehat{\theta}_C(t)-\theta_C(t)\right|-\sup_{t\in\mathcal{T}'}\left|\mathbb{G}_n\bar{\varphi}_t\right|\right|=O_P\left(\frac{\log n}{\sqrt{n\hbar}}+\sqrt{nh^{d+7}}+\sqrt{\frac{h^{d+3}\log n}{\hbar}}\right)$$

and

$$\left|\sqrt{nh^3b^d}\sup_{t\in\mathcal{T}'}\left|\widehat{m}_\theta(t)-m(t)\right|-\sup_{t\in\mathcal{T}'}\left|\mathbb{G}_n\varphi_t\right|\right|$$
$$=O_P\left(\sqrt{nh^{d+7}}+\sqrt{\frac{\log n}{n\hbar^2}}+\sqrt{\frac{h^{d+3}\log n}{\hbar}}+\sqrt{\frac{h^{d+3}}{\hbar^2}}\right).$$

The proof of Lemma 5 is in Section B.5, in which our key argument is to write $\widehat{m}_\theta(t)$ and $\widehat{\theta}_C(t)$ in the form of V-statistics (Shieh, 2014). We make two remarks to this crucial lemma.

REMARK 3 (Non-degeneracy and validity of pointwise confidence intervals). We prove in Lemma 13 of Section B.6 that the variances $\mathrm{Var}\left[\varphi_t(Y,T,\boldsymbol{S})\right]$ and $\mathrm{Var}\left[\bar{\varphi}_t(Y,T,\boldsymbol{S})\right]$ of the influence functions in (17) are positive for each $t\in\mathcal{T}'$ as long as $\mathrm{Var}(\epsilon)=\sigma^2>0$ in model (2). Thus, the asymptotic linearity results in Lemma 5 are non-degenerate. Similarly, one can show that the bootstrap estimates $\widehat{m}_\theta^*(t)$ and $\widehat{\theta}_C^*(t)$ are also asymptotically linear given the observed data $\{(Y_i,T_i,\boldsymbol{S}_i)\}_{i=1}^n$. It indicates that the pointwise bootstrap confidence intervals for $m(t)$ and $\theta(t)$ in Section 3.3 are asymptotically valid; see Lemma 23.3 in van der Vaart (1998) and related results in Arcones and Gine (1992); Tang and Westling (2024).

REMARK 4 (Bandwidth selection). In order for those remainder terms of the asymptotically linear forms in Lemma 5 to be of the order $o_P(1)$, we can choose the bandwidths $h\asymp b$ to be of the order $O\left(n^{-\frac{1}{d+5}}\right)$ and $\hbar$ to be of the order $O\left(n^{-\frac{1}{5}}\right)$, both of which match the outputs by the usual bandwidth selection methods for nonparametric regression (Wasserman, 2006; Schindler, 2011). In other words, we can tune $h,b,\hbar$ via the standard bandwidth selectors for estimating the regression function $\mu(t,\boldsymbol{s})$ and conditional CDF $p(\boldsymbol{s}|t)$ without any explicit undersmoothing.

Apart from asymptotic linearity, we establish couplings between $\sqrt{nh^3b^d}\sup_{t\in\mathcal{T}'}\left|\widehat{m}_\theta(t)-m(t)\right|$ and $\sup_{t\in\mathcal{T}'}\left|\mathbb{G}_n(\varphi_t)\right|$ (and similarly, for $\widehat{\theta}_C(t)$) in Lemma 5. These coupling results serve as key ingredients for deriving the Gaussian approximations for $\widehat{m}_\theta(t)$ and $\widehat{\theta}_C(t)$ in Theorem 6 below and their bootstrap consistencies. Consider two function classes

$$\mathcal{F}=\left\{(v,x,\boldsymbol{z})\mapsto\varphi_t(v,x,\boldsymbol{z}):t\in\mathcal{T}'\right\}\quad\text{and}\quad\mathcal{F}_\theta=\left\{(v,x,\boldsymbol{z})\mapsto\bar{\varphi}_t(v,x,\boldsymbol{z}):t\in\mathcal{T}'\right\}\tag{18}$$

with $\varphi_t,\bar{\varphi}_t$ defined in (17) respectively. Let $\mathbb{B},\bar{\mathbb{B}}$ be two Gaussian processes indexed by $\mathcal{F}$ and $\mathcal{F}_\theta$ respectively with zero means and covariance functions

$$\mathrm{Cov}(\mathbb{B}(f_1),\mathbb{B}(f_2))=\mathbb{E}\left[f_1(\boldsymbol{U})\cdot f_2(\boldsymbol{U})\right]\quad\text{and}\quad\mathrm{Cov}(\bar{\mathbb{B}}(g_1),\bar{\mathbb{B}}(g_2))=\mathbb{E}\left[g_1(\boldsymbol{U})\cdot g_2(\boldsymbol{U})\right]$$

with $\boldsymbol{U}=(Y,T,\boldsymbol{S})$ for any $f_1,f_2\in\mathcal{F}$ and $g_1,g_2\in\mathcal{F}_\theta$.

THEOREM 6 (Gaussian approximation). *Let $q\geq 2$ in the local polynomial regression for estimating $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ and $\mathcal{T}'\subset\mathcal{T}$ be a compact set so that $p_T(t)$ is uniformly bounded away from 0 within $\mathcal{T}'$. Suppose that Assumptions A1, A2, A3, A4, A5, and A6 hold. If $h\asymp b\asymp n^{-\frac{1}{\gamma}}$ and $\hbar\asymp n^{-\frac{1}{\varpi}}$ for some $\gamma\geq\varpi>0$ such that $\frac{nh^{d+5}}{\log n}\to c_1$ and $\frac{n\hbar^5}{\log n}\to c_2$ for some finite* *number* $c_1, c_2 \geq 0$ *and* $\frac{n\hbar^2}{\log n}, \frac{\hbar}{h^{d+3}\log n}, \hbar n^{\frac{1}{4}}, \frac{\hbar^2}{h^{d+3}} \to \infty$ *as* $n \to \infty$*, then there exist Gaussian processes* $\mathbb{B}, \bar{\mathbb{B}}$ *such that*

$$\sup_{u\geq 0}\left|\mathrm{P}\left(\sqrt{nh^3b^d}\cdot\sup_{t\in\mathcal{T}'}|\widehat{m}_\theta(t)-m(t)|\leq u\right)-\mathrm{P}\left(\sup_{f\in\mathcal{F}}|\mathbb{B}(f)|\leq u\right)\right| = O\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right),$$

$$\sup_{u\geq 0}\left|\mathrm{P}\left(\sqrt{nh^3b^d}\cdot\sup_{t\in\mathcal{T}'}\left|\widehat{\theta}_C(t)-\theta_C(t)\right|\leq u\right)-\mathrm{P}\left(\sup_{g\in\mathcal{F}_\theta}|\bar{\mathbb{B}}(g)|\leq u\right)\right| = O\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right),$$

where $\mathcal{F}, \mathcal{F}_\theta$ *are defined in* (18).

The proof of Theorem 6 is in Section B.6. It demonstrates that the distributions of $\sqrt{nh^3b^d}\cdot\sup_{t\in\mathcal{T}'}|\widehat{m}_\theta(t)-m(t)|$ and $\sqrt{nh^3b^d}\cdot\sup_{t\in\mathcal{T}'}\left|\widehat{\theta}_C(t)-\theta_C(t)\right|$ can be asymptotically approximated by the suprema of two separate Gaussian processes respectively. To establish the consistency of our bootstrap inference procedure in Section 3.3, it remains to show that the bootstrap versions of the differences $\sqrt{nh^3b^d}\cdot\sup_{t\in\mathcal{T}'}|\widehat{m}^*_\theta(t)-\widehat{m}_\theta(t)|$ and $\sqrt{nh^3b^d}\cdot\sup_{t\in\mathcal{T}'}\left|\widehat{\theta}^*_C(t)-\widehat{\theta}_C(t)\right|$, conditioning on the observed data $\boldsymbol{U}_n=\{(Y_i,T_i,\boldsymbol{S}_i)\}_{i=1}^n$, can be asymptotically approximated by the suprema of the same Gaussian processes, which are summarized in the following theorem.

THEOREM 7 (Bootstrap consistency). *Let* $\mathbb{U}_n=\{(Y_i,T_i,\boldsymbol{S}_i)\}_{i=1}^n$ *be the observed data. Under the same setup of Theorem 6, we have that*

$$\sup_{u\geq 0}\left|\mathrm{P}\left(\sqrt{nh^3b^d}\cdot\sup_{t\in\mathcal{T}'}|\widehat{m}^*_\theta(t)-\widehat{m}_\theta(t)|\leq u\middle|\mathbb{U}_n\right)-\mathrm{P}\left(\sup_{f\in\mathcal{F}}|\mathbb{B}(f)|\leq u\right)\right| = O_P\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right)$$

$$\sup_{u\geq 0}\left|\mathrm{P}\left(\sqrt{nh^3b^d}\cdot\sup_{t\in\mathcal{T}'}\left|\widehat{\theta}^*_C(t)-\widehat{\theta}_C(t)\right|\leq u\middle|\mathbb{U}_n\right)-\mathrm{P}\left(\sup_{g\in\mathcal{F}_\theta}|\bar{\mathbb{B}}(g)|\leq u\right)\right| = O_P\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right),$$

where $\widehat{m}^*_\theta(t)$ *and* $\widehat{\theta}^*_C(t)$ *are the integral estimator* (8) *and localized derivative estimator* (12) *based on a bootstrap sample* $\mathbb{U}^*_n=\{(Y^*_i,T^*_i,\boldsymbol{S}^*_i)\}_{i=1}^n$.

The proof of Theorem 7 is in Section B.7. These results, together with Theorem 6, imply the asymptotic validity of bootstrap uniform confidence bands in Section 3.3; see Corollary 8 below with its proof in Section B.8.

COROLLARY 8 (Uniform confidence band). *Under the setup of Theorem 7, we have that*

$$\mathrm{P}\left(\theta(t)\in\left[\widehat{\theta}_C(t)-\bar{\xi}^*_{1-\alpha},\widehat{\theta}_C(t)+\bar{\xi}^*_{1-\alpha}\right] \text{ for all } t\in\mathcal{T}'\right) = 1-\alpha+O\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right),$$

$$\mathrm{P}\left(m(t)\in\left[\widehat{m}_\theta(t)-\xi^*_{1-\alpha},\widehat{m}_\theta(t)+\xi^*_{1-\alpha}\right] \text{ for all } t\in\mathcal{T}'\right) = 1-\alpha+O\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right).$$

5. Experiments.

In this section, we evaluate the finite-sample performances of our proposed integral estimator (8) and localized derivative estimator (12) through several simulation studies. We also apply them to analyzing the causal effects of fine particulate matter on the cardiovascular mortality rate as a case study.

5.1. *Parameter Setup.* Throughout the experiments, we use the Epanechnikov kernel for K_T and K_S (with the product kernel technique) in the (partial) local quadratic regression (10). To choose its default bandwidth parameters $h, b > 0$, we implement the rule-of-thumb method in Appendix A of Yang and Tschernig (1999) as:

$$h_{ROT} = C_h \left[\frac{\left(\nu_0^{(T)}\right)^2 \widehat{R}\left(T_{(n)} - T_{(1)}\right)}{4\left(\kappa_2^{(T)}\right)^2 \widehat{C}_\mu} \right]^{\frac{1}{5}} n^{-\frac{1}{d+5}}, \; \boldsymbol{b}_{ROT} = C_b \left[\frac{d\left(\nu_{0,1}^{(S)}\right)^{2d} \widehat{R}\left(\boldsymbol{S}_{(n)} - \boldsymbol{S}_{(1)}\right)}{4n\left(\kappa_{2,1}^{(S)}\right)^2 \widehat{C}_\mu} \right]^{-\frac{1}{d+5}},$$

where $\nu_0^{(T)}, \nu_{0,1}^{(S)}, \kappa_2^{(T)}, \kappa_{2,1}^{(S)} > 0$ are defined in Assumption A6(a), $\widehat{R} = \frac{1}{n-5}\sum_{i=1}^n \left[Y - \widehat{m}(T_i, \boldsymbol{S}_i)\right]^2$ with $\widehat{m}(t, \boldsymbol{s}) = \widehat{\beta}_1(t, \boldsymbol{s})$ from the local polynomial regression (10) with $q = 4$ and $K_T = K_S \equiv 1$ (*i.e.*, global (partial) forth-order polynomial fitting), $\boldsymbol{S}_{(n)} - \boldsymbol{S}_{(1)}$ are the differences between coordinatewise maxima and minima that approximate the integration ranges, and $\widehat{C}_{\mu,T}, \widehat{C}_{\mu,\boldsymbol{S}}$ are the estimated density-weighted curvatures (or second-order partial derivatives) of μ. We obtain $\widehat{C}_{\mu,T}, \widehat{C}_{\mu,\boldsymbol{S}}$ by fitting coordinatewise global fourth-order polynomials and averaging the estimates over the observed data $\{(T_i, \boldsymbol{S}_i), i = 1, ..., n\}$; see also Section 4.3 in García Portugués (2023). Here, we set $\boldsymbol{b}_{ROT} \in [0, \infty)^d$ as a bandwidth vector instead of a scalar $b > 0$ so that it can be more adaptive to the scale of each covariate. In addition, unless stated otherwise, we set the scaling constants to be $C_h = 7, C_b = 3$ for all the experiments. Finally, to maintain fair comparisons between our proposed estimators and the usual regression adjustment estimators $\widehat{m}_{RA}(t)$ in (6) and $\widehat{\theta}_{RA}(t)$ in Remark 1, we obtain $\widehat{\mu}(t, \boldsymbol{s})$ and $\widehat{\beta}_2(t, \boldsymbol{s})$ using the (partial) local quadratic regression (10) with the same choices of bandwidth parameters h, b.

As for the conditional CDF estimator (11), we leverage the Gaussian kernel for $\bar{K}_T$ and set its default bandwidth parameter $\hbar$ to the normal reference rule in Chacón et al. (2011); Chen et al. (2016) as $\hbar_{NR} = \left(\frac{4}{3n}\right)^{\frac{1}{5}} \widehat{\sigma}_T$, where $\widehat{\sigma}_T$ is the sample standard deviation of $\{T_i, i = 1, ..., n\}$. This rule is obtained by assuming a normal distribution of $p_T(t)$ and optimizing the asymptotic bias-variance trade-off, which could potentially oversmooth the conditional CDF estimator (Sheather, 2004). However, since our proposed estimators of $m(t)$ and $\theta(t)$ are not very sensitive to the choice of $\hbar$ and the order of $\hbar_{NR}$ aligns with our theoretical requirement (see Remark 4), we would stick to $\hbar_{NR}$ for $\hbar$ in the subsequent analysis.

As for the bootstrap inference, we set the resampling time $B = 1000$ and the significance level $\alpha = 0.05$, *i.e.*, targeting at 95% confidence intervals and uniform bands for inferring the true dose-response curve $m(t)$ and its derivative $\theta(t)$.

5.2. *Simulation Studies.* We consider three different model settings under the additive confounding model (3) for our simulation studies. For each model setting, we generate $\{(Y_i, T_i, \boldsymbol{S}_i)\}_{i=1}^n$ with $n = 2000$ i.i.d. observations.

- **Single Confounder Model:** The data-generating model is described in (1).
- **Linear Confounding Model:** We consider the following linear effect model with

$$Y = T + 6S_1 + 6S_2 + \epsilon, \quad T = 2S_1 + S_2 + E, \quad \text{and} \quad \boldsymbol{S} = (S_1, S_2) \sim \text{Uniform}[-1,1]^2 \subset \mathbb{R}^2, \tag{19}$$

where $E \sim \text{Uniform}[-0.5, 0.5]$ and $\epsilon \sim \mathcal{N}(0,1)$. The marginal supports of T and $\boldsymbol{S}$ are $\mathcal{T} = [-3.5, 3.5]$ and $\mathcal{S} = [-1,1]^2$ respectively, and the support of conditional density $p(t|\boldsymbol{s})$ for any $\boldsymbol{s} \in \mathcal{S}$ is much narrower than $\mathcal{T}$.

- **Nonlinear Confounding Model:** We consider the following nonlinear effect model with

$$\begin{gathered} Y = T^2 + T + 10Z + \epsilon, \quad T = \cos\left(\pi Z^3\right) + \frac{Z}{4} + E, \quad Z = 4S_1 + S_2, \\ \text{and} \quad \boldsymbol{S} = (S_1, S_2) \sim \text{Uniform}[-1,1]^2 \subset \mathbb{R}^2, \end{gathered} \tag{20}$$

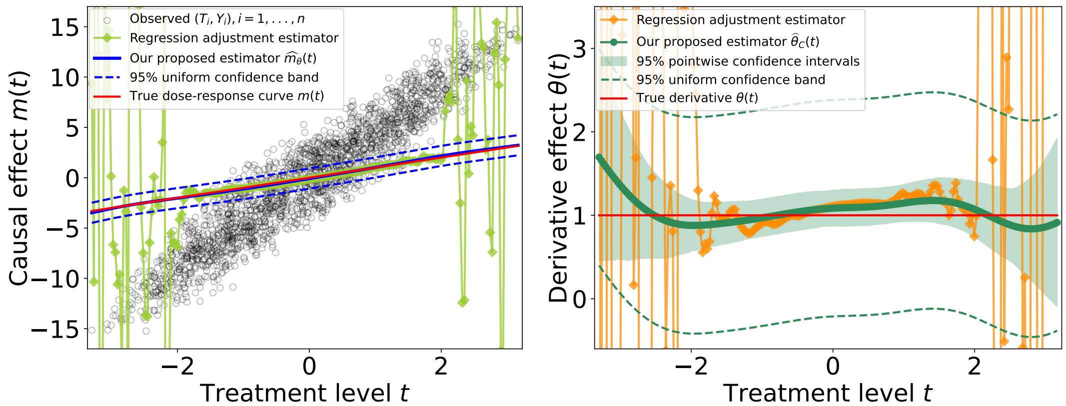

Fig 2: Simulation results under the linear confounding model (19). **Left:** The estimated dose-response curves by usual regression adjustment and our proposed estimators overlain with the true one $m(t)$. **Right:** The estimated derivatives of the dose-response curve by usual regression adjustment and our proposed estimators overlain with the true one $\theta(t)$. The middle and right panels also present the 95% confidence intervals and/or uniform confidence bands from our proposed estimators as shaded regions and dashed lines, respectively.

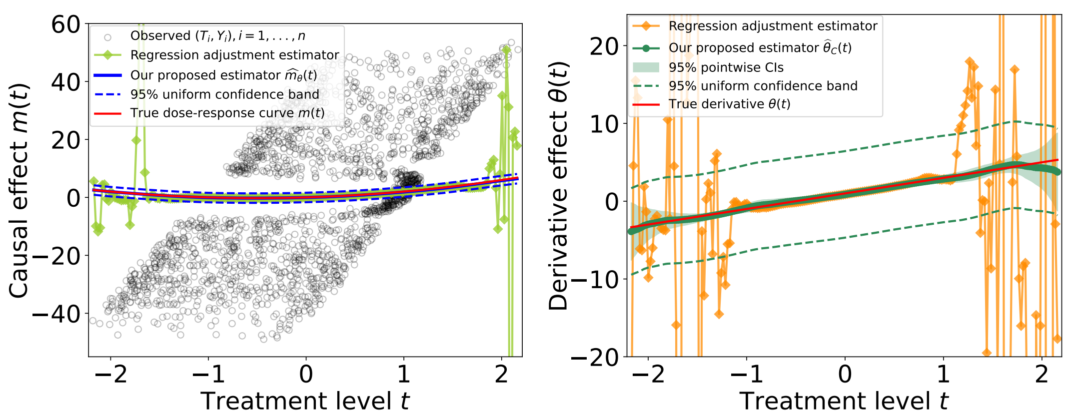

Fig 3: Simulation results under the nonlinear confounding model (20). See Fig 2 for the description of each panel.

where $E \sim \text{Uniform}[-0.1, 0.1]$ and $\epsilon \sim \mathcal{N}(0,1)$. Again, due to the nonlinear confounding effect $\cos(\pi Z)$ and small treatment variation E, the positivity condition (A0) fails on the generated data from (20). We also note that the debiased estimator of $m(t)$ proposed by Takatsu and Westling (2022) (and similarly, Kennedy et al. 2017) is incapable of recovering $m(t)$ under this data model, because some of their pseudo-outcomes have invalid values caused by nearly zero estimated conditional densities of $p(t|\boldsymbol{s})$.

Because of the presence of confounding variables, simply regressing Y_i's to T_i's in the generated data from each of the above models will give rise to biased estimates of $m(t)$. We apply both the proposed integral estimator $\widehat{m}_\theta(t)$ and localized derivative estimator $\widehat{\theta}_C(t)$ with bootstrap inferences in Section 3.3 to the generated data respectively. For comparisons, we also implement the usual regression adjustment estimators using the (partial) local quadratic regression (10) under the same choices of bandwidth parameters. The results are shown in Fig 1, Fig 2, and Fig 3. Note that the regression adjustment estimators are extremely unstable, especially around the boundary of the marginal support $\mathcal{T}$. Due to the failure of (A0), fine-tuning the bandwidth parameters h, b will not ameliorate the stability and consistency of regression adjustment estimators. On the contrary, our proposed estimators can recover the true dose-response curves and its derivatives with relatively narrow pointwise confidence intervals and uniform bands.

5.3. *Case Study: Effect of PM$_{2.5}$ on Cardiovascular Mortality Rate.* Air pollution, especially fine particulate matter with diameters less than 2.5 µm (PM$_{2.5}$), is known to contribute to an increase in cardiovascular diseases (Brook et al., 2010; Krittanawong et al., 2023). Some recent studies also identify a positive association between the PM$_{2.5}$ level (µg$/m^3$) and the county-level cardiovascular mortality rate (CMR; deaths/100,000 person-years) in the United States after controlling for socioeconomic factors (Wyatt et al., 2020a).

To showcase the applicability of our proposed integral and its derivative estimators, we apply them to analyzing the PM$_{2.5}$ and CMR data in Wyatt et al. (2020b). The data contain average annual CMR for outcome Y and PM$_{2.5}$ concentration for treatment/exposure T from 1990 to 2010 within $n = 2132$ counties, which were obtained from the US National Center for Health Statistics and Community Multiscale Air Quality modeling system, respectively. Our covariate vector $\boldsymbol{S} \in \mathbb{R}^{10}$ comprises two parts of confounding variables. The first part consists of two spatial confounding variables, latitude and longitude of each county, that helps incorporate the spatial dependence. The second part includes eight county-level socioeconomic factors acquired from the US census: population in 2000, civilian unemployment rate, median household income, percentage of female households with no spouse, percentage of vacant housing units, percentage of owner occupied housing units, percentage of high school graduates or above, and percentage of households below poverty. For each county in a given year, we use the closest records from a US census year (1990, 2000, or 2010) as the values of its socioeconomic variables and then average the values of $Y, T, \boldsymbol{S}$ for each county over these 21 years as the final data. Finally, similar to Takatsu and Westling (2022), we focus on the values of PM$_{2.5}$ between 2.5 µg$/m^3$ and 11.5 µg$/m^3$ to avoid boundary effects.

We apply our proposed integral estimator (8) and localized derivative estimator (12) to the final data with the choices of bandwidth parameters in Section 5.1. To take the magnitude differences of treatment T and covariates in $\boldsymbol{S}$ into consideration, we multiply the (coordinatewise) standard deviations of T and $\boldsymbol{S}$ on the data to $h_{ROT}, \boldsymbol{b}_{ROT}$ respectively and set $C_h = 7, C_b = 20$ to smooth out the estimated derivatives. To study the effects of confounding, we also consider regressing Y on T only via local quadratic regression estimators as well as fitting Y on T and spatial locations only via our proposed estimators. The results in the left panel of Fig 4 demonstrate that the effects of confounding. Before controlling any confounding variables, the fitted curve is not monotonic and peaks at around 8 µg$/m^3$. After controlling for spatial locations, it becomes flatter but still decreases when PM$_{2.5}$ level is above 8 µg$/m^3$. Only when we incorporate both spatial and socioeconomic covariates can the estimated relationship between PM$_{2.5}$ and CMR becomes monotonically increasing. The estimated changing rate of CMR with respect to PM$_{2.5}$ in the right panel of Fig 4 reveals the same conclusion from a different angle. More interestingly, after we adjust for all available confounding variables, the changing rate of CMR and its 95% confidence intervals flatten out and are unanimously above 0 when the PM$_{2.5}$ level is below 9 µg$/m^3$. This indicates a strong signal of a positive association between the PM$_{2.5}$ level and CMR. However, when the PM$_{2.5}$ level is above 10 µg$/m^3$, it is unclear whether PM$_{2.5}$ still positively contributes to cardiovascular mortality, possibly due to other competing risks (Leiser et al., 2019). Our findings generally align with the conclusions in Wyatt et al. (2020a); Takatsu and Westling (2022), but the superiority of our proposed estimators comes from two aspects. First, we estimate not only the relationship between PM$_{2.5}$ and CMR but also its derivative function, providing new

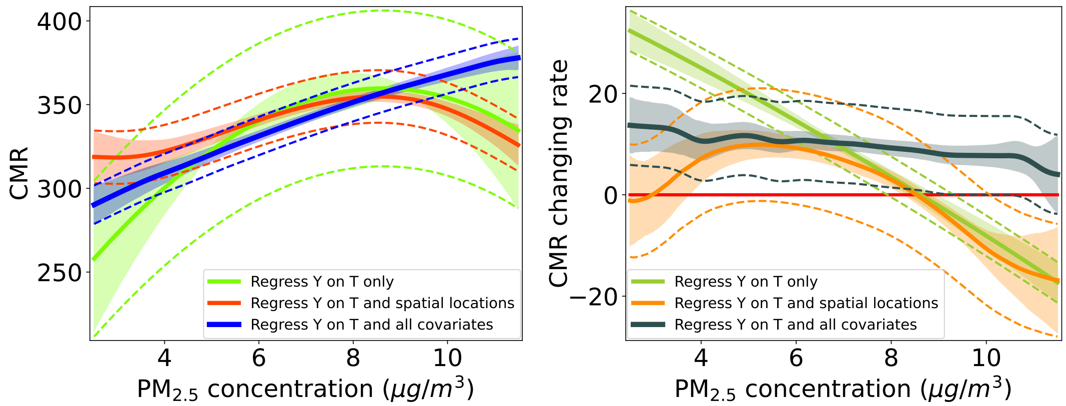

Fig 4: Estimated relationships between the $\mathrm{PM}_{2.5}$ concentration and CMR or its changing rate at the county level. **Left:** The estimated CMR with respect to the $\mathrm{PM}_{2.5}$ concentration. **Right:** The estimated changing rates of CMR with respect to the $\mathrm{PM}_{2.5}$ concentration. We also present the 95% confidence intervals and uniform confidence bands as shaded regions and dashed lines respectively for each regression scenario.

insights into their association. Second, our proposed estimators are capable of resolving the potential violation of the positivity condition (A0) and thus lead to convincing conclusions in this observational study.

6. Discussion. In summary, this paper studies nonparametric inference on the dose-response curve and its derivative function via innovative integral and localized derivative estimators. The main advantage of our proposed estimators is that they can consistently recover and infer the dose-response curve and its derivative without assuming the unrealistically strong positivity condition (A0) under the context of continuous treatments. We establish the consistency and bootstrap validity of the proposed estimators when nuisance functions are estimated by kernel smoothing methods but without requiring explicit undersmoothing on bandwidth parameters. Simulation studies and real-world applications demonstrate the effectiveness of our proposed estimators in addressing the failure of (A0). There are several future directions that can further advance the impacts of our proposed estimators.

1. Estimation of nuisance parameters/functions: The proposed integral estimator (8) and localized derivative estimator (12) require us to estimate two nuisance functions, the partial derivative $\frac{\partial}{\partial t}\mu(t, \boldsymbol{s})$ and the conditional CDF $P(\boldsymbol{s}|t)$. When we focus on the kernel smoothing methods, the finite-sample performances of our integral estimator (8) and localized derivative estimator (12) depend on the choices of three bandwidth parameters $h, b, \hbar > 0$. We only consider the rule-of-thumb and normal reference bandwidth selection methods in this paper, but it would be of research interest to study how the bandwidth choices can be improved through the plug-in rule (Ruppert et al., 1995) or cross-validation (Li and Racine, 2004). More broadly, other estimators, such as regression splines for $\frac{\partial}{\partial t}\mu(t, \boldsymbol{s})$ (Friedman, 1991; Zhou and Wolfe, 2000) and nearest neighbor-type or local logistic approaches for $P(\boldsymbol{s}|t)$ (Stute, 1986; Hall et al., 1999), are also be applied to estimating the nuisance functions and may lead to better performances of our proposed estimators $\widehat{\theta}_C(t)$ and $\widehat{m}_\theta(t)$.

2. IPTW and doubly robust estimators: Our proposed estimators (8) and (12) are based on the idea of regression adjustment. Naturally, it would also be interesting to investigate how to leverage our integral and localized techniques to resolve the positivity requirement in the existing generalized propensity score-based approaches (Hirano and Imbens, 2004; Imai and van Dyk, 2004) and doubly robust methods (Kennedy et al., 2017; Westling et al., 2020;

Colangelo and Lee, 2020; Semenova and Chernozhukov, 2021; Bonvini and Kennedy, 2022) for estimating the dose-response curve and its derivative function.

3. Additive model diagnostics: Assumption A2 is vital to the consistency of our proposed estimators, and it is unclear if this assumption holds beyond the additive confounding model (3). However, it is possible to utilize our estimators to test the correctness of the additive structure in (3). Under the positivity condition (A0), one can test (3) by estimating the absolute difference of $\left|\mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,\boldsymbol{S})\right] - \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,\boldsymbol{S})\middle|T=t\right]\right|$ from 0 via the regression adjustment estimator $\widehat{\theta}_{RA}(t)$ in Remark 1 and our estimator $\widehat{\theta}_C(t)$ in (12). Other testing procedures, such as the marginal integration regression method (Linton and Nielsen, 1995), are also applicable. When the positivity condition (A0) is violated, the partial derivative $\beta_2(t,\boldsymbol{s}) = \frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ is still independent of $\boldsymbol{s}$ for any $t \in \mathcal{T}$ under (3), and our estimator $\widehat{\beta}_2(t,\boldsymbol{s})$ in (10) should not depend on $\boldsymbol{s}$ asymptotically. It suggests that statistical quantities, such as $\Theta(t) = \sup_{\boldsymbol{s}\in\mathcal{S}(t)}\left|\widehat{\beta}_2(t,\boldsymbol{s}) - \widehat{\theta}_C(t)\right|$ and $\Theta = \int_{\mathcal{T}}\Theta(t)\,dt$, should converge to 0 as $n \to \infty$. Thus, one may derive the limiting distribution of $\Theta(t)$ or Θ to carry out a procedure for statistically testing the additive confounding model (3) without assuming (A0).

4. Violation of ignorability (Assumption A1(b)): In observational studies, there could be some unmeasured confounding variables that go beyond our covariate vector $\boldsymbol{S} \in \mathcal{S} \subset \mathbb{R}^d$ and bias our inference (VanderWeele, 2008). Hence, it is important to analyze the sensitivities of the dose-response curve and its derivative with respect to the violation of ignorability. In this scenario, the instrumental variable model (Kilbertus et al., 2020) and the Riesz-Frechet representation technique (Chernozhukov et al., 2022) might be useful for the analysis. As an alternative, one can also incorporate an additional high-dimensional covariate vector $\boldsymbol{Z} \in \mathbb{R}^{d'}$ into our model (2) to avoid unmeasured confounding variables. However, since nonparametric estimations on $m(t)$ and $\theta(t)$ have slow rates of convergence when d' is large, one may consider imposing a linear form on $\boldsymbol{Z}$ so that

$$\mathbb{E}\left(Y|T=t,\boldsymbol{S}=\boldsymbol{s},\boldsymbol{Z}=\boldsymbol{z}\right) = m(t) + \eta(\boldsymbol{s}) + \vartheta^T\boldsymbol{z}$$

or utilize some orthogonalization techniques in Section 4.3 of Rothenhäusler and Yu (2019) to conduct valid inference. We will leave them as future works.

Acknowledgements. YZ is supported in part by YC's NSF grant DMS-2141808 and AG's NSF grant DMS-2310578. YC is supported by NSF grants DMS-1952781, 2112907, 2141808, and NIH U24-AG07212. AG is supported by NSF grant DMS-2310578.

REFERENCES

ARCONES, M. A. and GINE, E. (1992). On the Bootstrap of U and V Statistics. *The Annals of Statistics* **20** 655 – 674.

BAHADUR, R. R. (1966). A note on quantiles in large samples. *The Annals of Mathematical Statistics* **37** 577–580.

BASHTANNYK, D. M. and HYNDMAN, R. J. (2001). Bandwidth selection for kernel conditional density estimation. *Computational Statistics & Data Analysis* **36** 279–298.

BJERVE, S., DOKSUM, K. A. and YANDELL, B. S. (1985). Uniform confidence bounds for regression based on a simple moving average. *Scandinavian Journal of Statistics* 159–169.

BONVINI, M. and KENNEDY, E. H. (2022). Fast convergence rates for dose-response estimation. *arXiv preprint arXiv:2207.11825*.

BROOK, R. D., RAJAGOPALAN, S., POPE, C. A., BROOK, J. R., BHATNAGAR, A., DIEZ-ROUX, A. V., HOLGUIN, F., HONG, Y., LUEPKER, R. V., MITTLEMAN, M. A., PETERS, A., SISCOVICK, D., SMITH, S. C., WHITSEL, L. and KAUFMAN, J. D. (2010). Particulate matter air pollution and cardiovascular disease: an update to the scientific statement from the American Heart Association. *Circulation* **121** 2331–2378.

BUJA, A., HASTIE, T. and TIBSHIRANI, R. (1989). Linear Smoothers and Additive Models. *The Annals of Statistics* **17** 453 – 510.

BUSSO, M., DINARDO, J. and MCCRARY, J. (2014). New evidence on the finite sample properties of propensity score reweighting and matching estimators. *Review of Economics and Statistics* **96** 885–897.

CALONICO, S., CATTANEO, M. D. and FARRELL, M. H. (2018). On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association* **113** 767–779.

CATTANEO, M. D., CRUMP, R. K. and JANSSON, M. (2010). Robust data-driven inference for density-weighted average derivatives. *Journal of the American Statistical Association* **105** 1070–1083.

CHACÓN, J. E., DUONG, T. and WAND, M. (2011). Asymptotics for general multivariate kernel density derivative estimators. *Statistica Sinica* 807–840.

CHAMBERLAIN, G. (1986). Asymptotic efficiency in semi-parametric models with censoring. *Journal of Econometrics* **32** 189–218.

CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2015). Asymptotic theory for density ridges. *The Annals of Statistics* **43** 1896–1928.

CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2016). A comprehensive approach to mode clustering. *Electronic Journal of Statistics* **10** 210 – 241.

CHEN, Y.-C., GENOVESE, C. R. and WASSERMAN, L. (2017). Density level sets: Asymptotics, inference, and visualization. *Journal of the American Statistical Association* **112** 1684–1696.

CHENG, G. and CHEN, Y.-C. (2019). Nonparametric inference via bootstrapping the debiased estimator. *Electronic Journal of Statistics* **13** 2194 – 2256.

CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2014). Gaussian approximation of suprema of empirical processes. *The Annals of Statistics* **42** 1564–1597.

CHERNOZHUKOV, V., CINELLI, C., NEWEY, W., SHARMA, A. and SYRGKANIS, V. (2022). Long story short: Omitted variable bias in causal machine learning Technical Report, National Bureau of Economic Research.

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

COLE, S. R. and HERNÁN, M. A. (2008). Constructing inverse probability weights for marginal structural models. *American Journal of Epidemiology* **168** 656–664.

COX, D. R. (1958). *Planning of Experiments*. Wiley.

CRUMP, R. K., HOTZ, V. J., IMBENS, G. W. and MITNIK, O. A. (2009). Dealing with limited overlap in estimation of average treatment effects. *Biometrika* **96** 187–199.

DEHEJIA, R. H. and WAHBA, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American statistical Association* **94** 1053–1062.

DÍAZ, I. and VAN DER LAAN, M. J. (2013). Targeted data adaptive estimation of the causal dose–response curve. *Journal of Causal Inference* **1** 171–192.

DVORETZKY, A., KIEFER, J. and WOLFOWITZ, J. (1956). Asymptotic Minimax Character of the Sample Distribution Function and of the Classical Multinomial Estimator. *The Annals of Mathematical Statistics* **27** 642 – 669.

D'AMOUR, A., DING, P., FELLER, A., LEI, L. and SEKHON, J. (2021). Overlap in observational studies with high-dimensional covariates. *Journal of Econometrics* **221** 644–654.

EFRON, B. (1979). Bootstrap Methods: Another Look at the Jackknife. *The Annals of Statistics* **7** 1 – 26.

EINMAHL, U. and MASON, D. M. (2005). Uniform in bandwidth consistency of kernel-type function estimators. *The Annals of Statistics* **33** 1380 – 1403.

FAN, J. and GIJBELS, I. (1996). *Local polynomial modelling and its applications* **66**. Chapman & Hall/CRC.

FAN, Y. and GUERRE, E. (2015). Multivariate Local Polynomial Estimators: Boundary Properties and Uniform Asymptotic Linear Representation. *Advances in Econometrics*.

FAN, J., GIJBELS, I., HU, T.-C. and HUANG, L.-S. (1996). A study of variable bandwidth selection for local polynomial regression. *Statistica Sinica* **6** 113–127.

FLORES, C. (2007). Estimation of dose-response functions and optimal doses with a continuous treatment Technical Report, Department of Economics, University of Miami.

FREEDMAN, D. A. (1981). Bootstrapping regression models. *The Annals of Statistics* **9** 1218–1228.

FRIEDMAN, J. H. (1991). Multivariate adaptive regression splines. *The Annals of Statistics* **19** 1–67.

GARCÍA PORTUGUÉS, E. (2023). Notes for nonparametric statistics. (Accessed on March 26, 2024).

GIESSING, A. (2023). Gaussian and Bootstrap Approximations for Suprema of Empirical Processes. *arXiv preprint arXiv:2309.01307*.

GILBERT, B., DATTA, A., CASEY, J. A. and OGBURN, E. L. (2023). A causal inference framework for spatial confounding. *arXiv preprint arXiv:2112.14946*.

GILL, R. D. and ROBINS, J. M. (2001). Causal inference for complex longitudinal data: the continuous case. *Annals of Statistics* **29** 1785–1811.

GINÉ, E. and GUILLOU, A. (2002). Rates of strong uniform consistency for multivariate kernel density estimators. *Annales de l'Institut Henri Poincare (B) Probability and Statistics* **38** 907–921.

HALL, P. (1992). On bootstrap confidence intervals in nonparametric regression. *The Annals of Statistics* **20** 695–711.

HALL, P., WOLFF, R. C. and YAO, Q. (1999). Methods for estimating a conditional distribution function. *Journal of the American Statistical Association* **94** 154–163.

HÄRDLE, W. and STOKER, T. M. (1989). Investigating smooth multiple regression by the method of average derivatives. *Journal of the American statistical Association* **84** 986–995.

HASTIE, T. and LOADER, C. (1993). Local regression: Automatic kernel carpentry. *Statistical Science* 120–129.

HINES, O., DIAZ-ORDAZ, K. and VANSTEELANDT, S. (2023). Optimally weighted average derivative effects. *arXiv preprint arXiv:2308.05456*.

HIRANO, K. and IMBENS, G. W. (2004). The propensity score with continuous treatments. *Applied Bayesian modeling and causal inference from incomplete-data perspectives* **226164** 73–84.

HIRSHBERG, D. A. and WAGER, S. (2020). Debiased inference of average partial effects in single-index models: Comment on wooldridge and zhu. *Journal of Business & Economic Statistics* **38** 19–24.

HOLMES, M. P., GRAY, A. G. and ISBELL, C. L. (2012). Fast nonparametric conditional density estimation. *arXiv preprint arXiv:1206.5278*.

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

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

KAMMANN, E. and WAND, M. P. (2003). Geoadditive models. *Journal of the Royal Statistical Society Series C: Applied Statistics* **52** 1–18.

KELLER, J. P. and SZPIRO, A. A. (2020). Selecting a scale for spatial confounding adjustment. *Journal of the Royal Statistical Society Series A: Statistics in Society* **183** 1121–1143.

KENNEDY, E. H. (2019). Nonparametric causal effects based on incremental propensity score interventions. *Journal of the American Statistical Association* **114** 645–656.

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

KHAN, S. and TAMER, E. (2010). Irregular identification, support conditions, and inverse weight estimation. *Econometrica* **78** 2021–2042.

KILBERTUS, N., KUSNER, M. J. and SILVA, R. (2020). A class of algorithms for general instrumental variable models. *Advances in Neural Information Processing Systems* **33** 20108–20119.

KONG, E., LINTON, O. and XIA, Y. (2010). Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory* **26** 1529–1564.

KRITTANAWONG, C., QADEER, Y. K., HAYES, R. B., WANG, Z., THURSTON, G. D., VIRANI, S. and LAVIE, C. J. (2023). PM2.5 and cardiovascular diseases: State-of-the-Art review. *International Journal of Cardiology Cardiovascular Risk and Prevention* **19** 200217.

LÉGER, M., CHATTON, A., LE BORGNE, F., PIRRACCHIO, R., LASOCKI, S. and FOUCHER, Y. (2022). Causal inference in case of near-violation of positivity: comparison of methods. *Biometrical Journal* **64** 1389–1403.

LEISER, C. L., SMITH, K. R., VANDERSLICE, J. A., GLOTZBACH, J. P., FARRELL, T. W. and HANSON, H. A. (2019). Evaluation of the sex-and-age-specific effects of PM2. 5 on hospital readmission in the presence of the competing risk of mortality in the medicare population of Utah 1999–2009. *Journal of Clinical Medicine* **8** 2114.

LI, Q. and RACINE, J. (2004). Cross-validated local linear nonparametric regression. *Statistica Sinica* 485–512.

LINTON, O. and NIELSEN, J. P. (1995). A Kernel Method of Estimating Structured Nonparametric Regression Based on Marginal Integration. *Biometrika* **82** 93–100.

LU, Z.-Q. (1996). Multivariate locally weighted polynomial fitting and partial derivative estimation. *Journal of Multivariate Analysis* **59** 187–205.

MA, X. and WANG, J. (2020). Robust inference using inverse probability weighting. *Journal of the American Statistical Association* **115** 1851–1860.

MANSKI, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review* **80** 319–323.

MASSART, P. (1990). The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality. *The Annals of Probability* **18** 1269–1283.

NEUGEBAUER, R. and VAN DER LAAN, M. (2007). Nonparametric causal effects based on marginal structural models. *Journal of Statistical Planning and Inference* **137** 419–434.

NEWEY, W. K. (1994). Kernel estimation of partial means and a general variance estimator. *Econometric Theory* **10** 1–21.

NEWEY, W. K. and STOKER, T. M. (1993). Efficiency of weighted average derivative estimators and index models. *Econometrica* **61** 1199–1223.

NOLAN, D. and POLLARD, D. (1987). U-Processes: Rates of Convergence. *The Annals of Statistics* **15** 780 – 799.

PACIOREK, C. J. (2010). The importance of scale for spatial-confounding bias and precision of spatial regression estimators. *Statistical Science* **25** 107-125.

PETERSEN, M. L., PORTER, K. E., GRUBER, S., WANG, Y. and VAN DER LAAN, M. J. (2012). Diagnosing and responding to violations in the positivity assumption. *Statistical methods in medical research* **21** 31–54.

POWELL, J. L., STOCK, J. H. and STOKER, T. M. (1989). Semiparametric estimation of index coefficients. *Econometrica* 1403–1430.

ROBINS, J. (1986). A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect. *Mathematical modelling* **7** 1393–1512.

ROBINS, J. M., HERNAN, M. A. and BRUMBACK, B. (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology* **11** 550–560.

ROTHE, C. (2017). Robust confidence intervals for average treatment effects under limited overlap. *Econometrica* **85** 645–660.

ROTHENHÄUSLER, D. and YU, B. (2019). Incremental causal effects. *arXiv preprint arXiv:1907.13258*.

RUBIN, D. B. (1974). Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology* **66** 688-701.

RUBIN, D. B. (1980). Randomization analysis of experimental data: The Fisher randomization test comment. *Journal of the American Statistical Association* **75** 591–593.

RUPPERT, D., SHEATHER, S. J. and WAND, M. P. (1995). An effective bandwidth selector for local least squares regression. *Journal of the American Statistical Association* **90** 1257–1270.

RUPPERT, D. and WAND, M. P. (1994). Multivariate locally weighted least squares regression. *The Annals of Statistics* **22** 1346–1370.

SCHINDLER, A. (2011). Bandwidth selection in nonparametric kernel estimation, PhD thesis, Göttingen, Georg-August Universität, Diss.

SCHNELL, P. and PAPADOGEORGOU, G. (2020). Mitigating unobserved spatial confounding when estimating the effect of supermarket access on cardiovascular disease deaths. *Annals of Applied Statistics* **14** 2069-2095.

SEMENOVA, V. and CHERNOZHUKOV, V. (2021). Debiased machine learning of conditional average treatment effects and other causal functions. *The Econometrics Journal* **24** 264–289.

SHAO, J. (2003). *Mathematical Statistics*. Springer Science & Business Media.

SHEATHER, S. J. (2004). Density estimation. *Statistical Science* **19** 588–597.

SHIEH, G. S. (2014). U-and V-statistics. *Wiley StatsRef: Statistics Reference Online*.

STONE, C. J. (1980). Optimal rates of convergence for nonparametric estimators. *The Annals of Statistics* **8** 1348–1360.

STONE, C. J. (1982). Optimal global rates of convergence for nonparametric regression. *The Annals of Statistics* **10** 1040–1053.

STUTE, W. (1986). Conditional empirical processes. *The Annals of Statistics* **14** 638–647.

TAKATSU, K. and WESTLING, T. (2022). Debiased inference for a covariate-adjusted regression function. *arXiv preprint arXiv:2210.06448*.

TANG, Z. and WESTLING, T. (2024). Consistency of the bootstrap for asymptotically linear estimators based on machine learning. *arXiv preprint arXiv:2404.03064*.

THADEN, H. and KNEIB, T. (2018). Structural equation models for dealing with spatial confounding. *The American Statistician* **72** 239–252.

VAN DER LAAN, M. J. and ROBINS, J. M. (2003). *Unified methods for censored longitudinal data and causality*. Springer.

VAN DER VAART, A. (1991). On differentiable functionals. *The Annals of Statistics* **19** 178–204.

VAN DER VAART, A. W. (1998). *Asymptotic Statistics*. *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press.

VAN DER VAART, A. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. *Springer Series in Statistics*. Springer.

VANDERWEELE, T. J. (2008). The sign of the bias of unmeasured confounding. *Biometrics* **64** 702–706.

WAND, M. P. and JONES, M. C. (1994). *Kernel smoothing*. CRC press.

WASSERMAN, L. (2006). *All of Nonparametric Statistics*. Springer Science & Business Media.

WAUD, D. R. (1975). Analysis of dose-response curves. In *Smooth muscle* (E. E. Daniel and D. M. Paton, eds.) 471–506. Springer US, Boston, MA.

WESTLING, T., GILBERT, P. and CARONE, M. (2020). Causal isotonic regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **82** 719–747.

WESTREICH, D. and COLE, S. R. (2010). Invited commentary: positivity in practice. *American Journal of Epidemiology* **171** 674–677.

WIECHA, N. and REICH, B. J. (2024). Two-stage Spatial Regression Models for Spatial Confounding. *arXiv preprint arXiv:2404.09358*.

WU, C.-F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *The Annals of Statistics* **14** 1261–1295.

WYATT, L. H., PETERSON, G. C., WADE, T. J., NEAS, L. M. and RAPPOLD, A. G. (2020a). The contribution of improved air quality to reduced cardiovascular mortality: Declines in socioeconomic differences over time. *Environment international* **136** 105430.

WYATT, L. H., PETERSON, G. C. L., WADE, T. J., NEAS, L. M. and RAPPOLD, A. G. (2020b). Annual PM2. 5 and cardiovascular mortality rate data: Trends modified by county socioeconomic status in 2,132 US counties. *Data in Brief* **30** 105318.

YANG, S. and DING, P. (2018). Asymptotic inference of causal effects with observational studies trimmed by the estimated propensity scores. *Biometrika* **105** 487–493.

YANG, L. and TSCHERNIG, R. (1999). Multivariate bandwidth selection for local linear regression. *Journal of the Royal Statistical Society Series B: Statistical Methodology* **61** 793–815.

ZHOU, S. and WOLFE, D. A. (2000). On derivative estimation in spline regression. *Statistica Sinica* **10** 93–108.

SUPPLEMENT TO "NONPARAMETRIC INFERENCE ON DOSE-RESPONSE CURVES WITHOUT THE POSITIVITY CONDITION"

CONTENTS

The supplementary material contains the proofs of main results in the paper and other auxiliary technical results.

APPENDIX A: NONPARAMETRIC BOUNDS UNDER ZERO TREATMENT VARIATIONS

In this section, we study nonparametric bounds on the dose-response curve $m(t)$ and its derivative function $\theta(t)$ under the additive confounding model (3) when the treatment variation E has zero variance. We will leave the derivations of nonparametric bounds under the general model (2) as a future direction.

A.1. Nonparametric Bound on $m(t)$. Recall from Section 2.2 that when $\mathrm{Var}(E) = 0$, the conditional mean outcome function $\mu(t, \boldsymbol{s})$ can only be identified on the lower dimensional surface $\{(t, \boldsymbol{s}) \in \mathcal{T} \times \mathcal{S} : t = f(\boldsymbol{s})\}$ so that

$$\mu\left(f(\boldsymbol{s}), \boldsymbol{s}\right) = \bar{m}\left(f(\boldsymbol{s})\right) + \eta(\boldsymbol{s}) = m\left(f(\boldsymbol{s})\right) + \eta(\boldsymbol{s}). \tag{21}$$

Given that $\mathrm{Var}(E) = 0$, we can always fit the relation $T = f(\boldsymbol{S})$ from the observed data. Hence, we assume that the function f is known for any $\boldsymbol{s} \in \mathcal{S}$ in the sequel. To obtain a nonparametric bound for all the possible values that $m(t)$ can take, we impose the following assumption on the magnitude of the random effect function $\eta : \mathcal{S} \to \mathbb{R}$ in (3); see Manski (1990) for related discussions.

ASSUMPTION A7 (Bounded random effect). Let $L_f(t) = \{\boldsymbol{s} \in \mathcal{S} : f(\boldsymbol{s}) = t\}$ be a level set of the function $f : \mathcal{S} \to \mathbb{R}$ at $t \in \mathcal{T}$. There exists a constant $\rho_1 > 0$ such that

$$\rho_1 \geq \max\left\{\sup_{t\in\mathcal{T}} \sup_{\boldsymbol{s}\in L_f(t)} |\eta(\boldsymbol{s})|, \; \frac{\sup_{t\in\mathcal{T}} \sup_{\boldsymbol{s}\in L_f(t)} \mu(f(\boldsymbol{s}), \boldsymbol{s}) - \inf_{t\in\mathcal{T}} \inf_{\boldsymbol{s}\in L_f(t)} \mu(f(\boldsymbol{s}), \boldsymbol{s})}{2}\right\}.$$

By (21) and the first lower on $\rho_1 \geq \sup_{t\in\mathcal{T}} \sup_{\boldsymbol{s}\in L_f(t)} |\eta(\boldsymbol{s})|$ in Assumption A7, we know that

$$|\mu(f(\boldsymbol{s}),\boldsymbol{s}) - m(t)| = |\eta(\boldsymbol{s})| \leq \rho_1$$

for any $\boldsymbol{s} \in L_f(t)$. It also implies that

$$\begin{aligned} m(t) &\in \bigcap_{\boldsymbol{s}\in L_f(t)} \left[\mu(f(\boldsymbol{s}),\boldsymbol{s}) - \rho_1,\, \mu(f(\boldsymbol{s}),\boldsymbol{s}) + \rho_1\right] \\ &= \left[\sup_{\boldsymbol{s}\in L_f(t)} \mu(f(\boldsymbol{s}),\boldsymbol{s}) - \rho_1,\, \inf_{\boldsymbol{s}\in L_f(t)} \mu(f(\boldsymbol{s}),\boldsymbol{s}) + \rho_1\right], \end{aligned} \tag{22}$$

which is the nonparametric bound on $m(t)$ that contains all the possible values of $m(t)$ for any fixed $t \in \mathcal{T}$ when $\mathrm{Var}(E) = 0$. Notice that this bound (22) is well-defined and nonempty under the second lower bound on ρ_1 in Assumption A7.

A.2. Nonparametric Bound on $\boldsymbol{\theta(t)}$. We already know that $\mu(f(\boldsymbol{s}),\boldsymbol{s})$ and $f(\boldsymbol{s})$ can be identified even when $\mathrm{Var}(E) = 0$, so their derivatives with respect to $\boldsymbol{s}$ are also identifiable. Hence, we assume that the functions $\nabla_{\boldsymbol{s}}\mu(f(\boldsymbol{s}),\boldsymbol{s})$ and $\nabla f(\boldsymbol{s})$ are known for any $\boldsymbol{s} \in \mathcal{S}$ in the sequel. The nonparametric bound on the dose-response curve $m(t)$ involves an upper bound on the random effect function $\eta : \mathcal{S} \to \mathbb{R}$ (recall Assumption A7), while the nonparametric bound on the derivative function $\theta(t)$ requires some constraints on the gradient $\nabla\eta(\boldsymbol{s})$ of the random effect function as follows.

ASSUMPTION A8 (Constraints on the random effect gradient). Let $L_f(t) = \{\boldsymbol{s} \in \mathcal{S} : f(\boldsymbol{s}) = t\}$ be a level set of the function $f : \mathcal{S} \to \mathbb{R}$ at $t \in \mathcal{T}$. There exist constants $\rho_2, \rho_3 > 0$ such that

$$\rho_3 \leq \inf_{t\in\mathcal{T}} \inf_{\boldsymbol{s}\in L_f(t)} ||\nabla\eta(\boldsymbol{s})||_{\min} \leq \sup_{t\in\mathcal{T}} \sup_{\boldsymbol{s}\in L_f(t)} ||\nabla\eta(\boldsymbol{s})||_{\max} \leq \rho_2$$

and

$$\sup_{t\in\mathcal{T}} \sup_{\boldsymbol{s}\in L_f(t)} \max_{j=1,\dots,d} \left[\frac{v_j(\boldsymbol{s}) - \mathrm{sign}(g_j(\boldsymbol{s}))\cdot\rho_2}{g_j(\boldsymbol{s})}\right] \leq \inf_{t\in\mathcal{T}} \inf_{\boldsymbol{s}\in L_f(t)} \min_{j=1,\dots,d} \left[\frac{v_j(\boldsymbol{s}) + \mathrm{sign}(g_j(\boldsymbol{s}))\cdot\rho_2}{g_j(\boldsymbol{s})}\right],$$

where $||\nabla\eta(\boldsymbol{s})||_{\min} = \min_{j=1,\dots,d}\left|\frac{\partial}{\partial s_j}\eta(\boldsymbol{s})\right|$, $||\nabla\eta(\boldsymbol{s})||_{\max} = \max_{j=1,\dots,d}\left|\frac{\partial}{\partial s_j}\eta(\boldsymbol{s})\right|$, $v_j(\boldsymbol{s}) = \frac{\partial}{\partial s_j}\mu(f(\boldsymbol{s}),\boldsymbol{s})$, and $g_j(\boldsymbol{s}) = \frac{\partial}{\partial s_j} f(\boldsymbol{s})$ for $j = 1,\dots,d$.

The first inequality in Assumption A8 ensures that we can derive a finite and meaningful nonparametric bound on $\theta(t)$ for any $t \in \mathcal{T}$. In particular, a direct calculation shows that

$$\begin{aligned} \nabla_{\boldsymbol{s}}\mu(f(\boldsymbol{s}),\boldsymbol{s}) &= \nabla_{\boldsymbol{s}} m(f(\boldsymbol{s})) + \nabla\eta(\boldsymbol{s}) \\ &= m'(f(\boldsymbol{s}))\cdot\nabla f(\boldsymbol{s}) + \nabla\eta(\boldsymbol{s}) \\ &= \theta(t)\cdot\nabla f(\boldsymbol{s}) + \nabla\eta(\boldsymbol{s}) \end{aligned}$$

for all $\boldsymbol{s} \in L_f(t)$. This indicates that

$$\left|\frac{\partial}{\partial s_j}\mu(f(\boldsymbol{s}),\boldsymbol{s}) - \theta(t)\cdot\frac{\partial}{\partial s_j} f(\boldsymbol{s})\right| = |v_j(\boldsymbol{s}) - \theta(t)\cdot g_j(\boldsymbol{s})| = \left|\frac{\partial}{\partial s_j}\eta(\boldsymbol{s})\right| \leq \rho_2$$

for $j = 1,\dots,d$ and therefore, a nonparametric bound on $\theta(t)$ is given by

$$\theta(t) \in \bigcap_{j=1}^{d} \left[\frac{v_j(\boldsymbol{s}) - \mathrm{sign}(g_j(\boldsymbol{s}))\cdot\rho_2}{g_j(\boldsymbol{s})},\, \frac{v_j(\boldsymbol{s}) + \mathrm{sign}(g_j(\boldsymbol{s}))\cdot\rho_2}{g_j(\boldsymbol{s})}\right]. \tag{23}$$

This bound is well-defined and nonempty under the second inequality in Assumption A8. More importantly, the nonparametric bound (23) sheds light on the critical role of the covariate effect function f of $\boldsymbol{S}$ on treatment T in bounding the possible value of $\theta(t)$. Specifically, when the variation of f is higher (*i.e.*, $\nabla f(\boldsymbol{s})$ is larger in its magnitude), the nonparametric bound (23) will be tighter.

APPENDIX B: PROOFS

This section consists of the proofs for theorems, lemmas, and propositions in the main paper as well as other auxiliary results.

B.1. Proof of Proposition 1.

PROPOSITION 1 (Properties of the additive confounding model (3)). *Let* $\theta_M(t) = \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,\boldsymbol{S})\right]$ *and* $\theta_C(t) = \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,\boldsymbol{S})\middle|T=t\right]$, *where* $\mu(t,\boldsymbol{s}) = \mathbb{E}\left[Y|T=t,\boldsymbol{S}=\boldsymbol{s}\right]$. *Under the additive confounding model* (3) *with* $\mathbb{E}\left[\eta(\boldsymbol{S})\right]=0$, *the following results hold for all* $t\in\mathcal{T}$*:*

(a) $m(t) = \bar{m}(t)$.
(b) $m(t) \neq \mathbb{E}(Y|T=t) = \bar{m}(t) + \mathbb{E}\left[\eta(\boldsymbol{S})|T=t\right]$ *when* $\mathbb{E}\left[\eta(\boldsymbol{S})|T=t\right] \neq 0$.
(c) $\theta(t) = \theta_M(t) = \theta_C(t)$.
(d) $\theta(t) \neq \frac{d}{dt}\mathbb{E}\left[\mu(t,\boldsymbol{S})|T=t\right] = \theta(t) + \frac{d}{dt}\mathbb{E}\left[\eta(\boldsymbol{S})|T=t\right]$ *when* $\frac{d}{dt}\mathbb{E}\left[\eta(\boldsymbol{S})|T=t\right]\neq 0$.
(e) $\mathbb{E}\left[\mu(T,\boldsymbol{S})\right] = \mathbb{E}\left[m(T)\right]$ *even when* $\mathbb{E}\left[\eta(\boldsymbol{S})\right]\neq 0$.

The above results hold even if the treatment variation $E=0$ *almost surely.*

PROOF OF PROPOSITION 1. All the results follow from some simple calculations and the fact that

$$\mu(t,\boldsymbol{s}) = \mathbb{E}\left(Y|T=t,\boldsymbol{S}=\boldsymbol{s}\right) = \bar{m}(t) + \eta(\boldsymbol{s})$$

under the additive confounding model (3).

(a) Given that $\mathbb{E}\left[\eta(\boldsymbol{S})\right]=0$, we note that

$$m(t) = \mathbb{E}\left[\mu(t,\boldsymbol{S})\right] = \mathbb{E}\left[\mathbb{E}\left(Y|T=t,\boldsymbol{S}\right)\right] = \mathbb{E}\left[\bar{m}(t)+\eta(\boldsymbol{S})\right] = \bar{m}(t).$$

(b) By (a), we know that

$$\mathbb{E}\left(Y|T=t\right) = \bar{m}(t) + \mathbb{E}\left[\eta(\boldsymbol{S})|T=t\right] \neq \bar{m}(t) = m(t)$$

when $\mathbb{E}\left[\eta(\boldsymbol{S})|T=t\right]\neq 0$.

(c) First, we notice that

$$\theta_M(t) \equiv \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,\boldsymbol{S})\right] = \mathbb{E}\left[\bar{m}'(t)\right] = \bar{m}'(t) = m'(t) = \theta(t).$$

In addition, we also have that

$$\theta_C(t) \equiv \mathbb{E}\left[\frac{\partial}{\partial t}\mu(t,\boldsymbol{S})\middle|T=t\right] = \mathbb{E}\left[\bar{m}'(t)|T=t\right] = \bar{m}'(t) = \theta(t).$$

(d) From (c), we know that

$$\frac{d}{dt}\mathbb{E}\left[\mu(t,\boldsymbol{S})|T=t\right] = \frac{d}{dt}\left[\bar{m}(t) + \mathbb{E}\left(\eta(\boldsymbol{S})|T=t\right)\right] = \bar{m}'(t) + \frac{d}{dt}\mathbb{E}\left(\eta(\boldsymbol{S})|T=t\right) \neq \theta(t)$$

when $\frac{d}{dt}\mathbb{E}\left[\eta(\boldsymbol{S})|T=t\right]\neq 0$.

(e) We know that the left hand side is equal to

$$\mathbb{E}\left[\mu(T,\boldsymbol{S})\right]=\mathbb{E}\left[\bar{m}(T)+\eta(\boldsymbol{S})\right]=\mathbb{E}\left[\bar{m}(T)\right]+\mathbb{E}\left[\eta(\boldsymbol{S})\right],$$

while the right hand side is given by

$$\mathbb{E}\left[m(T)\right]=\mathbb{E}\left\{\bar{m}(T)+\mathbb{E}\left[\eta(\boldsymbol{S})\right]\right\}=\mathbb{E}\left[\bar{m}(T)\right]+\mathbb{E}\left[\eta(\boldsymbol{S})\right].$$

This result thus holds even when $\mathbb{E}\left[\eta(\boldsymbol{S})\right]\neq 0$.

Finally, when deriving (a)-(e), we do not impose any condition on the treatment variation E, the results hold no matter what distribution of E is. $\square$

B.2. Proof of Lemma 2. Before proving Lemma 2, we give a complete statement of the asymptotic expressions for the conditional variance and bias of $\widehat{\beta}_2(t,\boldsymbol{s})$ as follows.

LEMMA 2 (Pointwise convergence of $\widehat{\beta}_2(t,\boldsymbol{s})$). *Suppose that Assumptions A3, A4, A5, and A6(a) hold under the confounding model (2). Let $\widehat{\beta}_2(t,\boldsymbol{s})$ be the second element of $\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s})\in\mathbb{R}^{q+1}$, $\beta_2(t,\boldsymbol{s})=\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$, and $\mathbb{X}=\{(T_i,\boldsymbol{S}_i)\}_{i=1}^n$. Then, for any $(t,\boldsymbol{s})\in\mathcal{E}^\circ$ and $q>0$, as $h,b,\frac{b^4}{h}\to 0$ and $nh^3b^d\to\infty$, we know that the asymptotic conditional covariance of $\widehat{\beta}_2(t,\boldsymbol{s})\in\mathbb{R}$ is given by*

$$\mathrm{Var}\left[\widehat{\beta}_2(t,\boldsymbol{s})\middle|\mathbb{X}\right]=\frac{\sigma^2}{nh^3b^d\cdot p(t,\boldsymbol{s})}\left[\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{M}_q^*\boldsymbol{M}_q^{-1}\boldsymbol{e}_2+O\left(\max\{h,b\}\right)+O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right],$$

and the asymptotic conditional bias is given by

$$\mathrm{Bias}\left[\widehat{\beta}_2(t,\boldsymbol{s})\middle|\mathbb{X}\right]=\begin{cases}\frac{1}{p(t,\boldsymbol{s})}\left[h^{q+1}\tau_q^{odd}+b^2\tau_q^*+O\left(\frac{b^4}{h}\right)+o_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] & q \text{ is odd},\\ \frac{1}{p(t,\boldsymbol{s})}\left[h^{q}\tau_q^{even}+b^2\tau_q^*+O\left(\frac{b^4}{h}\right)+o_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] & q \text{ is even}.\end{cases}$$

Here, $\mathcal{E}^\circ$ is the interior of the support $\mathcal{E}\subset\mathcal{T}\times\mathcal{S}$, $\boldsymbol{e}_2=(0,1,0,...,0)^T\in\mathbb{R}^{q+1}$, $\boldsymbol{M}_q=\left(\kappa_{i+j-2}^{(T)}\right)_{1\leq i,j\leq q+1}\in\mathbb{R}^{(q+1)\times(q+1)}$, $\boldsymbol{M}_q^*=\left(\nu_{i+j-2}^{(T)}\nu_0^{(S)}\right)_{1\leq i,j\leq q+1}\in\mathbb{R}^{(q+1)\times(q+1)}$, $\tau_q^{odd}=\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\tau}_q}{h}$ when q is odd and $\tau_q^{even}=\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\tau}_q$ when q is even, as well as $\tau_q^*=\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\tau}_q^*}{b}$, where

$$\boldsymbol{\tau}_q=\left[\frac{\partial^{q+1}}{\partial t^{q+1}}\mu(t,\boldsymbol{s})\right]\frac{1}{(q+1)!}\begin{bmatrix}\left(\kappa_{q+j}^{(T)}\cdot p(t,\boldsymbol{s})+h\cdot\kappa_{q+j+1}^{(T)}\cdot\frac{\partial}{\partial t}p(t,\boldsymbol{s})\right)_{1\leq j\leq q+1}\\ \left(b\kappa_{q+1}^{(T)}\kappa_{2,j-q-1}^{(S)}\cdot\frac{\partial}{\partial s_{j-q-1}}p(t,\boldsymbol{s})\right)_{q+1<j\leq q+1+d}\end{bmatrix}\in\mathbb{R}^{(q+1+d)\times(q+1+d)},$$

$$\boldsymbol{\tau}_q^*=\begin{bmatrix}\left(b\sum_{\ell=1}^d\left[\frac{\partial^2}{\partial t\partial s_\ell}\mu(t,\boldsymbol{s})\right]\kappa_j^{(T)}\kappa_{2,\ell}^{(S)}\cdot\frac{\partial}{\partial s_\ell}p(t,\boldsymbol{s})\right)_{1\leq j\leq q+1}\\ \left(h\left[\frac{\partial^2}{\partial t\partial s_{j-q-1}}\mu(t,\boldsymbol{s})\right]\kappa_2^{(T)}\kappa_{2,j-q-1}^{(S)}\cdot\frac{\partial}{\partial t}p(t,\boldsymbol{s})\right)_{q+1<j\leq q+1+d}\end{bmatrix}\in\mathbb{R}^{(q+1+d)\times(q+1+d)},$$

and

$$\widetilde{\boldsymbol{\tau}}_q=\begin{bmatrix}\left(\kappa_{j-1}^{(T)}p(t,\boldsymbol{s})\sum_{\ell=1}^d\frac{\kappa_{2,\ell}^{(S)}}{2}\left(\frac{\partial^2}{\partial s_\ell^2}\mu(t,\boldsymbol{s})\right)\right)_{1\leq j\leq q+1}\\ \left(b\sum_{\ell=1}^d\frac{\kappa_{2,j-q-1,\ell}^{(S)}}{2}\left[\frac{\partial^2}{\partial s_\ell^2}\mu(t,\boldsymbol{s})\right]\cdot\frac{\partial}{\partial s_{j-q-1}}p(t,\boldsymbol{s})\right)_{q+1<j\leq q+1+d}\end{bmatrix}\in\mathbb{R}^{(q+1+d)\times(q+1+d)}.$$

Furthermore, as $h, b, \frac{\max\{h,b\}^4}{h} \to 0$, *and* $nh^3 b^d \to \infty$, *we have that*

$$\begin{aligned} &\widehat{\beta}_2(t,\boldsymbol{s}) - \beta_2(t,\boldsymbol{s}) \\ &= \begin{cases} O\left(h^{q+1} + b^2 + \frac{b^4}{h}\right) + O_P\left(\sqrt{\frac{1}{nh^3 b^d}}\right) & \text{if } q \text{ is odd and } (t,\boldsymbol{s}) \in \mathcal{E}^\circ, \\ O\left(h^{q} + b^2 + \frac{b^4}{h}\right) + O_P\left(\sqrt{\frac{1}{nh^3 b^d}}\right) & \text{if } q \text{ is even and } (t,\boldsymbol{s}) \in \mathcal{E}^\circ, \\ O\left(h^{q} + \frac{\max\{h,b\}^4}{h}\right) + O_P\left(\sqrt{\frac{1}{nh^3 b^d}}\right) & \text{if } q \text{ is an integer and } (t,\boldsymbol{s}) \in \partial\mathcal{E}. \end{cases} \end{aligned}$$

PROOF OF LEMMA 2. Recall from (10) that

$$\begin{aligned} \left(\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}), \widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\right)^T &= \left[\boldsymbol{X}^T(t,\boldsymbol{s})\boldsymbol{W}(t,\boldsymbol{s})\boldsymbol{X}(t,\boldsymbol{s})\right]^{-1} \boldsymbol{X}(t,\boldsymbol{s})^T \boldsymbol{W}(t,\boldsymbol{s})\boldsymbol{Y} \\ &\equiv \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1} \boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y}. \end{aligned}$$

The proof of Lemma 2 has two major components, where we consider the cases when $(t,\boldsymbol{s})$ is an interior point or a boundary point of the support $\mathcal{E}$. When $(t,\boldsymbol{s})$ is an interior point of $\mathcal{E}$, we also divide the arguments into three steps that deal with the rates of convergence for the term $\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}$, the conditional covariance term $\mathrm{Cov}\left[\left(\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}), \widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\right)^T\right]$, and the conditional bias term $\mathrm{Bias}\left[\left(\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}), \widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\right)^T \middle| \mathbb{X}\right]$ separately.

We first consider the case when $(t,\boldsymbol{s})$ is an interior point of the support $\mathcal{E}$, *i.e.*, for any $(t_1,\boldsymbol{s}_1) \in \mathcal{T} \times \mathcal{S}$, $\frac{t_1 - t}{h}$ and $\frac{\boldsymbol{s}_1 - \boldsymbol{s}}{b}$ lie in the supports of K_T and K_S respectively when h, b are small. Notice that under the confounding model (2), we have that

$$\mathrm{Cov}\left[\left(\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}), \widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\right)^T \middle| \mathbb{X}\right] = \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1} \boldsymbol{X}^T\boldsymbol{\Sigma}\boldsymbol{X}\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}$$

and

$$\begin{aligned} \mathrm{Bias}\left[\left(\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}), \widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\right)^T \middle| \mathbb{X}\right] &= \mathbb{E}\left[\left(\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}), \widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\right)^T\right] - \left(\boldsymbol{\beta}(t,\boldsymbol{s}), \boldsymbol{\alpha}(t,\boldsymbol{s})\right)^T \\ &= \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1} \boldsymbol{X}^T\boldsymbol{W}\left[\begin{pmatrix} \mu(T_1,\boldsymbol{S}_1) \\ \vdots \\ \mu(T_n,\boldsymbol{S}_n) \end{pmatrix} - \boldsymbol{X}\begin{pmatrix} \boldsymbol{\beta}(t,\boldsymbol{s}) \\ \boldsymbol{\alpha}(t,\boldsymbol{s}) \end{pmatrix}\right], \end{aligned}$$

where $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$ is a diagonal matrix as $\mathtt{Diag}\left(\sigma^2 K_T^2\left(\frac{T_1 - t}{h}\right) K_S^2\left(\frac{\boldsymbol{S}_1 - \boldsymbol{s}}{b}\right), ..., \sigma^2 K_T^2\left(\frac{T_n - t}{h}\right) K_S^2\left(\frac{\boldsymbol{S}_n - \boldsymbol{s}}{b}\right)\right)$ and by Assumption A3 and Taylor's expansion,

$$\left(\boldsymbol{\beta}(t,\boldsymbol{s}), \boldsymbol{\alpha}(t,\boldsymbol{s})\right)^T \equiv \left(\mu(t,\boldsymbol{s}), \frac{\partial}{\partial t}\mu(t,\boldsymbol{s}), ..., \frac{1}{q!}\cdot\frac{\partial^q}{\partial t^q}\mu(t,\boldsymbol{s}), \frac{\partial}{\partial s_1}\mu(t,\boldsymbol{s}), ..., \frac{\partial}{\partial s_d}\mu(t,\boldsymbol{s})\right)^T.$$

Step 1: Common term $\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}$. Before deriving the asymptotic behaviors of the above conditional covariance matrix and bias, we first study the rates of convergence of $\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X} \in \mathbb{R}^{(q+1+d)\times(q+1+d)}$.

By direct calculations, we know that

(24)

$$
\begin{aligned}
&\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j} \\
&= \begin{cases}
\sum_{k=1}^n (T_k-t)^{i+j-2} K_T\left(\frac{T_k-t}{h}\right) K_S\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right), & 1\le i,j\le q+1,\\
\sum_{k=1}^n (T_k-t)^{i-1}\left(S_{k,j-q-1}-s_{j-q-1}\right) K_T\left(\frac{T_k-t}{h}\right) K_S\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right), & 1\le i\le q+1 \text{ and } q+1<j\le q+1+d,\\
\sum_{k=1}^n \left(S_{k,i-q-1}-s_{i-q-1}\right)(T_k-t)^{j-1} K_T\left(\frac{T_k-t}{h}\right) K_S\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right), & q+1<i\le q+1+d \text{ and } 1\le j\le q+1,\\
\sum_{k=1}^n \left(S_{k,i-q-1}-s_{i-q-1}\right)\left(S_{k,j-q-1}-s_{j-q-1}\right) K_T\left(\frac{T_k-t}{h}\right) K_S\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right), & q+1<i,j\le q+1+d.
\end{cases}
\end{aligned}
$$

Here, each sample $\boldsymbol{S}_k\in\mathcal{S}\subset\mathbb{R}^d$ is written as $\boldsymbol{S}_k=(S_{k,1},...,S_{k,d})^T\in\mathbb{R}^d$. We now derive the asymptotic rates of convergence of the expectation and variance for each term in (24) under Assumptions A3, A4, and A6.

• **Case I:** $1\le i,j\le q+1$. We compute that

$$
\begin{aligned}
&\mathbb{E}\left[\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j}\right]\\
&= n\int_{\mathbb{R}\times\mathbb{R}^d}\left(\tilde{t}-t\right)^{i+j-2} K_T\left(\frac{\tilde{t}-t}{h}\right) K_S\left(\frac{\tilde{\boldsymbol{s}}-\boldsymbol{s}}{b}\right) p(\tilde{t},\tilde{\boldsymbol{s}})\,d\tilde{t}d\tilde{\boldsymbol{s}}\\
&\overset{\text{(i)}}{=} nh^{i+j-1}b^d\int_{\mathbb{R}\times\mathbb{R}^d} u^{i+j-2}K_T(u)K_S(\boldsymbol{v})\cdot p(t+uh,\boldsymbol{s}+b\boldsymbol{v})\,dud\boldsymbol{v}\\
&\overset{\text{(ii)}}{=} nh^{i+j-1}b^d\int_{\mathbb{R}\times\mathbb{R}^d} u^{i+j-2}K_T(u)K_S(\boldsymbol{v})\left[p(t,\boldsymbol{s})+uh\cdot\frac{\partial}{\partial t}p(t,\boldsymbol{s})+b\boldsymbol{v}^T\frac{\partial}{\partial\boldsymbol{s}}p(t,\boldsymbol{s})+O\left(\max\{h,b\}^2\right)\right]dud\boldsymbol{v}\\
&\overset{\text{(iii)}}{=} nh^{i+j-1}b^d\left[\kappa_{i+j-2}^{(T)}\cdot p(t,\boldsymbol{s})+h\cdot\kappa_{i+j-1}^{(T)}\cdot\frac{\partial}{\partial t}p(t,\boldsymbol{s})+O\left(\max\{h,b\}^2\right)\right],
\end{aligned}
$$

where (i) utilizes the changes of variables $u=\frac{\tilde{t}-t}{h}$ and $\boldsymbol{v}=\frac{\tilde{\boldsymbol{s}}-\boldsymbol{s}}{b}$, (ii) leverages the differentiability of $p(t,\boldsymbol{s})$ and apply Taylor's expansion, as well as (iii) uses the symmetric properties of K_T,K_S with notations in Assumption A6(a). In addition, we calculate that

$$
\begin{aligned}
&\mathrm{Var}\left[\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j}\right]\\
&= n\cdot\mathrm{Var}\left[(T_1-t)^{i+j-2}K_T\left(\frac{T_1-t}{h}\right)K_S\left(\frac{\boldsymbol{S}_1-\boldsymbol{s}}{b}\right)\right]\\
&\le n\int_{\mathbb{R}\times\mathbb{R}^d}(\tilde{t}-t)^{2i+2j-4}K_T^2\left(\frac{\tilde{t}-t}{h}\right)K_S^2\left(\frac{\tilde{\boldsymbol{s}}-\boldsymbol{s}}{b}\right)p(\tilde{t},\tilde{\boldsymbol{s}})\,d\tilde{t}d\tilde{\boldsymbol{s}}\\
&= nh^{2i+2j-3}b^d\int_{\mathbb{R}\times\mathbb{R}^d}u^{2i+2j-4}K_T^2(u)K_S^2(v)p(t+uh,\boldsymbol{s}+b\boldsymbol{v})\,dud\boldsymbol{v}\\
&= nh^{2i+2j-3}b^d\int_{\mathbb{R}\times\mathbb{R}^d}u^{2i+2j-4}K_T^2(u)K_S^2(v)\left[p(t,\boldsymbol{s})+uh\cdot\frac{\partial}{\partial t}p(t,\boldsymbol{s})+b\boldsymbol{v}^T\frac{\partial}{\partial\boldsymbol{s}}p(t,\boldsymbol{s})+O\left(\max\{h,b\}^2\right)\right]dud\boldsymbol{v}\\
&= nh^{2i+2j-3}b^d\left[\nu_{2i+2j-4}^{(T)}\nu_0^{(S)}\cdot p(t,\boldsymbol{s})+h\cdot\nu_{2i+2j-3}^{(T)}\nu_0^{(S)}\cdot\frac{\partial}{\partial t}p(t,\boldsymbol{s})+O\left(\max\{h,b\}^2\right)\right].
\end{aligned}
$$

The above calculations on $\mathbb{E}\left[\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j}\right]$ and $\mathrm{Var}\left[\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j}\right]$ imply that

$$
\begin{aligned}
&\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j}\\
&= \mathbb{E}\left[\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j}\right] + O_P\left(\sqrt{\mathrm{Var}\left[(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})_{i,j}\right]}\right)\\
&= nh^{i+j-1}b^d\left[\kappa_{i+j-2}^{(T)}\cdot p(t,\boldsymbol{s}) + h\cdot\kappa_{i+j-1}^{(T)}\cdot\frac{\partial}{\partial t}p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right]
\end{aligned}
$$

for any $1\le i,j\le q+1$.

• **Case II:** $1\le i\le q+1$ and $q+1<j\le q+1+d$. We compute that

$$
\begin{aligned}
&\mathbb{E}\left[\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j}\right]\\
&= n\int_{\mathbb{R}\times\mathbb{R}^d}\left(\widetilde{t}-t\right)^{i-1}(\widetilde{s}_{j-q-1}-s_{j-q-1})K_T\left(\frac{\widetilde{t}-t}{h}\right)K_S\left(\frac{\widetilde{\boldsymbol{s}}-\boldsymbol{s}}{b}\right)p(\widetilde{t},\widetilde{\boldsymbol{s}})\,d\widetilde{t}d\widetilde{\boldsymbol{s}}\\
&= nh^ib^{d+1}\int_{\mathbb{R}\times\mathbb{R}^d}u^{i-1}v_{j-q-1}K_T(u)K_S(\boldsymbol{v})p(t+hu,\boldsymbol{s}+b\boldsymbol{v})\,dud\boldsymbol{v}\\
&= nh^ib^{d+1}\int_{\mathbb{R}\times\mathbb{R}^d}u^{i-1}v_{j-q-1}K_T(u)K_S(\boldsymbol{v})\left[p(t,\boldsymbol{s}) + uh\cdot\frac{\partial}{\partial t}p(t,\boldsymbol{s}) + b\boldsymbol{v}^T\frac{\partial}{\partial\boldsymbol{s}}p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right)\right]dud\boldsymbol{v}\\
&= nh^ib^{d+1}\left[b\cdot\kappa_{i-1}^{(T)}\kappa_{2,j-q-1}^{(S)}\cdot\frac{\partial}{\partial s_{j-q-1}}p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right)\right]
\end{aligned}
$$

and

$$
\begin{aligned}
&\mathrm{Var}\left[\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j}\right]\\
&= n\cdot\mathrm{Var}\left[(T_1-t)^{i-1}(S_{1,j-q-1}-s_{j-q-1})K_T\left(\frac{T_1-t}{h}\right)K_S\left(\frac{\boldsymbol{S}_1-\boldsymbol{s}}{b}\right)\right]\\
&\le n\int_{\mathbb{R}\times\mathbb{R}^d}(\widetilde{t}-t)^{2i-2}(\widetilde{s}_{j-q-1}-s_{j-q-1})^2K_T^2\left(\frac{\widetilde{t}-t}{h}\right)K_S^2\left(\frac{\widetilde{\boldsymbol{s}}-\boldsymbol{s}}{b}\right)p(\widetilde{t},\widetilde{\boldsymbol{s}})\,d\widetilde{t}d\widetilde{\boldsymbol{s}}\\
&= nh^{2i-1}b^{d+2}\int_{\mathbb{R}\times\mathbb{R}^d}u^{2i-2}v_{j-q-1}^2K_T^2(u)K_S^2(\boldsymbol{v})\left[p(t,\boldsymbol{s}) + uh\cdot\frac{\partial}{\partial t}p(t,\boldsymbol{s}) + b\boldsymbol{v}^T\frac{\partial}{\partial\boldsymbol{s}}p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right)\right]dud\boldsymbol{v}\\
&= nh^{2i-1}b^{d+2}\left[\nu_{2i-2}^{(T)}\nu_{2,j-q-1}^{(S)}\cdot p(t,\boldsymbol{s}) + O(h) + O\left(\max\{h,b\}^2\right)\right].
\end{aligned}
$$

These two terms indicate that

$$
\begin{aligned}
\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j} &= \mathbb{E}\left[\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j}\right] + O_P\left(\sqrt{\mathrm{Var}\left[(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})_{i,j}\right]}\right)\\
&= nh^ib^{d+1}\left[b\cdot\kappa_{i-1}^{(T)}\kappa_{2,j-q-1}^{(S)}\cdot\frac{\partial}{\partial s_{j-q-1}}p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right]
\end{aligned}
$$

for any $1\le i\le q+1$ and $q+1<j\le q+1+d$.

• **Case III:** $q+1<i\le q+1+d$ and $1\le j\le q+1$. By swapping the roles of i and j in our calculations for **Case 2**, we obtain that

$$\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j}=nh^jb^{d+1}\left[b\cdot\kappa_{j-1}^{(T)}\kappa_{2,i-q-1}^{(S)}\cdot\frac{\partial}{\partial s_{i-q-1}}p(t,\boldsymbol{s})+O\left(\max\{h,b\}^2\right)+O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right]$$

for any $q+1<i\le q+1+d$ and $1\le j\le q+1$.

• **Case IV:** $q+1<i,j\le q+1+d$. We compute that

$$\begin{aligned}
&\mathbb{E}\left[\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j}\right]\\
&=n\int_{\mathbb{R}\times\mathbb{R}^d}(\widetilde{s}_{i-q-1}-s_{i-q-1})(\widetilde{s}_{j-q-1}-s_{j-q-1})K_T\left(\frac{\widetilde{t}-t}{h}\right)K_S\left(\frac{\widetilde{\boldsymbol{s}}-\boldsymbol{s}}{b}\right)p(\widetilde{t},\widetilde{\boldsymbol{s}})\,d\widetilde{t}d\widetilde{\boldsymbol{s}}\\
&=nhb^{d+2}\int_{\mathbb{R}\times\mathbb{R}^d}v_{i-q-1}\cdot v_{j-q-1}K_T(u)K_S(\boldsymbol{v})\cdot p(t+uh,\boldsymbol{s}+b\boldsymbol{v})\,dud\boldsymbol{v}\\
&=nhb^{d+2}\int_{\mathbb{R}\times\mathbb{R}^d}v_{i-q-1}\cdot v_{j-q-1}K_T(u)K_S(\boldsymbol{v})\left[p(t,\boldsymbol{s})+uh\cdot\frac{\partial}{\partial t}p(t,\boldsymbol{s})+b\boldsymbol{v}^T\frac{\partial}{\partial\boldsymbol{s}}p(t,\boldsymbol{s})+O\left(\max\{h,b\}^2\right)\right]dud\boldsymbol{v}\\
&=nhb^{d+2}\left[\kappa_{2,i-q-1}^{(S)}\mathbb{1}_{\{i=j\}}\cdot p(t,\boldsymbol{s})+O\left(\max\{h,b\}^2\right)\right]
\end{aligned}$$

and

$$\begin{aligned}
&\mathrm{Var}\left[\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j}\right]\\
&=n\cdot\mathrm{Var}\left[(S_{1,i-q-1}-s_{i-q-1})(S_{1,j-q-1}-s_{j-q-1})K_T\left(\frac{T_1-t}{h}\right)K_S\left(\frac{\boldsymbol{S}_1-\boldsymbol{s}}{b}\right)\right]\\
&\le n\int_{\mathbb{R}\times\mathbb{R}^d}(\widetilde{s}_{i-q-1}-s_{i-q-1})^2(\widetilde{s}_{j-q-1}-s_{j-q-1})^2K_T^2\left(\frac{\widetilde{t}-t}{h}\right)K_S^2\left(\frac{\widetilde{\boldsymbol{s}}-\boldsymbol{s}}{b}\right)p(\widetilde{t},\widetilde{\boldsymbol{s}})\,d\widetilde{t}d\widetilde{\boldsymbol{s}}\\
&=nhb^{d+4}\int_{\mathbb{R}\times\mathbb{R}^d}v_{i-q-1}^2v_{j-q-1}^2K_T^2(u)K_S^2(\boldsymbol{v})\left[p(t,\boldsymbol{s})+uh\cdot\frac{\partial}{\partial t}p(t,\boldsymbol{s})+b\boldsymbol{v}^T\frac{\partial}{\partial\boldsymbol{s}}p(t,\boldsymbol{s})+O\left(\max\{h,b\}^2\right)\right]dud\boldsymbol{v}\\
&=nhb^{d+4}\left[\nu_{2,i-q-1}^{(S)}\nu_{2,j-q-1}^{(S)}\cdot p(t,\boldsymbol{s})+O\left(\max\{h,b\}^2\right)\right].
\end{aligned}$$

The above calculations imply that

$$\begin{aligned}
\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j}&=\mathbb{E}\left[\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j}\right]+O_P\left(\sqrt{\mathrm{Var}\left[(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X})_{i,j}\right]}\right)\\
&=nhb^{d+2}\left[\kappa_{2,i-q-1}^{(S)}\mathbb{1}_{\{i=j\}}\cdot p(t,\boldsymbol{s})+O\left(\max\{h,b\}^2\right)+O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right]
\end{aligned}$$

for any $q+1<i,j\le q+1+d$.

Therefore, we summarize all the above cases as:

$$\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)_{i,j}$$

$$
= \begin{cases} nh^{i+j-1}b^d \left[\kappa_{i+j-2}^{(T)} \cdot p(t,\boldsymbol{s}) + h \cdot \kappa_{i+j-1}^{(T)} \cdot \frac{\partial}{\partial t} p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right], \\ \qquad\qquad 1 \le i,j \le q+1, \\ nh^i b^{d+1} \left[b \cdot \kappa_{i-1}^{(T)} \kappa_{2,j-q-1}^{(S)} \cdot \frac{\partial}{\partial s_{j-q-1}} p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right], \\ \qquad\qquad 1 \le i \le q+1 \text{ and } q+1 < j \le q+1+d, \\ nh^j b^{d+1} \left[b \cdot \kappa_{j-1}^{(T)} \kappa_{2,i-q-1}^{(S)} \cdot \frac{\partial}{\partial s_{i-q-1}} p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right], \\ \qquad\qquad q+1 < i \le q+1+d \text{ and } 1 \le j \le q+1, \\ nhb^{d+2} \left[\kappa_{2,i-q-1}^{(S)} \mathbb{1}_{\{i=j\}} \cdot p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right], \quad q+1 < i,j \le q+1+d. \end{cases}
$$

Let $\boldsymbol{H} = \mathtt{Diag}(1,h,...,h^q,b,...,b) \in \mathbb{R}^{(q+1+d)\times(q+1+d)}$. We also denote

$$
\boldsymbol{M}_q = \begin{pmatrix} \left(\kappa_{i+j-2}^{(T)}\right)_{1\le i,j\le q+1} & \boldsymbol{0} \\ \boldsymbol{0} & \left(\kappa_{2,i-q-1}^{(S)} \mathbb{1}_{\{i=j\}}\right)_{q+1<i,j\le q+1+d} \end{pmatrix} \in \mathbb{R}^{(q+1+d)\times(q+1+d)}
$$

and

$$
\begin{aligned}
&\widetilde{\boldsymbol{M}}_{q,h,b} \\
&= \begin{pmatrix} h \cdot \frac{\partial}{\partial t} p(t,\boldsymbol{s}) \left(\kappa_{i+j-1}^{(T)}\right)_{1\le i,j\le q+1} & b \cdot \mathtt{Diag}\left(\frac{\partial}{\partial \boldsymbol{s}} p(t,\boldsymbol{s})\right) \left(\kappa_{i-1}^{(T)} \kappa_{2,j-q-1}^{(S)}\right)_{1\le i\le q+1, q+1<j\le q+1+d} \\ b \cdot \mathtt{Diag}\left(\frac{\partial}{\partial \boldsymbol{s}} p(t,\boldsymbol{s})\right) \left(\kappa_{j-1}^{(T)} \kappa_{2,i-q-1}^{(S)}\right)_{q+1<i\le q+1+d, 1\le j\le q+1} & \boldsymbol{0} \end{pmatrix} \\
&\in \mathbb{R}^{(q+1+d)\times(q+1+d)}.
\end{aligned}
$$

Then, we can rewrite the asymptotic behaviors of $\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X}$ in its matrix form as:

(25)

$$
\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X} = nhb^d \cdot p(t,\boldsymbol{s}) \cdot \boldsymbol{H} \left[\boldsymbol{M}_q + \frac{\widetilde{\boldsymbol{M}}_{q,h,b}}{p(t,\boldsymbol{s})} + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \boldsymbol{H},
$$

where an abuse of notation is applied when we use $O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb}}\right)$ to denote a matrix whose entries are of this order. By the matrix inversion formula

$$
(A + \max\{h,b\} \cdot B)^{-1} = A^{-1} - \max\{h,b\} \cdot A^{-1} B A^{-1} + O\left(\max\{h,b\}^2\right),
$$

we know that

$$
\begin{aligned}
&\left(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X}\right)^{-1} \\
&= \frac{1}{nhb^d \cdot p(t,\boldsymbol{s})} \cdot \boldsymbol{H}^{-1} \left[\boldsymbol{M}_q^{-1} - \boldsymbol{M}_q^{-1} \cdot \frac{\widetilde{\boldsymbol{M}}_{q,h,b}}{p(t,\boldsymbol{s})} \cdot \boldsymbol{M}_q^{-1} + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \boldsymbol{H}^{-1}.
\end{aligned}
$$

Step 2: Conditional covariance term $\mathrm{Cov}\left[\left(\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}), \widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\right)^T\right]$**.** Following our calculations for $\left(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X}\right)_{i,j}$ for $1 \le i,j \le q+1+d$, we can similarly derive that

$$
\left(\boldsymbol{X}^T \boldsymbol{\Sigma} \boldsymbol{X}\right)_{i,j}
$$

$$= \begin{cases} \sigma^2 \sum_{k=1}^n (T_k - t)^{i+j-2} K_T^2\left(\frac{T_k-t}{h}\right) K_S^2\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right), & 1 \leq i,j \leq q+1, \\ \sigma^2 \sum_{k=1}^n (T_k - t)^{i-1} \left(S_{k,j-q-1} - s_{j-q-1}\right) K_T^2\left(\frac{T_k-t}{h}\right) K_S^2\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right), & 1 \leq i \leq q+1 \text{ and } q+1 < j \leq q+1+d, \\ \sigma^2 \sum_{k=1}^n \left(S_{k,i-q-1} - s_{i-q-1}\right) (T_k - t)^{j-1} K_T^2\left(\frac{T_k-t}{h}\right) K_S^2\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right), & q+1 < i \leq q+1+d \text{ and } 1 \leq j \leq q+1, \\ \sigma^2 \sum_{k=1}^n \left(S_{k,i-q-1} - s_{i-q-1}\right)\left(S_{k,j-q-1} - s_{j-q-1}\right) K_T^2\left(\frac{T_k-t}{h}\right) K_S^2\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right), & q+1 < i,j \leq q+1+d, \end{cases}$$

and its asymptotic behaviors become

$$\left(\boldsymbol{X}^T \boldsymbol{\Sigma} \boldsymbol{X}\right)_{i,j}$$

$$= \begin{cases} \sigma^2 n h^{i+j-1} b^d \left[\nu_{i+j-2}^{(T)} \nu_0^{(S)} \cdot p(t,\boldsymbol{s}) + h \cdot \nu_{i+j-1}^{(T)} \nu_0^{(S)} \cdot \frac{\partial}{\partial t} p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right], \\ \qquad 1 \leq i,j \leq q+1, \\ n h^i b^{d+1} \left[b \cdot \nu_{i-1}^{(T)} \nu_{2,j-q-1}^{(S)} \cdot \frac{\partial}{\partial s_{j-q-1}} p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right], \\ \qquad 1 \leq i \leq q+1 \text{ and } q+1 < j \leq q+1+d, \\ n h^j b^{d+1} \left[b \cdot \nu_{j-1}^{(T)} \nu_{2,i-q-1}^{(S)} \cdot \frac{\partial}{\partial s_{i-q-1}} p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right], \\ \qquad q+1 < i \leq q+1+d \text{ and } 1 \leq j \leq q+1, \\ n h b^{d+2} \left[\nu_0^{(T)} \nu_{2,i-q-1}^{(S)} \mathbb{1}_{\{i=j\}} \cdot p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right], \quad q+1 < i,j \leq q+1+d. \end{cases}$$

If we denote

$$\boldsymbol{M}_q^* = \begin{pmatrix} \left(\nu_{i+j-2}^{(T)} \nu_0^{(S)}\right)_{1 \leq i,j \leq q+1} & \boldsymbol{0} \\ \boldsymbol{0} & \left(\nu_0^{(T)} \nu_{2,i-q-1}^{(S)} \mathbb{1}_{\{i=j\}}\right)_{q+1 < i,j \leq q+1+d} \end{pmatrix} \in \mathbb{R}^{(q+1+d)\times(q+1+d)}$$

and

$$\widetilde{\boldsymbol{M}}_{q,h,b}^*$$
$$= \begin{pmatrix} h \cdot \frac{\partial}{\partial t} p(t,\boldsymbol{s}) \cdot \left(\nu_{i+j-1}^{(T)} \nu_0^{(S)}\right)_{1 \leq i,j \leq q+1} & b \cdot \mathtt{Diag}\left(\frac{\partial}{\partial \boldsymbol{s}} p(t,\boldsymbol{s})\right) \left(\nu_{i-1}^{(T)} \nu_{2,j-q-1}^{(S)}\right)_{1 \leq i \leq q+1, q+1 < j \leq q+1+d} \\ b \cdot \mathtt{Diag}\left(\frac{\partial}{\partial \boldsymbol{s}} p(t,\boldsymbol{s})\right) \left(\nu_{j-1}^{(T)} \nu_{2,i-q-1}^{(S)}\right)_{q+1 < i \leq q+1+d, 1 \leq j \leq q+1} & \boldsymbol{0} \end{pmatrix}$$
$$\in \mathbb{R}^{(q+1+d)\times(q+1+d)},$$

then the asymptotic behavior of $\boldsymbol{X}^T \boldsymbol{\Sigma} \boldsymbol{X}$ in its matrix form is

$$\boldsymbol{X}^T \boldsymbol{\Sigma} \boldsymbol{X} = nhb^d \sigma^2 \cdot p(t,\boldsymbol{s}) \boldsymbol{H} \left[\boldsymbol{M}_q^* + \frac{\widetilde{\boldsymbol{M}}_{q,h,b}^*}{p(t,\boldsymbol{s})} + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \boldsymbol{H}.$$

With our results for $\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X}$ in (25) and Assumption A4, we conclude that the asymptotic conditional covariance matrix of $\left(\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}), \widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\right)^T \in \mathbb{R}^{q+1+d}$ is

$$\mathrm{Cov}\left[\left(\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}), \widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\right)^T \Big| \mathbb{X}\right]$$
$$= \left(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X}\right)^{-1} \boldsymbol{X}^T \boldsymbol{\Sigma} \boldsymbol{X} \left(\boldsymbol{X}^T \boldsymbol{W} \boldsymbol{X}\right)^{-1}$$
$$= \frac{\sigma^2}{nhb^d \cdot p(t,\boldsymbol{s})} \cdot \boldsymbol{H}^{-1} \left[\boldsymbol{M}_q^{-1} - \boldsymbol{M}_q^{-1} \cdot \frac{\widetilde{\boldsymbol{M}}_{q,h,b}}{p(t,\boldsymbol{s})} \cdot \boldsymbol{M}_q^{-1} + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right]$$

$$\times\left[\boldsymbol{M}_q^* + \widetilde{\boldsymbol{M}}_{q,h,b}^* + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right]$$

$$\times\left[\boldsymbol{M}_q^{-1} - \boldsymbol{M}_q^{-1}\cdot\frac{\widetilde{\boldsymbol{M}}_{q,h,b}}{p(t,\boldsymbol{s})}\cdot\boldsymbol{M}_q^{-1} + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right]\boldsymbol{H}^{-1}$$

$$= \frac{\sigma^2}{nhb^d\cdot p(t,\boldsymbol{s})}\cdot\boldsymbol{H}^{-1}\left[\boldsymbol{M}_q^{-1}\boldsymbol{M}_q^*\boldsymbol{M}_q^{-1} + O\left(\max\{h,b\}\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right]\boldsymbol{H}^{-1}.$$

In particular, we know that the asymptotic conditional variance of $\widehat{\beta}_2(t,\boldsymbol{s})$ is given by

$$\mathrm{Var}\left[\widehat{\beta}_2(t,\boldsymbol{s})\middle|\mathbb{X}\right] = \frac{\sigma^2}{nh^3b^d\cdot p(t,\boldsymbol{s})}\left[\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{M}_q^*\boldsymbol{M}_q^{-1}\boldsymbol{e}_2 + O\left(\max\{h,b\}\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right],$$

where $\{\boldsymbol{e}_1,...,\boldsymbol{e}_{q+1+d}\}$ is the standard basis in $\mathbb{R}^{q+1+d}$. Since the only entries in $\boldsymbol{M}_q, \boldsymbol{M}_q^*$ that affects $\mathrm{Var}\left[\widehat{\beta}_2(t,\boldsymbol{s})\middle|\mathbb{X}\right]$ are those in the first $(q+1+d)$ rows and columns, the results follow by restricting $\boldsymbol{M}_q, \boldsymbol{M}_q^*$ to their first $(q+1)\times(q+1)$ block matrices and taking $\boldsymbol{e}_2 = (0,1,0,...,0)^T \in \mathbb{R}^{q+1}$.

Step 3: Conditional bias term $\mathrm{Bias}\left[\left(\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}),\widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\right)^T\middle|\mathbb{X}\right]$**.** By Assumption A3 and Taylor's expansion, we have that

$$\mu(T,\boldsymbol{S}) = \mu(t,\boldsymbol{s}) + \sum_{j=1}^q \frac{1}{q!}\cdot\frac{\partial^j}{\partial t^j}\mu(t,\boldsymbol{s})\cdot(T-t)^j + \sum_{\ell=1}^d \frac{\partial}{\partial s_\ell}\mu(t,\boldsymbol{s})\cdot(S_\ell - s_\ell) + r(T,\boldsymbol{S}), \tag{26}$$

where the reminder term $r(T,\boldsymbol{S})$ is given by

$$r(T,\boldsymbol{S}) = \frac{(T-t)^{q+1}}{(q+1)!}\left[\frac{\partial^{q+1}}{\partial t^{q+1}}\mu(t,\boldsymbol{s})\right] + \sum_{\ell=1}^d\left[\frac{\partial^2}{\partial t\partial s_\ell}\mu(t,\boldsymbol{s})\right](T-t)(S_\ell - s_\ell)$$
$$+ \frac{1}{2}(\boldsymbol{S}-\boldsymbol{s})^T\left[\frac{\partial^2}{\partial\boldsymbol{s}\partial\boldsymbol{s}^T}\mu(t,\boldsymbol{s})\right](\boldsymbol{S}-\boldsymbol{s}) + o\left(|T-t|^{q+1} + |T-t|\,||\boldsymbol{S}-\boldsymbol{s}||_2 + ||\boldsymbol{S}-\boldsymbol{s}||_2^2\right).$$

Thus, the conditional bias can be written as:

$$\mathrm{Bias}\left[\left(\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}),\widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\right)^T\middle|\mathbb{X}\right] = \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\left[\begin{pmatrix}\mu(T_1,\boldsymbol{S}_1)\\ \vdots\\ \mu(T_n,\boldsymbol{S}_n)\end{pmatrix} - \boldsymbol{X}\begin{pmatrix}\boldsymbol{\beta}(t,\boldsymbol{s})\\ \boldsymbol{\alpha}(t,\boldsymbol{s})\end{pmatrix}\right]$$
$$= \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\begin{pmatrix}r(T_1,\boldsymbol{S}_1)\\ \vdots\\ r(T_n,\boldsymbol{S}_n)\end{pmatrix}.$$

Now, we note that

$$\boldsymbol{X}^T\boldsymbol{W}\begin{pmatrix}r(T_1,\boldsymbol{S}_1)\\ \vdots\\ r(T_n,\boldsymbol{S}_n)\end{pmatrix}$$

$$
= \begin{bmatrix} \begin{pmatrix} \sum_{k=1}^{n} \left[\frac{\partial^{q+1}}{\partial t^{q+1}}\mu(t,\boldsymbol{s})\right] \frac{(T_k-t)^{q+j}}{(q+1)!} \cdot K_T\left(\frac{T_k-t}{h}\right) K_S\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right) \\ + \sum_{k=1}^{n}\sum_{\ell=1}^{d} \left[\frac{\partial^2}{\partial t \partial s_\ell}\mu(t,\boldsymbol{s})\right] (T_k-t)^j (S_{k,\ell}-s_\ell) K_T\left(\frac{T_k-t}{h}\right) K_S\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right) \\ + \sum_{k=1}^{n} (T_k-t)^{j-1}\cdot \frac{1}{2}(\boldsymbol{S}_k-\boldsymbol{s})^T \left[\frac{\partial^2}{\partial \boldsymbol{s}\partial \boldsymbol{s}^T}\mu(t,\boldsymbol{s})\right] (\boldsymbol{S}_k-\boldsymbol{s}) K_T\left(\frac{T_k-t}{h}\right) K_S\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right) \end{pmatrix}_{1\le j\le q+1} \\ \begin{pmatrix} \sum_{k=1}^{n} \left[\frac{\partial^{q+1}}{\partial t^{q+1}}\mu(t,\boldsymbol{s})\right] \frac{(T_k-t)^{q+1}}{(q+1)!} \cdot (S_{k,j-q-1}-s_{j-q-1}) \cdot K_T\left(\frac{T_k-t}{h}\right) K_S\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right) \\ + \sum_{k=1}^{n}\sum_{\ell=1}^{d} \left[\frac{\partial^2}{\partial t \partial s_\ell}\mu(t,\boldsymbol{s})\right] (T_k-t)(S_{k,\ell}-s_\ell)(S_{k,j-q-1}-s_{j-q-1}) K_T\left(\frac{T_k-t}{h}\right) K_S\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right) \\ + \sum_{k=1}^{n} (S_{k,j-q-1}-s_{j-q-1})\cdot \frac{1}{2}(\boldsymbol{S}_k-\boldsymbol{s})^T \left[\frac{\partial^2}{\partial \boldsymbol{s}\partial \boldsymbol{s}^T}\mu(t,\boldsymbol{s})\right] (\boldsymbol{S}_k-\boldsymbol{s}) K_T\left(\frac{T_k-t}{h}\right) K_S\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right) \end{pmatrix}_{q+1<j\le q+1+d} \end{bmatrix}
$$

$$
= \begin{bmatrix} \begin{pmatrix} nh^{q+1+j}b^d \left[\frac{\partial^{q+1}}{\partial t^{q+1}}\mu(t,\boldsymbol{s})\right] \frac{1}{(q+1)!}\left[\kappa_{q+j}^{(T)} p(t,\boldsymbol{s}) + \kappa_{q+j+1}^{(T)} h \frac{\partial}{\partial t}p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \\ + nh^{j+1}b^{d+1}\left[b\cdot \sum_{\ell=1}^{d}\left[\frac{\partial^2}{\partial t\partial s_\ell}\mu(t,\boldsymbol{s})\right]\kappa_j^{(T)}\kappa_{2,\ell}^{(S)}\cdot \frac{\partial}{\partial s_\ell}p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \\ + nh^j b^{d+2}\left[\kappa_{j-1}^{(T)} p(t,\boldsymbol{s}) \sum_{\ell=1}^{d} \frac{\kappa_{2,\ell}^{(S)}}{2}\left(\frac{\partial^2}{\partial s_\ell^2}\mu(t,\boldsymbol{s})\right) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \end{pmatrix}_{1\le j\le q+1} \\ \begin{pmatrix} nh^{q+2}b^{d+1} \left[\frac{\partial^{q+1}}{\partial t^{q+1}}\mu(t,\boldsymbol{s})\right] \frac{1}{(q+1)!}\left[b\kappa_{q+1}^{(T)}\kappa_{2,j-q-1}^{(S)} \frac{\partial}{\partial s_{j-q-1}}p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \\ + nh^2 b^{d+2}\left[\frac{\partial^2}{\partial t\partial s_{j-q-1}}\mu(t,\boldsymbol{s})\right]\left[h\kappa_2^{(T)}\kappa_{2,j-q-1}^{(S)}\cdot \frac{\partial}{\partial t}p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \\ + nhb^{d+3}\left[b\sum_{\ell=1}^{d}\frac{\kappa_{2,j-q-1,\ell}^{(S)}}{2}\left(\frac{\partial^2}{\partial s_\ell^2}\mu(t,\boldsymbol{s})\right)\cdot \frac{\partial}{\partial s_{j-q-1}}p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \end{pmatrix}_{q+1<j\le q+1+d} \end{bmatrix}
$$

$$
= nhb^d \boldsymbol{H} \begin{bmatrix} \begin{pmatrix} h^{q+1} \left[\frac{\partial^{q+1}}{\partial t^{q+1}}\mu(t,\boldsymbol{s})\right] \frac{1}{(q+1)!}\left[\kappa_{q+j}^{(T)} p(t,\boldsymbol{s}) + \kappa_{q+j+1}^{(T)} h \frac{\partial}{\partial t}p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \\ + hb\left[b\cdot \sum_{\ell=1}^{d}\left(\frac{\partial^2}{\partial t\partial s_\ell}\mu(t,\boldsymbol{s})\right)\kappa_j^{(T)}\kappa_{2,\ell}^{(S)}\cdot \frac{\partial}{\partial s_\ell}p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \\ + b^2\left[\kappa_{j-1}^{(T)} p(t,\boldsymbol{s}) \sum_{\ell=1}^{d} \frac{\kappa_{2,\ell}^{(S)}}{2}\left(\frac{\partial^2}{\partial s_\ell^2}\mu(t,\boldsymbol{s})\right) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \end{pmatrix}_{1\le j\le q+1} \\ \begin{pmatrix} h^{q+1} \left[\frac{\partial^{q+1}}{\partial t^{q+1}}\mu(t,\boldsymbol{s})\right] \frac{1}{(q+1)!}\left[b\kappa_{q+1}^{(T)}\kappa_{2,j-q-1}^{(S)} \frac{\partial}{\partial s_{j-q-1}}p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \\ + hb\left[\frac{\partial^2}{\partial t\partial s_{j-q-1}}\mu(t,\boldsymbol{s})\right]\left[h\kappa_2^{(T)}\kappa_{2,j-q-1}^{(S)}\cdot \frac{\partial}{\partial t}p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \\ + b^2\left[b\sum_{\ell=1}^{d}\frac{\kappa_{2,j-q-1,\ell}^{(S)}}{2}\left(\frac{\partial^2}{\partial s_\ell^2}\mu(t,\boldsymbol{s})\right)\cdot \frac{\partial}{\partial s_{j-q-1}}p(t,\boldsymbol{s}) + O\left(\max\{h,b\}^2\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \end{pmatrix}_{q+1<j\le q+1+d} \end{bmatrix}.
$$

Therefore, we know from the above display and (25) that

$$
\begin{aligned}
&\text{Bias}\left[\left(\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}), \widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\right)^T \Big| \mathbb{X}\right] \\
&= \left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\begin{pmatrix} r(T_1,\boldsymbol{S}_1) \\ \vdots \\ r(T_n,\boldsymbol{S}_n)\end{pmatrix} \\
&= \frac{1}{p(t,\boldsymbol{s})}\cdot \boldsymbol{H}^{-1}\left[\boldsymbol{M}_q^{-1} + O\left(\max\{h,b\}\right) + O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right]
\end{aligned}
$$

$$\times \left[\begin{matrix} \begin{pmatrix} h^{q+1}\left[\frac{\partial^{q+1}}{\partial t^{q+1}}\mu(t,\boldsymbol{s})\right]\frac{1}{(q+1)!}\left[\kappa_{q+j}^{(T)}p(t,\boldsymbol{s})+\kappa_{q+j+1}^{(T)}h\frac{\partial}{\partial t}p(t,\boldsymbol{s})+O\left(\max\{h,b\}^2\right)+O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \\ +hb\left[b\cdot\sum_{\ell=1}^d\left[\frac{\partial^2}{\partial t\partial s_\ell}\mu(t,\boldsymbol{s})\right]\kappa_j^{(T)}\kappa_{2,\ell}^{(S)}\cdot\frac{\partial}{\partial s_\ell}p(t,\boldsymbol{s})+O\left(\max\{h,b\}^2\right)+O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \\ +b^2\left[\kappa_{j-1}^{(T)}p(t,\boldsymbol{s})\sum_{\ell=1}^d\frac{\kappa_{2,\ell}^{(S)}}{2}\left(\frac{\partial^2}{\partial s_\ell^2}\mu(t,\boldsymbol{s})\right)+O\left(\max\{h,b\}^2\right)+O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \end{pmatrix}_{1\le j\le q+1} \\ \begin{pmatrix} h^{q+1}\left[\frac{\partial^{q+1}}{\partial t^{q+1}}\mu(t,\boldsymbol{s})\right]\frac{1}{(q+1)!}\left[b\kappa_{q+1}^{(T)}\kappa_{2,j-q-1}^{(S)}\frac{\partial}{\partial s_{j-q-1}}p(t,\boldsymbol{s})+O\left(\max\{h,b\}^2\right)+O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \\ +hb\left[\frac{\partial^2}{\partial t\partial s_{j-q-1}}\mu(t,\boldsymbol{s})\right]\left[h\kappa_2^{(T)}\kappa_{2,j-q-1}^{(S)}\cdot\frac{\partial}{\partial t}p(t,\boldsymbol{s})+O\left(\max\{h,b\}^2\right)+O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \\ +b^2\left[b\sum_{\ell=1}^d\frac{\kappa_{2,j-q-1,\ell}^{(S)}}{2}\left(\frac{\partial^2}{\partial s_\ell^2}\mu(t,\boldsymbol{s})\right)\cdot\frac{\partial}{\partial s_{j-q-1}}p(t,\boldsymbol{s})+O\left(\max\{h,b\}^2\right)+O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \end{pmatrix}_{q+1<j\le q+1+d} \end{matrix}\right]$$

$$= \frac{1}{p(t,\boldsymbol{s})}\boldsymbol{H}^{-1}\left[\boldsymbol{M}_q^{-1}+O\left(\max\{h,b\}\right)+O_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right]$$
$$\times\left[h^{q+1}\boldsymbol{\tau}_q+hb\boldsymbol{\tau}_q^*+b^2\widetilde{\boldsymbol{\tau}}_q+O\left(\max\{h,b\}^4\right)+o_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right],$$

where

$$\boldsymbol{\tau}_q=\left[\frac{\partial^{q+1}}{\partial t^{q+1}}\mu(t,\boldsymbol{s})\right]\frac{1}{(q+1)!}\left[\begin{matrix}\left(\kappa_{q+j}^{(T)}\cdot p(t,\boldsymbol{s})+h\cdot\kappa_{q+j+1}^{(T)}\cdot\frac{\partial}{\partial t}p(t,\boldsymbol{s})\right)_{1\le j\le q+1}\\ \left(b\kappa_{q+1}^{(T)}\kappa_{2,j-q-1}^{(S)}\cdot\frac{\partial}{\partial s_{j-q-1}}p(t,\boldsymbol{s})\right)_{q+1<j\le q+1+d}\end{matrix}\right]\in\mathbb{R}^{(q+1+d)\times(q+1+d)},$$

$$\boldsymbol{\tau}_q^*=\left[\begin{matrix}\left(b\sum_{\ell=1}^d\left[\frac{\partial^2}{\partial t\partial s_\ell}\mu(t,\boldsymbol{s})\right]\kappa_j^{(T)}\kappa_{2,\ell}^{(S)}\cdot\frac{\partial}{\partial s_\ell}p(t,\boldsymbol{s})\right)_{1\le j\le q+1}\\ \left(h\left[\frac{\partial^2}{\partial t\partial s_{j-q-1}}\mu(t,\boldsymbol{s})\right]\kappa_2^{(T)}\kappa_{2,j-q-1}^{(S)}\cdot\frac{\partial}{\partial t}p(t,\boldsymbol{s})\right)_{q+1<j\le q+1+d}\end{matrix}\right]\in\mathbb{R}^{(q+1+d)\times(q+1+d)},$$

and

$$\widetilde{\boldsymbol{\tau}}_q=\left[\begin{matrix}\left(\kappa_{j-1}^{(T)}p(t,\boldsymbol{s})\sum_{\ell=1}^d\frac{\kappa_{2,\ell}^{(S)}}{2}\left(\frac{\partial^2}{\partial s_\ell^2}\mu(t,\boldsymbol{s})\right)\right)_{1\le j\le q+1}\\ \left(b\sum_{\ell=1}^d\frac{\kappa_{2,j-q-1,\ell}^{(S)}}{2}\left[\frac{\partial^2}{\partial s_\ell^2}\mu(t,\boldsymbol{s})\right]\cdot\frac{\partial}{\partial s_{j-q-1}}p(t,\boldsymbol{s})\right)_{q+1<j\le q+1+d}\end{matrix}\right]\in\mathbb{R}^{(q+1+d)\times(q+1+d)}.$$

Since $\kappa_{2j-1}^{(T)}=0$ for all $j=1,2,...,q+1$ by Assumption A6(a), we know that

$$\boldsymbol{M}_q=\begin{pmatrix}\left(\kappa_{i+j-2}^{(T)}\right)_{1\le i,j\le q+1} & \boldsymbol{0}\\ \boldsymbol{0} & \left(\kappa_{2,i-q-1}^{(S)}\mathbb{1}_{\{i=j\}}\right)_{q+1<i,j\le q+1+d}\end{pmatrix}$$

$$=\begin{pmatrix}\begin{matrix}\kappa_0^{(T)} & 0 & \kappa_2^{(T)} & 0 & \cdots & \kappa_q^{(T)}\\ 0 & \kappa_2^{(T)} & 0 & \kappa_4^{(T)} & \cdots & \kappa_{q+1}^{(T)}\\ \vdots & & \ddots & & & \vdots\\ \kappa_q^{(T)} & & \cdots & & & \kappa_{2q}^{(T)}\end{matrix} & \boldsymbol{0}\\ \boldsymbol{0} & \left(\kappa_{2,i-q-1}^{(S)}\mathbb{1}_{\{i=j\}}\right)_{q+1<i,j\le q+1+d}\end{pmatrix}, \tag{27}$$

i.e., each row/column of the upper $(q+1)\times(q+1)$ submatrix of M_q has interleaving nonzero and zero entries. More importantly, given that

$$\boldsymbol{M}_q^{-1}=\frac{1}{\det(\boldsymbol{M}_q)}\cdot\mathrm{adj}(\boldsymbol{M}_q)$$

with $\mathrm{adj}(\boldsymbol{M}_q) \in \mathbb{R}^{(q+1+d)\times(q+1+d)}$ being the adjugate of $\boldsymbol{M}_q$, it can be shown that $\boldsymbol{M}_q^{-1}$ has the identical sparsity structures as $\boldsymbol{M}_q$. In more details, any (i,j)-minor of $\boldsymbol{M}_q$ with $i+j$ being odd has two rows/columns that share an identical sparsity structure. Recalling the formula of calculating a determinant

$$\det(A) = \sum_{\sigma_q} \left(\mathrm{sign}(\sigma_q) \prod_{i=1}^{q+1+d} A_{i\sigma_q(i)} \right),$$

where the sum is over all $(q+1+d)!$ permutations σ_q, we know that each summand is zero when we compute the determinant of (i,j)-minor of $\boldsymbol{M}_q$ with $i+j$ being odd through the above formula. Hence, the determinant of (i,j)-minor of $\boldsymbol{M}_q$ with $i+j$ being odd is zero. A similar argument can be found in the proof of Theorem 2 in Fan et al. (1996).

Using this result, we derive the bias term $\mathrm{Bias}\left[\widehat{\beta}_2(t,\boldsymbol{s})\middle|\mathbb{X}\right]$ when q is either odd or even (recommended) as follows.

• **Case I:** q **is odd.** Then,

$$\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} = \Big(\underbrace{0,\star,0,\star,...,0,\star}_{1\le j\le q+1}, \underbrace{\boldsymbol{0}}_{q+1<j\le q+1+d} \Big) \in \mathbb{R}^{q+1+d},$$

where $\star \in \mathbb{R}$ stands for some nonzero bounded value (not necessarily equal for each entry). This implies that

$$\begin{aligned}
&\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\tau}_q \\
&= \Big(\underbrace{0,\star,0,\star,...,0,\star}_{1\le j\le q+1}, \underbrace{\boldsymbol{0}}_{q+1<j\le q+1+d} \Big) \left[\frac{\partial^{q+1}}{\partial t^{q+1}}\mu(t,\boldsymbol{s})\right] \frac{1}{(q+1)!} \begin{bmatrix} \begin{pmatrix} \kappa_{q+1}^{(T)} \cdot p(t,\boldsymbol{s}) \\ h\cdot \kappa_{q+3}^{(T)} \cdot \frac{\partial}{\partial t} p(t,\boldsymbol{s}) \\ \vdots \\ h\cdot \kappa_{2q+2}^{(T)} \cdot \frac{\partial}{\partial t} p(t,\boldsymbol{s}) \end{pmatrix}_{1\le j\le q+1} \\ \left(b\cdot \kappa_{q+1}^{(T)} \kappa_{2,j-q-1}^{(S)} \cdot \frac{\partial}{\partial s_{j-q-1}} p(t,\boldsymbol{s}) \right)_{q+1<j\le q+1+d} \end{bmatrix} \\
&= h\cdot \tau_q^{\mathrm{odd}},
\end{aligned}$$

where $\tau_q^{\mathrm{odd}} = \frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\tau}_q}{h}$ is a dominating term independent of h, b. Similarly, we know that

$$\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\tau}_q^* = b\cdot \tau_q^*, \quad \boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \widetilde{\tau}_q = 0,$$

where $\tau_q^* = \frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\tau}_q^*}{b}$ is a constant independent of b, h. Hence, when q is odd and $h, b \to 0, nhb^d \to \infty$, we obtain that the asymptotic conditional bias of $\widehat{\beta}_2(t,\boldsymbol{s})$ is given by

$$\begin{aligned}
&\mathrm{Bias}\left[\widehat{\beta}_2(t,\boldsymbol{s})\middle|\mathbb{X}\right] \\
&= \frac{1}{h\cdot p(t,\boldsymbol{s})} \left[h^{q+1} \boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\tau}_q + hb \boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\tau}_q^* + b^2 \boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \widetilde{\boldsymbol{\tau}}_q + o_P\left(\sqrt{\frac{1}{nhb^d}}\right) \right] \quad (28) \\
&= \frac{1}{p(t,\boldsymbol{s})} \left[h^{q+1} \tau_q^{\mathrm{odd}} + b^2 \tau_q^* + O\left(\frac{b^4}{h}\right) + o_P\left(\sqrt{\frac{1}{nhb^d}}\right) \right].
\end{aligned}$$

• **Case II: q is even.** Then,

$$\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} = \left(\underbrace{0, \star, 0, \star, ..., 0, \star, 0}_{1 \le j \le q+1}, \underbrace{\mathbf{0}}_{q+1<j\le q+1+d} \right) \in \mathbb{R}^{q+1+d},$$

where $\star \in \mathbb{R}$ stands for some nonzero bounded value (not necessarily equal for each entry). This indicates that

$$\begin{aligned}
&\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\tau}_q \\
&= \left(\underbrace{0, \star, 0, \star, ..., 0, \star, 0}_{1 \le j \le q+1}, \underbrace{\mathbf{0}}_{q+1<j\le q+1+d} \right) \left[\frac{\partial^{q+1}}{\partial t^{q+1}} \mu(t, \boldsymbol{s}) \right] \frac{1}{(q+1)!} \begin{bmatrix} \begin{pmatrix} h \cdot \kappa_{q+2}^{(T)} \cdot \frac{\partial}{\partial t} p(t, \boldsymbol{s}) \\ \kappa_{q+2}^{(T)} \cdot p(t, \boldsymbol{s}) \\ \vdots \\ h \cdot \kappa_{2q+2}^{(T)} \cdot \frac{\partial}{\partial t} p(t, \boldsymbol{s}) \end{pmatrix}_{1 \le j \le q+1} \\ \mathbf{0}_{q+1<j\le q+1+d} \end{bmatrix} \\
&= \tau_q^{\text{even}},
\end{aligned}$$

where $\tau_q^{\text{even}} = \boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\tau}_q$ is a dominating constant independent of h, b. Similarly, we know that

$$\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\tau}_q^* = b \cdot \tau_q^*, \quad \boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \widetilde{\tau}_q = 0,$$

where $\tau_q^* = \frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\tau}_q^*}{b}$ is a constant independent of b, h. Hence, when q is even and $h, b \to 0, nhb^d \to \infty$, we obtain that the asymptotic conditional bias of $\widehat{\beta}_2(t, \boldsymbol{s})$ is given by

$$\begin{aligned}
&\text{Bias}\left[\widehat{\beta}_2(t, \boldsymbol{s}) \middle| \mathbb{X} \right] \\
&= \frac{1}{h \cdot p(t, \boldsymbol{s})} \left[h^{q+1} \boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\tau}_q + hb \boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\tau}_q^* + b^2 \boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \widetilde{\boldsymbol{\tau}}_q + O\left(\max\{h, b\}^4 \right) + o_P\left(\sqrt{\frac{1}{nhb^d}} \right) \right] \\
&= \frac{1}{p(t, \boldsymbol{s})} \left[h^q \tau_q^{\text{even}} + b^2 \tau_q^* + O\left(\frac{b^4}{h} \right) + o_P\left(\sqrt{\frac{1}{nhb^d}} \right) \right].
\end{aligned}$$

In summary, as $h, b \to 0$ and $nh^3 b^d \to \infty$, we know that

$$\text{Bias}\left[\widehat{\beta}_2(t, \boldsymbol{s}) \middle| \mathbb{X} \right] = \begin{cases} \frac{1}{p(t, \boldsymbol{s})} \left[h^{q+1} \tau_q^{\text{odd}} + b^2 \tau_q^* + O\left(\frac{b^4}{h} \right) + o_P\left(\sqrt{\frac{1}{nhb^d}} \right) \right] & q \text{ is odd}, \\ \frac{1}{p(t, \boldsymbol{s})} \left[h^q \tau_q^{\text{even}} + b^2 \tau_q^* + O\left(\frac{b^4}{h} \right) + o_P\left(\sqrt{\frac{1}{nhb^d}} \right) \right] & q \text{ is even}. \end{cases}$$

Now, we consider the case when $(t, \boldsymbol{s})$ lies in the boundary region of $\mathcal{E}$. In this case, those dominating constants in $\text{Var}\left[\widehat{\beta}_2(t, \boldsymbol{s}) \middle| \mathbb{X} \right]$ and $\text{Bias}\left[\widehat{\beta}_2(t, \boldsymbol{s}) \middle| \mathbb{X} \right]$ would be different but the asymptotic rates remain the same due to Assumption A5(a). In more details,

$\kappa_j^{(T)}, \nu_j^{(T)}, \kappa_{j,\ell}^{(S)}, \nu_{j,k}^{(S)}$ are now defined as:

$$\begin{aligned}
\kappa_j^{(T)} &= \int_{\{u\in\mathbb{R}:(t+hu,\boldsymbol{s}+b\boldsymbol{v})\in\mathcal{E} \text{ for some } \boldsymbol{v}\in\mathbb{R}^d\}} u^j K_T(u)\,du < \infty, \\
\nu_j^{(T)} &= \int_{\{u\in\mathbb{R}:(t+hu,\boldsymbol{s}+b\boldsymbol{v})\in\mathcal{E} \text{ for some } \boldsymbol{v}\in\mathbb{R}^d\}} u^j K_T^2(u)\,du < \infty, \\
\kappa_{j,\ell}^{(S)} &= \int_{\{\boldsymbol{v}\in\mathbb{R}^d:(t+hu,\boldsymbol{s}+b\boldsymbol{v})\in\mathcal{E} \text{ for some } u\in\mathbb{R}\}} u_\ell^j K_S(\boldsymbol{u})\,d\boldsymbol{u} < \infty, \\
\text{and} \quad \nu_{j,k}^{(S)} &= \int_{\{\boldsymbol{v}\in\mathbb{R}^d:(t+hu,\boldsymbol{s}+b\boldsymbol{v})\in\mathcal{E} \text{ for some } u\in\mathbb{R}\}} u_\ell^j K_S^2(\boldsymbol{u})\,d\boldsymbol{u} < \infty.
\end{aligned} \tag{29}$$

These terms are again bounded even when $h, b \to 0$. Under the above new definitions, we can redefine $\boldsymbol{M}_q, \widetilde{\boldsymbol{M}}_{q,h,b}, \boldsymbol{M}_q^*, \widetilde{\boldsymbol{M}}_{q,h,b}^*, \boldsymbol{\tau}_q, \boldsymbol{\tau}_q^*, \widetilde{\boldsymbol{\tau}}_q$ accordingly, where those partial derivatives of $p(t, \boldsymbol{s})$ and the corresponding Taylor's expansion are defined at a nearby interior point by Assumption A5(a) and taking the limit to the boundary $\partial\mathcal{E}$. More importantly, the matrix $\boldsymbol{M}_q$ remains non-singular as $h, b \to 0$ due to Assumption A5(a) and Lemma 7.1 in Fan and Guerre (2015). A similar argument can also be found in Theorem 2.2 in Ruppert and Wand (1994).

However, the matrices $\boldsymbol{M}_q$ and $\boldsymbol{M}_q^{-1}$ no longer have the interleaving zero structures as in (27) because of the asymmetric integrated ranges for those terms in (29). Nevertheless, by our Assumption A5(b), we know that both $\boldsymbol{\tau}_q^* = \widetilde{\boldsymbol{\tau}}_q = \boldsymbol{0}$. Thus, no matter $q > 0$ is odd or even, the asymptotic conditional bias of $\widehat{\beta}_2(t, \boldsymbol{s})$ is given by

$$\begin{aligned}
&\text{Bias}\left[\widehat{\beta}_2(t,\boldsymbol{s})\middle|\mathbb{X}\right] \\
&= \frac{1}{h\cdot p(t,\boldsymbol{s})}\left[h^{q+1}\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\tau}_q + hb\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\tau}_q^* + b^2\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\widetilde{\boldsymbol{\tau}}_q + O\left(\max\{h,b\}^4\right) + o_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right] \\
&= \frac{1}{p(t,\boldsymbol{s})}\left[h^q\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\tau}_q + O\left(\frac{\max\{h,b\}^4}{h}\right) + o_P\left(\sqrt{\frac{1}{nhb^d}}\right)\right].
\end{aligned}$$

Finally, for any $(t, \boldsymbol{s}) \in \mathcal{E}$, the unconditional asymptotic rate of convergence of $\widehat{\beta}_2(t, \boldsymbol{s})$ can be derived by noting that

$$\mathbb{E}\left[\widehat{\beta}_2(t,\boldsymbol{s})\right] = \mathbb{E}\left[\mathbb{E}\left(\widehat{\beta}_2(t,\boldsymbol{s})\middle|\mathbb{X}\right)\right] \quad \text{and} \quad \text{Var}\left[\widehat{\beta}_2(t,\boldsymbol{s})\right] = \mathbb{E}\left[\text{Var}\left(\widehat{\beta}_2(t,\boldsymbol{s})\middle|\mathbb{X}\right)\right] + \text{Var}\left[\mathbb{E}\left(\widehat{\beta}_2(t,\boldsymbol{s})\middle|\mathbb{X}\right)\right].$$

Therefore, if $(t, \boldsymbol{s})$ is an interior point of $\mathcal{E}$, then we have that

$$\widehat{\beta}_2(t,\boldsymbol{s}) - \frac{\partial}{\partial t}\mu(t,\boldsymbol{s}) = \begin{cases} O\left(h^{q+1} + b^2 + \frac{b^4}{h}\right) + O_P\left(\sqrt{\frac{1}{nh^3b^d}}\right) & q \text{ is odd}, \\ O\left(h^q + b^2 + \frac{b^4}{h}\right) + O_P\left(\sqrt{\frac{1}{nh^3b^d}}\right) & q \text{ is even}, \end{cases}$$

as $h, b \to 0$ and $nh^3b^d \to \infty$. Otherwise, if $(t, \boldsymbol{s}) \in \partial\mathcal{E}$, then we have that

$$\widehat{\beta}_2(t,\boldsymbol{s}) - \frac{\partial}{\partial t}\mu(t,\boldsymbol{s}) = O\left(h^q + \frac{\max\{h,b\}^4}{h}\right) + O_P\left(\sqrt{\frac{1}{nh^3b^d}}\right)$$

as $h, b \to 0$ and $nh^3b^d \to \infty$. The results follow. □

B.3. Proof of Lemma 3.

LEMMA 3 (Uniform convergence of $\widehat{\beta}_2(t,\boldsymbol{s})$). *Let $q>0$. Suppose that Assumptions A3, A4, A5, and A6(a,b) hold under the confounding model (2). Let $\widehat{\beta}_2(t,\boldsymbol{s})$ be the second element of $\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s})\in\mathbb{R}^{q+1}$ and $\beta_2(t,\boldsymbol{s})=\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$. Then, as $h,b,\frac{\max\{h,b\}^4}{h}\to 0$ and $\frac{nh^3b^d}{|\log(hb^d)|},\frac{|\log(hb^d)|}{\log\log n}\to\infty$, we know that*

$$\sup_{(t,\boldsymbol{s})\in\mathcal{E}}\left|\widehat{\beta}_2(t,\boldsymbol{s})-\beta_2(t,\boldsymbol{s})\right|=O(h^q)+O\left(b^2\right)+O\left(\frac{\max\{h,b\}^4}{h}\right)+O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}}\right).$$

PROOF OF LEMMA 3. The proof of Lemma 3 is partially inspired by the proof of Lemma 7 in Cheng and Chen (2019). Recall from (10) that

$$\left(\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}),\widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\right)^T=\left[\boldsymbol{X}^T(t,\boldsymbol{s})\boldsymbol{W}(t,\boldsymbol{s})\boldsymbol{X}(t,\boldsymbol{s})\right]^{-1}\boldsymbol{X}(t,\boldsymbol{s})^T\boldsymbol{W}(t,\boldsymbol{s})\boldsymbol{Y}\equiv\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y}.$$

By direct calculations with $\boldsymbol{H}=\mathtt{Diag}(1,h,...,h^q,b,...,b)\in\mathbb{R}^{(q+1+d)\times(q+1+d)}$, we have that

(30)
$$\begin{aligned}\left(\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s}),\widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\right)^T&=\left(\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y}\\&=\frac{1}{nhb^d}\cdot\boldsymbol{H}^{-1}\left[\frac{1}{nhb^d}(\boldsymbol{X}\boldsymbol{H}^{-1})^T\boldsymbol{W}(\boldsymbol{X}\boldsymbol{H}^{-1})\right]^{-1}\boldsymbol{H}^{-1}\boldsymbol{X}^T\boldsymbol{W}\boldsymbol{Y}\\&=\boldsymbol{H}^{-1}\left[\frac{1}{nhb^d}(\boldsymbol{X}\boldsymbol{H}^{-1})^T\boldsymbol{W}(\boldsymbol{X}\boldsymbol{H}^{-1})\right]^{-1}\left[\frac{1}{nhb^d}(\boldsymbol{X}\boldsymbol{H}^{-1})^T\boldsymbol{W}\boldsymbol{Y}\right].\end{aligned}$$

We first derive the uniform rate of convergence for $\frac{1}{nhb^d}(\boldsymbol{X}\boldsymbol{H}^{-1})^T\boldsymbol{W}(\boldsymbol{X}\boldsymbol{H}^{-1})$. Recall from (24) in the proof of Lemma 2 that

(31)
$$\begin{aligned}&\left[\frac{1}{nhb^d}(\boldsymbol{X}\boldsymbol{H}^{-1})^T\boldsymbol{W}(\boldsymbol{X}\boldsymbol{H}^{-1})\right]_{i,j}\\&=\begin{cases}\frac{1}{nhb^d}\sum_{k=1}^n\left(\frac{T_k-t}{h}\right)^{i+j-2}K_T\left(\frac{T_k-t}{h}\right)K_S\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right), & 1\le i,j\le q+1,\\ \frac{1}{nhb^d}\sum_{k=1}^n\left(\frac{T_k-t}{h}\right)^{i-1}\left(\frac{S_{k,j-q-1}-s_{j-q-1}}{b}\right)K_T\left(\frac{T_k-t}{h}\right)K_S\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right), & 1\le i\le q+1\text{ and }q+1<j\le q+1+d,\\ \frac{1}{nhb^d}\sum_{k=1}^n\left(\frac{S_{k,i-q-1}-s_{i-q-1}}{b}\right)\left(\frac{T_k-t}{h}\right)^{j-1}K_T\left(\frac{T_k-t}{h}\right)K_S\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right), & q+1<i\le q+1+d\text{ and }1\le j\le q+1,\\ \frac{1}{nhb^d}\sum_{k=1}^n\left(\frac{S_{k,i-q-1}-s_{i-q-1}}{b}\right)\left(\frac{S_{k,j-q-1}-s_{j-q-1}}{b}\right)K_T\left(\frac{T_k-t}{h}\right)K_S\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right), & q+1<i,j\le q+1+d.\end{cases}\end{aligned}$$

By our notation in (14), we can obtain from (25) in the proof of Lemma 2 that

$$\begin{aligned}&\frac{1}{nhb^d}(\boldsymbol{X}\boldsymbol{H}^{-1})^T\boldsymbol{W}(\boldsymbol{X}\boldsymbol{H}^{-1})\\&=p(t,\boldsymbol{s})\boldsymbol{M}_q+O\left(\max\{h,b\}\right)+\underbrace{\frac{1}{nhb^d}(\boldsymbol{X}\boldsymbol{H}^{-1})^T\boldsymbol{W}(\boldsymbol{X}\boldsymbol{H}^{-1})-\mathbb{E}\left[\frac{1}{nhb^d}(\boldsymbol{X}\boldsymbol{H}^{-1})^T\boldsymbol{W}(\boldsymbol{X}\boldsymbol{H}^{-1})\right]}_{\textbf{Term A}},\end{aligned}$$

where we apply an abuse of notation when using $O\left(\max\{h,b\}\right)$ to denote a matrix whose entries are of this order.

By (31), each entry of the matrix in **Term A** is a mean-zero empirical process (scaled by $\frac{1}{\sqrt{nhb^d}}$) over a function in the class $\mathcal{K}_{q,d}$ defined in Assumption A6(b). By Theorem 2.3 in Giné and Guillou (2002) or Theorem 1 in Einmahl and Mason (2005), we know that

$$\sup_{(t,\boldsymbol{s})\in\mathcal{E}} \left|\left| \frac{1}{nhb^d}(\boldsymbol{X}\boldsymbol{H}^{-1})^T\boldsymbol{W}(\boldsymbol{X}\boldsymbol{H}^{-1}) - \mathbb{E}\left[\frac{1}{nhb^d}(\boldsymbol{X}\boldsymbol{H}^{-1})^T\boldsymbol{W}(\boldsymbol{X}\boldsymbol{H}^{-1})\right]\right|\right|_{\max} = O_P\left(\sqrt{\frac{|\log(hb^d)|}{nhb^d}}\right),$$

where $||\boldsymbol{A}||_{\max} = \max_{i,j}|\boldsymbol{A}_{ij}|$ for $\boldsymbol{A}\in\mathbb{R}^{(q+1+d)\times(q+1+d)}$. Therefore, we conclude that

$$\sup_{(t,\boldsymbol{s})\in\mathcal{E}} \left|\left| \frac{1}{nhb^d}(\boldsymbol{X}\boldsymbol{H}^{-1})^T\boldsymbol{W}(\boldsymbol{X}\boldsymbol{H}^{-1}) - p(t,\boldsymbol{s})\boldsymbol{M}_q\right|\right|_{\max} = O\left(\max\{h,b\}\right) + O_P\left(\sqrt{\frac{|\log(hb^d)|}{nhb^d}}\right)$$

and

$$\sup_{(t,\boldsymbol{s})\in\mathcal{E}} \left|\left| \left[\frac{1}{nhb^d}\left(\boldsymbol{X}\boldsymbol{H}^{-1}\right)^T\boldsymbol{W}\left(\boldsymbol{X}\boldsymbol{H}^{-1}\right)\right]^{-1} - \frac{\boldsymbol{M}_q^{-1}}{p(t,\boldsymbol{s})}\right|\right|_{\max} = O\left(\max\{h,b\}\right) + O_P\left(\sqrt{\frac{|\log(hb^d)|}{nhb^d}}\right).$$

Based on the above result and (30), we have that

$$\begin{aligned}
\widehat{\beta}_2(t,\boldsymbol{s}) &= \boldsymbol{e}_2^T\begin{bmatrix}\widehat{\boldsymbol{\beta}}(t,\boldsymbol{s})\\ \widehat{\boldsymbol{\alpha}}(t,\boldsymbol{s})\end{bmatrix}\\
&= \frac{1}{h}\cdot\boldsymbol{e}_2^T\left[\frac{1}{nhb^d}(\boldsymbol{X}\boldsymbol{H}^{-1})^T\boldsymbol{W}(\boldsymbol{X}\boldsymbol{H}^{-1})\right]^{-1}\left[\frac{1}{nhb^d}(\boldsymbol{X}\boldsymbol{H}^{-1})^T\boldsymbol{W}\boldsymbol{Y}\right] \qquad (32)\\
&= \frac{1}{h\cdot p(t,\boldsymbol{s})}\cdot\boldsymbol{e}_2^T\left[\boldsymbol{M}_q^{-1}+\boldsymbol{A}_{h,b}\right]\left[\frac{1}{nhb^d}(\boldsymbol{X}\boldsymbol{H}^{-1})^T\boldsymbol{W}\boldsymbol{Y}\right],
\end{aligned}$$

where each entry of $\boldsymbol{A}_{h,b}\in\mathbb{R}^{(q+1+d)\times(q+1+d)}$ is uniformly of the order $O\left(\max\{h,b\}\right) + O_P\left(\sqrt{\frac{|\log(hb^d)|}{nhb^d}}\right)$. Notice also that

$$\begin{aligned}
\frac{1}{nhb^d}(\boldsymbol{X}\boldsymbol{H}^{-1})^T\boldsymbol{W}\boldsymbol{Y} &= \begin{bmatrix}\left(\frac{1}{nhb^d}\sum_{k=1}^n Y_k\left(\frac{T_k-t}{h}\right)^{j-1}K_T\left(\frac{T_k-t}{h}\right)K_S\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right)\right)_{1\le j\le q+1}\\ \left(\frac{1}{nhb^d}\sum_{k=1}^n Y_k\left(\frac{S_{k,j-q-1}-s_{j-q-1}}{b}\right)K_T\left(\frac{T_k-t}{h}\right)K_S\left(\frac{\boldsymbol{S}_k-\boldsymbol{s}}{b}\right)\right)_{q+1<j\le q+1+d}\end{bmatrix}\\
&= \mathbb{P}_n\left(\frac{1}{hb^d}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right),
\end{aligned}$$

where $\mathbb{P}_n$ is the empirical measure associated with observations $\{(Y_k,T_k,\boldsymbol{S}_k)\}_{k=1}^n$ and $\boldsymbol{\Psi}_{t,\boldsymbol{s}}(y,z,\boldsymbol{v}) = y\cdot\boldsymbol{\psi}_{t,\boldsymbol{s}}(z,\boldsymbol{v})$ defined in (15). Plugging this result back into (32), we obtain that

$$\begin{aligned}
\widehat{\beta}_2(t,\boldsymbol{s}) &= \frac{1}{h}\cdot\boldsymbol{e}_2^T\left[\boldsymbol{M}_q^{-1}+\boldsymbol{A}_{h,b}\right]\mathbb{P}_n\left(\frac{1}{hb^d}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)\\
&= \frac{1}{h\cdot p(t,\boldsymbol{s})}\cdot\mathbb{P}_n\left(\frac{1}{hb^d}\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right) + \frac{1}{h\cdot p(t,\boldsymbol{s})}\cdot\mathbb{P}_n\left(\frac{1}{hb^d}\boldsymbol{e}_2^T\boldsymbol{A}_{h,b}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right).
\end{aligned}$$

Equivalently, we can scale and center $\widehat{\beta}_2(t,\boldsymbol{s})$ as:

(33)

$$\begin{aligned}
&\sqrt{nh^3b^d}\left\{\widehat{\beta}_2(t,\boldsymbol{s})-\mathbb{E}\left[\widehat{\beta}_2(t,\boldsymbol{s})\right]\right\}\\
&=\sqrt{nhb^d}\left[\mathbb{P}_n\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)-\mathrm{P}\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)\right]\\
&\quad+\sqrt{nhb^d}\left[\mathbb{P}_n\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\boldsymbol{e}_2^T\boldsymbol{A}_{h,b}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)-\mathrm{P}\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\boldsymbol{e}_2^T\boldsymbol{A}_{h,b}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)\right]\\
&=\sqrt{hb^d}\cdot\mathbb{G}_n\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)+\sqrt{hb^d}\cdot\mathbb{G}_n\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\boldsymbol{e}_2^T\boldsymbol{A}_{h,b}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right),
\end{aligned}$$

where $\mathbb{G}_n f=\sqrt{n}\,(\mathbb{P}_n-\mathrm{P})\,f=\frac{1}{\sqrt{n}}\sum_{k=1}^n\left\{f(Y_k,T_k,\boldsymbol{S}_k)-\mathbb{E}\left[f(Y_k,T_k,\boldsymbol{S}_k)\right]\right\}$. One auxiliary result that we derive here is a form of the uniform Bahadur representation (Bahadur, 1966; Kong et al., 2010) as:

$$\begin{aligned}
&\sup_{(t,\boldsymbol{s})\in\mathcal{E}}\left|\frac{\sqrt{nh^3b^d}\left\{\widehat{\beta}_2(t,\boldsymbol{s})-\mathbb{E}\left[\widehat{\beta}_2(t,\boldsymbol{s})\right]\right\}-\sqrt{hb^d}\cdot\mathbb{G}_n\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)}{\sqrt{hb^d}\cdot\mathbb{G}_n\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\boldsymbol{e}_2^T\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)}\right|\\
&=O\left(\max\{h,b\}\right)+O_P\left(\sqrt{\frac{|\log(hb^d)|}{nhb^d}}\right),
\end{aligned}$$

because each entry of $\boldsymbol{A}_{h,b}\in\mathbb{R}^{(q+1+d)\times(q+1+d)}$ is uniformly of the order $O\left(\max\{h,b\}\right)+O_P\left(\sqrt{\frac{|\log(hb^d)|}{nhb^d}}\right)$.

We remain to derive the uniform rate of convergence for $\sqrt{hb^d}\cdot\mathbb{G}_n\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\cdot\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)$ and $\sqrt{hb^d}\cdot\mathbb{G}_n\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\cdot\boldsymbol{e}_2^T\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)$. Recall from our notation in (15) that

$$\begin{aligned}
&\sqrt{hb^d}\cdot\mathbb{G}_n\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\cdot\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)\\
&=\frac{1}{\sqrt{nhb^d}\cdot p(t,\boldsymbol{s})}\sum_{k=1}^n\left\{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{s}}(Y_k,T_k,\boldsymbol{S}_k)-\mathbb{E}\left[\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{s}}(Y_k,T_k,\boldsymbol{S}_k)\right]\right\}\\
&=\frac{1}{\sqrt{nhb^d}\cdot p(t,\boldsymbol{s})}\sum_{k=1}^n\left\{Y_k\cdot\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\psi}_{t,\boldsymbol{s}}(T_k,\boldsymbol{S}_k)-\mathbb{E}\left[Y_k\cdot\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\psi}_{t,\boldsymbol{s}}(T_k,\boldsymbol{S}_k)\right]\right\}.
\end{aligned}$$

Since each entry of $(z,\boldsymbol{v})\mapsto\boldsymbol{\psi}_{t,\boldsymbol{s}}(z,\boldsymbol{v})$ is simply a linear combination of functions in $\mathcal{K}_{q,d}$ defined in Assumption A6(b), we can apply Theorem 4 in Einmahl and Mason (2005) to establish that

(34)

$$\begin{aligned}
&\sup_{(t,\boldsymbol{s})\in\mathcal{E}}\left|\sqrt{hb^d}\cdot\mathbb{G}_n\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\cdot\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)\right|\\
&=\sup_{(t,\boldsymbol{s})\in\mathcal{E}}\left|\frac{1}{\sqrt{nhb^d}\cdot p(t,\boldsymbol{s})}\sum_{k=1}^n\left\{Y_k\cdot\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\psi}_{t,\boldsymbol{s}}(T_k,\boldsymbol{S}_k)-\mathbb{E}\left[Y_k\cdot\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\psi}_{t,\boldsymbol{s}}(T_k,\boldsymbol{S}_k)\right]\right\}\right|\\
&=O_P\left(\sqrt{|\log(hb^d)|}\right)
\end{aligned}$$

when $\frac{|\log(hb^d)|}{\log\log n} \to \infty$. Similarly, we have that

$$\sup_{(t,\boldsymbol{s})\in\mathcal{E}} \left|\sqrt{hb^d}\cdot\mathbb{G}_n\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\cdot\boldsymbol{e}_2^T\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)\right| = O_P\left(\sqrt{|\log(hb^d)|}\right).$$

Plugging these two rates of convergence back into (33), we obtain that as $h,b\to 0$ and $\frac{nh^3b^d}{|\log(hb^d)|}\to\infty$,

(35)

$$\begin{aligned}
&\sup_{(t,\boldsymbol{s})\in\mathcal{E}}\left|\widehat{\beta}_2(t,\boldsymbol{s})-\mathbb{E}\left[\widehat{\beta}_2(t,\boldsymbol{s})\right]\right| \\
&= \sup_{(t,\boldsymbol{s})\in\mathcal{E}}\left|\frac{\sqrt{hb^d}\cdot\mathbb{G}_n\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\cdot\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)+\sqrt{hb^d}\cdot\mathbb{G}_n\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\cdot\boldsymbol{e}_2^T\boldsymbol{A}_{h,b}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)}{\sqrt{nh^3b^d}}\right| \\
&\le \sup_{(t,\boldsymbol{s})\in\mathcal{E}}\left|\frac{\sqrt{hb^d}\cdot\mathbb{G}_n\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\cdot\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)}{\sqrt{nh^3b^d}}\right| + \sup_{(t,\boldsymbol{s})\in\mathcal{E}}\left|\frac{\sqrt{hb^d}\cdot\mathbb{G}_n\left(\frac{1}{hb^d\cdot p(t,\boldsymbol{s})}\cdot\boldsymbol{e}_2^T\boldsymbol{A}_{h,b}\boldsymbol{\Psi}_{t,\boldsymbol{s}}\right)}{\sqrt{nh^3b^d}}\right| \\
&= O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}}\right)+O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}}\right)\left[O\left(\max\{h,b\}\right)+O_P\left(\sqrt{\frac{|\log(hb^d)|}{nhb^d}}\right)\right] \\
&= O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}}\right).
\end{aligned}$$

Together with the rate of convergence for the bias term $\mathbb{E}\left[\widehat{\beta}_2(t,\boldsymbol{s})\right]-\beta_2(t,\boldsymbol{s})$ in Lemma 2, we conclude that

$$\sup_{(t,\boldsymbol{s})\in\mathcal{E}}\left|\widehat{\beta}_2(t,\boldsymbol{s})-\beta_2(t,\boldsymbol{s})\right| = O\left(h^q+b^2+\frac{\max\{b,h\}^4}{h}\right)+O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}}\right),$$

where we combine the odd and even cases for $q>0$ by arguing that $O(h^{q+1})$ is dominated by $O(h^q)$ when h is small. The proof is thus completed. ☐

B.4. Proof of Theorem 4.

THEOREM 4 (Convergence of $\widehat{\theta}_C(t)$ and $\widehat{m}_\theta(t)$). *Let $q>0$ and $\mathcal{T}'\subset\mathcal{T}$ be a compact set so that $p_T(t)$ is uniformly bounded away from 0 within $\mathcal{T}'$. Suppose that Assumptions A1, A2, A3, A4, A5, and A6 hold. Then, as $h,b,\hbar,\frac{\max\{h,b\}^4}{h}\to 0$ and $\frac{nh^3b^d}{|\log(hb^d)|},\frac{|\log(hb^d)|}{\log\log n},\frac{n\hbar}{|\log\hbar|},\frac{|\log\hbar|}{\log\log n}\to\infty$, we know that*

$$\sup_{t\in\mathcal{T}'}\left|\widehat{\theta}_C(t)-\theta_C(t)\right| = O\left(h^q+b^2+\frac{\max\{b,h\}^4}{h}\right)+O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}}+\hbar^2+\sqrt{\frac{|\log\hbar|}{n\hbar}}\right)$$

and

$$\begin{aligned}
\sup_{t\in\mathcal{T}'}\left|\widehat{m}_\theta(t)-m(t)\right| &= O_P\left(\frac{1}{\sqrt{n}}\right)+O\left(h^q+b^2+\frac{\max\{b,h\}^4}{h}\right) \\
&\quad + O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}}+\hbar^2+\sqrt{\frac{|\log\hbar|}{n\hbar}}\right).
\end{aligned}$$

PROOF OF THEOREM 4. Assume, without loss of generality, that the set $\mathcal{T}'$ is connected. Otherwise, we focus our analysis on a connected component of $\mathcal{T}'$ and take the union afterwards.

Recall from (8) and (16) that the integral estimator is defined as:

$$\widehat{m}_\theta(t) = \frac{1}{n}\sum_{i=1}^n \left[Y_i + \int_{\widetilde{t}=T_i}^{\widetilde{t}=t} \widehat{\theta}_C(\widetilde{t})\, d\widetilde{t}\right] = \frac{1}{n}\sum_{i=1}^n Y_i + \widehat{\Delta}_{h,b}(t).$$

Moreover, under Assumption A2,

$$m(t) = \mathbb{E}\left[\mu(t,\boldsymbol{S})\right] = \mathbb{E}\left[\mu(T,\boldsymbol{S})\right] + \Delta(t),$$

where we define

$$\Delta(t) \equiv \mathbb{E}\left[\int_{\widetilde{t}=T}^{\widetilde{t}=t} \theta(\widetilde{t}) d\widetilde{t}\right] = \int \int_{\widetilde{t}=\tau}^{\widetilde{t}=t} \theta(\widetilde{t})\, d\widetilde{t}\, d\mathrm{P}_T(\tau) = \int \int_{\widetilde{t}=\tau}^{\widetilde{t}=t} \theta_C(\widetilde{t})\, d\widetilde{t}\, d\mathrm{P}_T(\tau)$$

as the population version of $\widehat{\Delta}_{h,b}(t)$ in (16) under Assumption A2. Here, P_T is the marginal probability distribution function of T. Under the boundedness of $\mu(t,\boldsymbol{s})$ by Assumption A3, we know that

$$\frac{1}{n}\sum_{i=1}^n Y_i - \mathbb{E}\left[\mu(T,\boldsymbol{S})\right] = O_P\left(\frac{1}{\sqrt{n}}\right).$$

Recall from (12) that $\widehat{\theta}_C(t) = \int \widehat{\beta}_2(t,\boldsymbol{s}) d\widehat{P}(\boldsymbol{s}|t)$ and $\theta(t) = \theta_C(t) = \int \beta_2(t,\boldsymbol{s}) dP(\boldsymbol{s}|t)$ with $\beta_2(t,\boldsymbol{s}) = \frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$. Then, we have the following decomposition as:

(36)

$$\begin{aligned}\widehat{\Delta}_{h,b}(t) - \Delta(t) &= \frac{1}{n}\sum_{i=1}^n \int_{\widetilde{t}=T_i}^{\widetilde{t}=t} \widehat{\theta}_C(\widetilde{t}) d\widetilde{t} - \int \int_{\widetilde{t}=\tau}^{\widetilde{t}=t} \theta_C(\widetilde{t}) d\widetilde{t} d\mathrm{P}_T(\tau) \\ &= \underbrace{\int \int_{\widetilde{t}=\tau}^{\widetilde{t}=t} \theta_C(\widetilde{t}) d\widetilde{t}\left[d\mathbb{P}_{n,T}(\tau) - d\mathrm{P}_T(\tau)\right]}_{\textbf{Term (I)}} + \underbrace{\int \int_{\widetilde{t}=\tau}^{\widetilde{t}=t} \left[\widehat{\theta}_C(\widetilde{t}) - \theta_C(\widetilde{t})\right] d\widetilde{t} d\mathbb{P}_{n,T}(\tau)}_{\textbf{Term (II)}},\end{aligned}$$

where $\mathbb{P}_{n,T}$ is the empirical measure associated with $\{T_1,...,T_n\}$.

- **Term (I):** By Dvoretzky-Kiefer-Wolfowitz inequality (Dvoretzky et al., 1956; Massart, 1990), we know that

$$\sup_{t\in\mathcal{T}'} \left|\mathbb{P}_{n,T}(t) - \mathrm{P}_T(t)\right| = O_P\left(\frac{1}{\sqrt{n}}\right).$$

In addition, under Assumption A3 and the compactness of the support $\mathcal{T}'$, we have that

$$\int \left[\int_{\widetilde{t}=\tau}^{\widetilde{t}=t} \theta_C(\widetilde{t}) d\widetilde{t}\right]^2 d\mathrm{P}_T(\tau) \leq \int |t-\tau|^2 \sup_{t\in\mathcal{T}} |\theta_C(t)|^2\, d\mathrm{P}_T(\tau) < \infty$$

for any $t \in \mathcal{T}$. Thus, the (uniform) rate of convergence of Term (I) in (36) is

$$\sup_{t\in\mathcal{T}'} \int \int_{\widetilde{t}=\tau}^{\widetilde{t}=t} \theta_C(\widetilde{t}) d\widetilde{t}\left[d\mathbb{P}_{n,T}(\tau) - d\mathrm{P}_T(\tau)\right] = O_P\left(\frac{1}{\sqrt{n}}\right).$$

• **Term (II):** We first focus on establishing the (uniform) rate of convergence for $\widehat{\theta}_C(t) - \theta_C(t)$. For any $t \in \mathcal{T}$, we note that

$$
\begin{aligned}
&\widehat{\theta}_C(t) - \theta_C(t) \\
&= \int \widehat{\beta}_2(t,\boldsymbol{s}) d\widehat{P}(\boldsymbol{s}|t) - \int \beta_2(t,\boldsymbol{s}) dP(\boldsymbol{s}|t) \\
&= \frac{\sum_{i=1}^n \left[\widehat{\beta}_2(t,\boldsymbol{S}_i) - \beta_2(t,\boldsymbol{S}_i)\right] \bar{K}_T\left(\frac{t-T_i}{\hbar}\right)}{\sum_{j=1}^n \bar{K}_T\left(\frac{t-T_j}{\hbar}\right)} + \frac{\sum_{i=1}^n \beta_2(t,\boldsymbol{S}_i)\cdot \bar{K}_T\left(\frac{t-T_i}{\hbar}\right)}{\sum_{j=1}^n \bar{K}_T\left(\frac{t-T_j}{\hbar}\right)} - \int \beta_2(t,\boldsymbol{s}) dP(\boldsymbol{s}|t) \\
&= \underbrace{\sum_{(t,\boldsymbol{S}_i)\in \mathcal{E}\cap(\partial\mathcal{E}\ominus\hbar)} \frac{\left[\widehat{\beta}_2(t,\boldsymbol{S}_i) - \beta_2(t,\boldsymbol{S}_i)\right] \bar{K}_T\left(\frac{t-T_i}{\hbar}\right)}{\sum_{j=1}^n \bar{K}_T\left(\frac{t-T_j}{\hbar}\right)}}_{\textbf{Term A}} + \underbrace{\sum_{(t,\boldsymbol{S}_i)\in \partial\mathcal{E}\oplus\hbar} \frac{\left[\widehat{\beta}_2(t,\boldsymbol{S}_i) - \beta_2(t,\boldsymbol{S}_i)\right] \bar{K}_T\left(\frac{t-T_i}{\hbar}\right)}{\sum_{j=1}^n \bar{K}_T\left(\frac{t-T_j}{\hbar}\right)}}_{\textbf{Term B}} \\
&\quad + \underbrace{\int \beta_2(t,\boldsymbol{s}) \left[d\widehat{P}(\boldsymbol{s}|t) - dP(\boldsymbol{s}|t)\right]}_{\textbf{Term C}},
\end{aligned}
$$

where $\partial\mathcal{E}\ominus\hbar = \left\{\boldsymbol{z}\in\mathbb{R}^{d+1} : \inf_{\boldsymbol{x}\in\partial\mathcal{E}} ||\boldsymbol{z}-\boldsymbol{x}||_2 \geq \hbar\right\}$ and $\partial\mathcal{E}\oplus\hbar = \left\{\boldsymbol{z}\in\mathbb{R}^{d+1} : \inf_{\boldsymbol{x}\in\partial\mathcal{E}} ||\boldsymbol{z}-\boldsymbol{x}||_2 \leq \hbar\right\}$.

As for **Term A**, we know from Lemma 3 that

$$
\sup_{(t,\boldsymbol{s})\in\mathcal{E}} \left|\widehat{\beta}_2(t,\boldsymbol{s}) - \beta_2(t,\boldsymbol{s})\right| = O\left(h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right) + O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}}\right)
$$

when $q>0$ is even. Hence, it has the uniform rate of convergence as:

$$
\begin{aligned}
\textbf{Term A} &\leq \left[\sup_{(t,\boldsymbol{s})\in\mathcal{E}} \left|\widehat{\beta}_2(t,\boldsymbol{s}) - \beta_2(t,\boldsymbol{s})\right|\right] \sum_{(t,\boldsymbol{S}_i)\in\mathcal{E}\cap(\partial\mathcal{E}\ominus\hbar)} \frac{\bar{K}_T\left(\frac{t-T_i}{\hbar}\right)}{\sum_{j=1}^n \bar{K}_T\left(\frac{t-T_j}{\hbar}\right)} \\
&= O\left(h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right) + O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}}\right)
\end{aligned}
$$

when $q>0$ is even, $h,b,\hbar$, $\frac{\max\{b,h\}^4}{h}\to 0$, and $\frac{nh^3b^d}{|\log(hb^d)|}, \frac{|\log(hb^d)|}{\log\log n}, \frac{n\hbar}{|\log\hbar|}, \frac{|\log\hbar|}{\log\log n}\to\infty$.

As for **Term B**, we leverage the compact support of $\bar{K}_T$ (Assumption A6(c)) to argue that $|t-T_i| = O(\hbar)$ for any $(t,\boldsymbol{S}_i)\in\partial\mathcal{E}\oplus\hbar$. Then, under Assumptions A3 and A6(a), we know that both $\widehat{\beta}_2(t,\boldsymbol{s})$ and $\beta_2(t,\boldsymbol{s})$ are Lipschitz so that

$$
\begin{aligned}
\left|\widehat{\beta}_2(t,\boldsymbol{S}_i) - \beta_2(t,\boldsymbol{S}_i)\right| &\leq \left|\widehat{\beta}_2(t,\boldsymbol{S}_i) - \widehat{\beta}_2(T_i,\boldsymbol{S}_i)\right| + \left|\widehat{\beta}_2(T_i,\boldsymbol{S}_i) - \beta_2(T_i,\boldsymbol{S}_i)\right| + \left|\beta_2(T_i,\boldsymbol{S}_i) - \beta_2(t,\boldsymbol{S}_i)\right| \\
&\leq O_P(\hbar) + O\left(h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right) + O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}}\right)
\end{aligned}
$$

when $q>0$ is even under Lemma 3. This upper bound continues to hold if we take the supremum over $\partial\mathcal{E}\oplus\hbar$. In addition, by Assumption A5(c), we know that the proportion of $(t,\boldsymbol{S}_i), i=1,...,n$ that lie in $\partial\mathcal{E}\oplus\hbar$ is of order $O_P(\hbar)$. Thus, $\sum_{(t,\boldsymbol{S}_i)\in\partial\mathcal{E}\oplus\hbar} \frac{\bar{K}_T\left(\frac{t-T_i}{\hbar}\right)}{\sum_{j=1}^n \bar{K}_T\left(\frac{t-T_j}{\hbar}\right)} = O_P(\hbar)$ and

$$
\textbf{Term B} \leq \sum_{(t,\boldsymbol{S}_i)\in\partial\mathcal{E}\oplus\hbar} \frac{\left|\widehat{\beta}_2(t,\boldsymbol{S}_i) - \beta_2(t,\boldsymbol{S}_i)\right| \bar{K}_T\left(\frac{t-T_i}{\hbar}\right)}{\sum_{j=1}^n \bar{K}_T\left(\frac{t-T_j}{\hbar}\right)}
$$

$$\leq \left[O_P(\hbar) + O\left(h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right) + O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}}\right)\right] \cdot O_P(\hbar)$$
$$= O_P\left(\hbar^2 + h^q\hbar + b^2\hbar + \frac{\max\{b,h\}^4\hbar}{h} + \sqrt{\frac{|\log(hb^d)|\hbar}{nh^3b^d}}\right)$$

when $q > 0$ is even, $h, b, \hbar, \frac{\max\{b,h\}^4}{h} \to 0$, and $\frac{nh^3b^d}{|\log(hb^d)|}, \frac{|\log(hb^d)|}{\log\log n}, \frac{n\hbar}{|\log\hbar|}, \frac{|\log\hbar|}{\log\log n} \to \infty$.

As for **Term C**, under Assumptions A3 and A6(d), we utilize Theorem 3 in Einmahl and Mason (2005) to derive its uniform rate of convergence as:

$$\textbf{Term C} \leq \sup_{(t,\boldsymbol{s})\in\mathcal{T}\times\mathcal{S}} |\beta_2(t,\boldsymbol{s})| \cdot \sup_{(t,\boldsymbol{s})\in\mathcal{T}'\times\mathcal{S}} \left|\widehat{P}(\boldsymbol{s}|t) - P(\boldsymbol{s}|t)\right| = O(\hbar^2) + O_P\left(\sqrt{\frac{|\log\hbar|}{n\hbar}}\right).$$

Therefore,

$$\sup_{t\in\mathcal{T}'}\left|\widehat{\theta}_C(t) - \theta_C(t)\right|$$
$$= O\left(h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right) + O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}} + \hbar^2 + \sqrt{\frac{|\log\hbar|}{n\hbar}}\right)$$

when q is even, $h, b, \hbar, \frac{\max\{b,h\}^4}{h} \to 0$, and $\frac{nh^3b^d}{|\log(hb^d)|}, \frac{|\log(hb^d)|}{\log\log n}, \frac{n\hbar}{|\log\hbar|}, \frac{|\log\hbar|}{\log\log n} \to \infty$.

Finally, plugging the results back into (36), we obtain that

$$\sup_{t\in\mathcal{T}'}\left|\widehat{\Delta}_{h,b,\hbar}(t) - \Delta(t)\right|$$
$$\leq \sup_{t\in\mathcal{T}'}\left|\int\int_{\tilde{t}=\tau}^{\tilde{t}=t} \theta_C(\tilde{t})d\tilde{t}\left[d\mathbb{P}_{n,T}(\tau) - d\mathrm{P}_T(\tau)\right]\right| + \sup_{t\in\mathcal{T}'}\left|\frac{1}{n}\sum_{i=1}^n\int_{T_i}^t\left[\widehat{\theta}_C(\tilde{t}) - \theta_C(\tilde{t})\right]d\tilde{t}\right|$$
$$\leq \sup_{t\in\mathcal{T}'}\left|\int\int_{\tilde{t}=\tau}^{\tilde{t}=t} \theta_C(\tilde{t})d\tilde{t}\left[d\mathbb{P}_{n,T}(\tau) - d\mathrm{P}_T(\tau)\right]\right| + \sup_{t_1,t_2\in\mathcal{T}}|t_1-t_2| \cdot \sup_{\tilde{t}\in\mathcal{T}'}\left|\widehat{\theta}_C(\tilde{t}) - \theta_C(\tilde{t})\right|$$
$$= O_P\left(\frac{1}{\sqrt{n}}\right) + O\left(h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right) + O_P\left(\sqrt{\frac{|\log(hb^d)|}{nh^3b^d}} + \hbar^2 + \sqrt{\frac{|\log\hbar|}{n\hbar}}\right),$$

where $\sup_{t_1,t_2\in\mathcal{T}}|t_1-t_2| < \infty$ due to the compactness of $\mathcal{T} = \mathrm{proj}_T(\mathcal{E})$. The result follows. $\square$

B.5. Proof of Lemma 5.

LEMMA 5 (Asymptotic linearity). *Let $q \geq 2$ in the local polynomial regression for estimating $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ and $\mathcal{T}' \subset \mathcal{T}$ be a compact set so that $p_T(t)$ is uniformly bounded away from 0 within $\mathcal{T}'$. Suppose that Assumptions A1, A2, A3, A4, A5, and A6 hold. Then, if $h \asymp b \asymp n^{-\frac{1}{\gamma}}$ and $\hbar \asymp n^{-\frac{1}{\varpi}}$ for some $\gamma \geq \varpi > 0$ such that $\frac{nh^{d+5}}{\log n} \to c_1$ and $\frac{n\hbar^5}{\log n} \to c_2$ for some finite number $c_1, c_2 \geq 0$ and $\frac{\log n}{n\hbar^2}, \frac{h^{d+3}\log n}{\hbar}, \frac{h^{d+3}}{\hbar^2} \to 0$ as $n \to \infty$, then for any $t \in \mathcal{T}'$, we have that*

$$\sqrt{nh^3b^d}\left[\widehat{\theta}_C(t) - \theta_C(t)\right] = \mathbb{G}_n\bar{\varphi}_t + o_P(1) \quad \textit{and} \quad \sqrt{nh^3b^d}\left[\widehat{m}_\theta(t) - m(t)\right] = \mathbb{G}_n\varphi_t + o_P(1),$$

where $\bar{\varphi}_t(Y,T,\boldsymbol{S}) = \mathbb{E}_{(T_{i_3},\boldsymbol{S}_{i_3})}\left[\frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{S}_{i_3}}(Y,T,\boldsymbol{S})}{\sqrt{hb^d}\cdot p(t,\boldsymbol{S}_{i_3})\cdot p_T(t)}\cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{t-T_{i_3}}{\hbar}\right)\right]$ *and*

$$\begin{aligned}\varphi_t\left(Y,T,\boldsymbol{S}\right) &= \mathbb{E}_{T_{i_2}}\left[\int_{T_{i_2}}^t \bar{\varphi}_{\tilde{t}}(Y,T,\boldsymbol{S})\,d\tilde{t}\right]\\ &= \mathbb{E}_{T_{i_2}}\left\{\int_{T_{i_2}}^t \mathbb{E}_{(T_{i_3},\boldsymbol{S}_{i_3})}\left[\frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\tilde{t},\boldsymbol{S}_{i_3}}(Y,T,\boldsymbol{S})}{\sqrt{hb^d}\cdot p(\tilde{t},\boldsymbol{S}_{i_3})\cdot p_T(\tilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\tilde{t}-T_{i_3}}{\hbar}\right)\right]d\tilde{t}\right\}.\end{aligned}$$

Furthermore, we have the following uniform results as:

$$\left|\sqrt{nh^3b^d}\sup_{t\in\mathcal{T}'}\left|\widehat{\theta}_C(t)-\theta_C(t)\right| - \sup_{t\in\mathcal{T}'}\left|\mathbb{G}_n\bar{\varphi}_t\right|\right| = O_P\left(\frac{\log n}{\sqrt{n\hbar}}+\sqrt{nh^{d+7}}+\sqrt{\frac{h^{d+3}\log n}{\hbar}}\right)$$

and

$$\begin{aligned}&\left|\sqrt{nh^3b^d}\sup_{t\in\mathcal{T}'}\left|\widehat{m}_\theta(t)-m(t)\right| - \sup_{t\in\mathcal{T}'}\left|\mathbb{G}_n\varphi_t\right|\right|\\ &= O_P\left(\sqrt{nh^{d+7}}+\sqrt{\frac{\log n}{n\hbar^2}}+\sqrt{\frac{h^{d+3}\log n}{\hbar}}+\sqrt{\frac{h^{d+3}}{\hbar^2}}\right).\end{aligned}$$

PROOF OF LEMMA 5. The entire proof consists of three steps, where we first tackle the asymptotic linearity of $\widehat{m}_\theta(t)$ and then inherit the arguments to establish the simpler asymptotic linearity of $\widehat{\theta}_C(t)$. **Step 1** establishes the asymptotic linearity of $\sqrt{nh^3b^d}\left[\widehat{m}_\theta(t)-m(t)\right]$ for any $t\in\mathcal{T}'$ by writing it as a V-statistic. **Step 2** derives the exact rate of convergence for the coupling between $\sqrt{nh^3b^d}\sup_{t\in\mathcal{T}'}\left|\widehat{m}_\theta(t)-m(t)\right|$ and $\sup_{t\in\mathcal{T}'}\left|\mathbb{G}_n\varphi_t\right|$. **Step 3** adopts the arguments in previous two steps to prove the asymptotic linearity of $\sqrt{nh^3b^d}\left[\widehat{\theta}_C(t)-\theta_C(t)\right]$.

• **Step 1:** Recalling from (8) and (16) for the definition of the integral estimator $\widehat{m}_\theta(t)$, we know that

$$\begin{aligned}\widehat{m}_\theta(t)-m(t) &= \frac{1}{n}\sum_{i=1}^n\left[Y_i+\int_{\tilde{t}=T_i}^{\tilde{t}=t}\widehat{\theta}_C(\tilde{t})d\tilde{t}\right]-\mathbb{E}\left[\mu(t,\boldsymbol{S})\right]\\ &= \underbrace{\frac{1}{n}\sum_{i=1}^n Y_i-\mathbb{E}\left[\mu(T,\boldsymbol{S})\right]}_{\textbf{Term I}}+\underbrace{\widehat{\Delta}_{h,b}(t)-\mathbb{E}\left[\int_T^t\theta_C(\tilde{t})\,d\tilde{t}\right]}_{\textbf{Term II}},\end{aligned}$$

where we use (5) in Assumption A2 to obtain the second equality above.

As for **Term I**, we know that its rate of convergence is $O_P\left(\frac{1}{\sqrt{n}}\right)$ under the boundedness of $\mu(t,\boldsymbol{s})$ by Assumption A3 and will be asymptotically negligible compared with **Term II**. Furthermore, **Term I** is independent of the choice of $t\in\mathcal{T}'$.

As for **Term II**, recall the notations from (15) and (33) in the proof of Lemma 3 that

Term II

$$= \frac{1}{n}\sum_{i=1}^n\int_{T_i}^t\widehat{\theta}_C(\tilde{t})\,d\tilde{t}-\mathbb{E}\left[\int_T^t\theta_C(\tilde{t})\,d\tilde{t}\right]$$

$$
\begin{aligned}
&= \frac{1}{n}\sum_{i=1}^{n}\int_{T_i}^{t} \frac{\sum_{k=1}^{n}\widehat{\beta}_2(\widetilde{t},\boldsymbol{S}_k)\cdot \bar{K}_T\left(\frac{\widetilde{t}-T_k}{\hbar}\right)}{\sum_{j=1}^{n}\bar{K}_T\left(\frac{\widetilde{t}-T_j}{\hbar}\right)}\, d\widetilde{t} - \mathbb{E}\left[\int_T^t \theta_C(\widetilde{t})\, d\widetilde{t}\right]\\
&\overset{\text{(i)}}{=} \frac{1}{n^2}\sum_{i=1}^{n}\int_{T_i}^{t}\sum_{k=1}^{n} \frac{\mathbb{P}_n\left(\frac{1}{hb^d}\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_k}\right)}{h\cdot p(\widetilde{t},\boldsymbol{S}_k)}\cdot \frac{\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_k}{\hbar}\right)}{\frac{1}{n\hbar}\sum_{j=1}^{n}\bar{K}_T\left(\frac{\widetilde{t}-T_j}{\hbar}\right)}\, d\widetilde{t} - \mathbb{E}\left[\int_T^t \theta_C(\widetilde{t})\, d\widetilde{t}\right]\\
&\overset{\text{(ii)}}{=} \frac{1}{n^2}\sum_{i=1}^{n}\int_{T_i}^{t}\sum_{k=1}^{n} \frac{\mathbb{P}_n\left(\frac{1}{hb^d}\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_k}\right)}{h\cdot p(\widetilde{t},\boldsymbol{S}_k)\cdot p_T(\widetilde{t})}\cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_k}{\hbar}\right) d\widetilde{t} - \mathbb{E}\left[\int_T^t \theta_C(\widetilde{t})\, d\widetilde{t}\right]\\
&\quad + \left[O_P(\hbar^2)+O_P\left(\sqrt{\frac{\log n}{n\hbar}}\right)\right]\left[O_P(1)+O_P\left(\frac{1}{\sqrt{n}}\right)+O\left(h^q+b^2+\frac{\max\{h,b\}^4}{h}\right)\right.\\
&\qquad\qquad\qquad \left. + O_P\left(\sqrt{\frac{\log n}{nh^3b^d}}+\hbar^2+\sqrt{\frac{\log n}{n\hbar}}\right)\right]\\
&= \frac{1}{n^3}\sum_{i_1=1}^{n}\sum_{i_2=1}^{n}\sum_{i_3=1}^{n}\int_{T_{i_1}}^{t} \frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_2}}(Y_{i_3},T_{i_3},\boldsymbol{S}_{i_3})}{h^2b^d\cdot p(\widetilde{t},\boldsymbol{S}_{i_2})\cdot p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right) d\widetilde{t} - \mathbb{E}\left[\int_T^t \theta_C(\widetilde{t})\, d\widetilde{t}\right]\\
&\quad + O_P\left(\hbar^2+\sqrt{\frac{\log n}{n\hbar}}\right),
\end{aligned}
$$

where the equality (i) only keeps the dominating term $\frac{1}{h}\mathbb{P}_n\left(\frac{1}{hb^d}\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_k}\right)$ from (33) and uses the result (35), the equality (ii) leverages the standard (uniform) rate of convergence for the kernel density estimator $\frac{1}{n\hbar}\sum_{j=1}^{n}\bar{K}_T\left(\frac{t-T_j}{\hbar}\right) = O\left(\hbar^2\right)+O_P\left(\sqrt{\frac{\log n}{n\hbar}}\right)$ for all $t\in\mathcal{T}'$ as well as uses our condition $h\asymp b\asymp n^{-\frac{1}{\gamma}}$ and Theorem 4 to argue that the rate of convergence for the first term in (ii) is given by

$$
O_P(1)+O_P\left(\frac{1}{\sqrt{n}}\right)+O\left(h^q+b^2+\frac{\max\{h,b\}^4}{h}\right)+O_P\left(\sqrt{\frac{\log n}{nh^3b^d}}+\hbar^2+\sqrt{\frac{\log n}{n\hbar}}\right).
$$

Recall that we denote the data sample by $\boldsymbol{U}_i=(Y_i,T_i,\boldsymbol{S}_i), i=1,...,n$ with $\mathbb{P}_n$ as the empirical measure. To deal with the triple summation term in the above display, we define via notations in (15) that

$$
\begin{aligned}
\lambda_t(\boldsymbol{U}_{i_1},\boldsymbol{U}_{i_2},\boldsymbol{U}_{i_3}) &\equiv \int_{T_{i_1}}^{t}\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_2}}(Y_{i_3},T_{i_3},\boldsymbol{S}_{i_3})}{h^2b^d\cdot p(\widetilde{t},\boldsymbol{S}_{i_2})\cdot p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right) d\widetilde{t}\\
&= \int_{T_{i_1}}^{t}\frac{Y_{i_3}\cdot \boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\psi}_{\widetilde{t},\boldsymbol{S}_{i_2}}(T_{i_3},\boldsymbol{S}_{i_3})}{h^2b^d\cdot p(\widetilde{t},\boldsymbol{S}_{i_2})\cdot p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right) d\widetilde{t}
\end{aligned}
$$

for any $t\in\mathcal{T}$. We also symmetrize it by considering

$$
\begin{aligned}
\Lambda_t(\boldsymbol{U}_{i_1},\boldsymbol{U}_{i_2},\boldsymbol{U}_{i_3}) \equiv \frac{1}{6}\Big[&\lambda_t(\boldsymbol{U}_{i_1},\boldsymbol{U}_{i_2},\boldsymbol{U}_{i_3})+\lambda_t(\boldsymbol{U}_{i_1},\boldsymbol{U}_{i_3},\boldsymbol{U}_{i_2})+\lambda_t(\boldsymbol{U}_{i_2},\boldsymbol{U}_{i_1},\boldsymbol{U}_{i_3})+\lambda_t(\boldsymbol{U}_{i_2},\boldsymbol{U}_{i_3},\boldsymbol{U}_{i_1})\\
&+\lambda_t(\boldsymbol{U}_{i_3},\boldsymbol{U}_{i_1},\boldsymbol{U}_{i_2})+\lambda_t(\boldsymbol{U}_{i_3},\boldsymbol{U}_{i_2},\boldsymbol{U}_{i_1})\Big].
\end{aligned}
$$

Under this notation, we know that **Term II** is a V-statistic with some small bias terms under a symmetric "kernel" $(\boldsymbol{U}_{i_1}, \boldsymbol{U}_{i_2}, \boldsymbol{U}_{i_3}) \mapsto \Lambda_t(\boldsymbol{U}_{i_1}, \boldsymbol{U}_{i_2}, \boldsymbol{U}_{i_3})$. By Pascal's rule, we obtain that

$$
\begin{aligned}
&\widehat{m}_\theta(t) - m(t) \\
&= \mathbb{P}_n^3 \Lambda_t - \mathbb{E}\left[\int_T^t \theta_C(\widetilde{t})\, d\widetilde{t}\right] + O_P\left(\hbar^2 + \sqrt{\frac{\log n}{n\hbar}}\right) \\
&= \mathrm{P}^3\Lambda_t + 3\left(\mathbb{P}_n - \mathrm{P}\right)\mathrm{P}^2\Lambda_t + 3\left(\mathbb{P}_n - \mathrm{P}\right)^2 \mathrm{P}\Lambda_t + \left(\mathbb{P}_n - \mathrm{P}\right)^3\Lambda_t - \mathbb{E}\left[\int_T^t \theta_C(\widetilde{t})\, d\widetilde{t}\right] \\
&\quad + O_P\left(\hbar^2 + \sqrt{\frac{\log n}{n\hbar}}\right),
\end{aligned} \tag{37}
$$

where we use the shorthand notations $\mathrm{P}\Lambda_t$ and $\mathrm{P}^2\Lambda_t$ to refer to the functions

$$
(\boldsymbol{U}_{i_1}, \boldsymbol{U}_{i_2}) \mapsto \int \Lambda_t(\boldsymbol{U}_{i_1}, \boldsymbol{U}_{i_2}, \boldsymbol{u}_{i_3})\, d\mathrm{P}(\boldsymbol{u}_{i_3}) \quad \text{and} \quad \boldsymbol{U}_{i_1} \mapsto \int\int \Lambda_t(\boldsymbol{U}_{i_1}, \boldsymbol{u}_{i_2}, \boldsymbol{u}_{i_3})\, d\mathrm{P}(\boldsymbol{u}_{i_2}) d\mathrm{P}(\boldsymbol{u}_{i_3}),
$$

respectively. In addition, based on the bias terms in the proof of Lemma 3, we know that

$$
\begin{aligned}
&\mathrm{P}^3\Lambda_t - \mathbb{E}\left[\int_T^t \theta_C(\widetilde{t})\, d\widetilde{t}\right] \\
&= \mathbb{E}_{T_{i_1}}\left\{\int_{T_{i_1}}^t \mathbb{E}_{(T_{i_2}, \boldsymbol{S}_{i_2})}\left[\frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\boldsymbol{\Psi}_{\widetilde{t}, \boldsymbol{S}_{i_2}}(Y_{i_3}, T_{i_3}, \boldsymbol{S}_{i_3})\right]}{h^2 b^d \cdot p(\widetilde{t}, \boldsymbol{S}_{i_2}) \cdot p_T(\widetilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t} - T_{i_2}}{\hbar}\right)\right] d\widetilde{t}\right\} - \mathbb{E}\left[\int_T^t \theta_C(\widetilde{t})\, d\widetilde{t}\right] \\
&= \mathbb{E}_{T_{i_1}}\left\{\int_{T_{i_1}}^t \mathbb{E}_{(T_{i_2}, \boldsymbol{S}_{i_2})}\left[\frac{\left[\beta_2(\widetilde{t}, \boldsymbol{S}_{i_2}) + O\left(h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right)\right]}{p_T(\widetilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t} - T_{i_2}}{\hbar}\right)\right] d\widetilde{t}\right\} \\
&\quad - \mathbb{E}\left[\int_T^t \theta_C(\widetilde{t})\, d\widetilde{t}\right] + O(\hbar^2) \\
&= \mathbb{E}_{T_{i_1}}\left\{\int_{T_{i_1}}^t \mathbb{E}_{(T_{i_2}, \boldsymbol{S}_{i_2})}\left[\frac{\beta_2(\widetilde{t}, \boldsymbol{S}_{i_2})}{p_T(\widetilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t} - T_{i_2}}{\hbar}\right)\right] d\widetilde{t}\right\} \\
&\quad + \mathbb{E}_{T_{i_1}}\left\{\int_{T_{i_1}}^t \mathbb{E}_{(T_{i_2}, \boldsymbol{S}_{i_2})}\left[\frac{O\left(h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right)}{p_T(\widetilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t} - T_{i_2}}{\hbar}\right)\right] d\widetilde{t}\right\} \\
&\quad - \mathbb{E}\left[\int_T^t \theta_C(\widetilde{t})\, d\widetilde{t}\right] + O(\hbar^2) \\
&\overset{\text{(iv)}}{=} \mathbb{E}_{T_{i_1}}\left[\int_{T_{i_1}}^t \frac{\int_{\mathcal{S}(\widetilde{t})} \beta_2(\widetilde{t}, \boldsymbol{s}) \cdot p(\widetilde{t}, \boldsymbol{s}) d\boldsymbol{s} + O(\hbar^2)}{p_T(\widetilde{t})}\, d\widetilde{t}\right] \\
&\quad + \left[O\left(h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right)\right] \mathbb{E}_{T_{i_1}}\left[\int_{T_{i_1}}^t \frac{p_T(\widetilde{t}) + O(\hbar^2)}{p_T(\widetilde{t})}\, d\widetilde{t}\right] - \mathbb{E}\left[\int_T^t \theta_C(\widetilde{t})\, d\widetilde{t}\right] + O(\hbar^2) \\
&= \mathbb{E}\left\{\int_T^t \mathbb{E}\left[\beta_2(\widetilde{t}, \boldsymbol{S}) | T = \widetilde{t}\right] d\widetilde{t}\right\} + O\left(\hbar^2 + h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right) - \mathbb{E}\left[\int_T^t \theta_C(\widetilde{t})\, d\widetilde{t}\right] \\
&= O\left(\hbar^2 + h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right)
\end{aligned}
$$

where (iv) leverages the following argument. For any fixed $\widetilde{t} \in \mathcal{T}'$, we compute via a similar argument in the proof of Theorem 4 that

$$\begin{aligned}
&\mathbb{E}_{(T_{i_2},\boldsymbol{S}_{i_2})}\left[\frac{\beta_2(\widetilde{t},\boldsymbol{S}_{i_2})}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right)\right]\\
&=\mathbb{E}_{(T_{i_2},\boldsymbol{S}_{i_2})}\left[\frac{\beta_2(\widetilde{t},\boldsymbol{S}_{i_2})}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right)\mathbb{1}_{\left\{(\widetilde{t},s_2)\in\mathcal{E}\cap(\partial\mathcal{E}\ominus\hbar)\right\}}\right]\\
&\quad+\mathbb{E}_{(T_{i_2},\boldsymbol{S}_{i_2})}\left[\frac{\beta_2(\widetilde{t},\boldsymbol{S}_{i_2})}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right)\mathbb{1}_{\left\{(\widetilde{t},s_2)\in\partial\mathcal{E}\oplus\hbar\right\}}\right]\\
&=\int_{\mathcal{T}\times\mathcal{S}}\frac{\beta_2(\widetilde{t},\boldsymbol{s}_2)}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-t_2}{\hbar}\right)\cdot p(t_2,\boldsymbol{s}_2)\cdot\mathbb{1}_{\left\{(\widetilde{t},s_2)\in\mathcal{E}\cap(\partial\mathcal{E}\ominus\hbar)\right\}}dt_2d\boldsymbol{s}_2\\
&\quad+\mathbb{E}_{(T_{i_2},\boldsymbol{S}_{i_2})}\left[\frac{\beta_2(T_{i_2},\boldsymbol{S}_{i_2})}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right)\mathbb{1}_{\left\{(\widetilde{t},s_2)\in\partial\mathcal{E}\oplus\hbar\right\}}\right]\\
&\quad+\mathbb{E}_{(T_{i_2},\boldsymbol{S}_{i_2})}\left\{\frac{\left[\beta_2(T_{i_2},\boldsymbol{S}_{i_2})-\beta_2(\widetilde{t},\boldsymbol{S}_{i_2})\right]}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right)\mathbb{1}_{\left\{(\widetilde{t},s_2)\in\partial\mathcal{E}\oplus\hbar\right\}}\right\}\\
&\le\int_{\mathbb{R}\times\mathcal{S}}\frac{\beta_2(\widetilde{t},\boldsymbol{s}_2)}{p_T(\widetilde{t})}\cdot\bar{K}_T(u)\cdot p(\widetilde{t}+u\hbar,\boldsymbol{s}_2)\cdot\mathbb{1}_{\left\{(\widetilde{t},s_2)\in\mathcal{E}\cap(\partial\mathcal{E}\ominus\hbar)\right\}}du\,d\boldsymbol{s}_2\\
&\quad+\int_{\mathbb{R}\times\mathcal{S}}\frac{\beta_2(\widetilde{t}+u\hbar,\boldsymbol{s}_2)}{p_T(\widetilde{t})}\cdot\bar{K}_T(u)\cdot p(\widetilde{t}+u\hbar,\boldsymbol{s}_2)\cdot\mathbb{1}_{\left\{(\widetilde{t},s_2)\in\partial\mathcal{E}\oplus\hbar\right\}}du\,d\boldsymbol{s}_2\\
&\quad+\mathbb{E}_{(T_{i_2},\boldsymbol{S}_{i_2})}\left[\frac{C_\mu|T_{i_2}-\widetilde{t}|}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right)\mathbb{1}_{\left\{(\widetilde{t},s_2)\in\partial\mathcal{E}\oplus\hbar\right\}}\right]\\
&\le\int_{\mathbb{R}\times\mathcal{S}}\frac{\beta_2(\widetilde{t},\boldsymbol{s}_2)}{p_T(\widetilde{t})}\cdot\bar{K}_T(u)\left[p(\widetilde{t},\boldsymbol{s}_2)+\hbar u\cdot\frac{\partial}{\partial t}p(t,\boldsymbol{s})+O(\hbar^2)\right]\mathbb{1}_{\left\{(\widetilde{t},s_2)\in\mathcal{E}\cap(\partial\mathcal{E}\ominus\hbar)\right\}}du\,d\boldsymbol{s}_2\\
&\quad+\int_{\mathbb{R}\times\mathcal{S}}\frac{\beta_2(\widetilde{t},\boldsymbol{s}_2)}{p_T(\widetilde{t})}\cdot\bar{K}_T(u)\cdot p(\widetilde{t}+u\hbar,\boldsymbol{s}_2)\cdot\mathbb{1}_{\left\{(\widetilde{t},s_2)\in\partial\mathcal{E}\oplus\hbar\right\}}du\,d\boldsymbol{s}_2\\
&\quad+\int_{\mathbb{R}\times\mathcal{S}}\frac{C_\mu\hbar|u|}{p_T(\widetilde{t})}\cdot\bar{K}_T(u)\cdot p(\widetilde{t}+u\hbar,\boldsymbol{s}_2)\cdot\mathbb{1}_{\left\{(\widetilde{t},s_2)\in\partial\mathcal{E}\oplus\hbar\right\}}du\,d\boldsymbol{s}_2\\
&\quad+\int_{\mathcal{T}\times\mathcal{S}}\frac{C_\mu|t_2-\widetilde{t}|}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-t_2}{\hbar}\right)\mathbb{1}_{\left\{(\widetilde{t},s_2)\in\partial\mathcal{E}\oplus\hbar\right\}}dt_2d\boldsymbol{s}_2\\
&\overset{\text{(v)}}{\le}\int_{\mathbb{R}\times\mathcal{S}}\frac{\beta_2(\widetilde{t},\boldsymbol{s}_2)}{p_T(\widetilde{t})}\cdot\bar{K}_T(u)\cdot p(\widetilde{t},\boldsymbol{s}_2)\cdot\mathbb{1}_{\left\{(\widetilde{t},s_2)\in\mathcal{E}\oplus\hbar\right\}}du\,d\boldsymbol{s}_2\\
&\quad+\int_{\mathbb{R}\times\mathcal{S}}\frac{\beta_2(\widetilde{t},\boldsymbol{s}_2)}{p_T(\widetilde{t})}\cdot\bar{K}_T(u)\left[\bar{p}(\widetilde{t}+u\hbar,\boldsymbol{s}_2)-\bar{p}(\widetilde{t},\boldsymbol{s}_2)\right]\mathbb{1}_{\left\{(\widetilde{t},s_2)\in\partial\mathcal{E}\oplus\hbar\right\}}du\,d\boldsymbol{s}_2+O(\hbar^2)\\
&=\int_{\mathcal{S}(\widetilde{t})}\beta_2(\widetilde{t},\boldsymbol{s}_2)\cdot p(\boldsymbol{s}_2|\widetilde{t})\,d\boldsymbol{s}_2+O(\hbar^2),
\end{aligned}$$

where $C_\mu=\sup_{(t,\boldsymbol{s})\in\mathcal{T}\times\mathcal{S}}\left|\frac{\partial}{\partial t}\beta_2(t,\boldsymbol{s})\right|=\sup_{(t,\boldsymbol{s})\in\mathcal{T}\times\mathcal{S}}\left|\frac{\partial^2}{\partial t^2}\mu(t,\boldsymbol{s})\right|<\infty$ by Assumption A3 and the equality (v) follows from Assumption A4(d) with the Lebesgue measure $|\partial\mathcal{E}\oplus\hbar|=$

$O(\hbar)$. Here, we introduce a smooth density $\bar{p}(\widetilde{t},\boldsymbol{s})$ that shares the same value with $p(\widetilde{t},\boldsymbol{s})$ within the support $\mathcal{E}$ but smoothly decays to 0 outside of $\mathcal{E}$.

Therefore, under our choices of $h,b,\hbar>0$ and $q\geq 2$, we proceed (37) as:

$$
\begin{aligned}
&\sup_{t\in\mathcal{T}}\left|\widehat{m}_\theta(t)-m(t)\right|\\
&=\sup_{t\in\mathcal{T}}\left|3\left(\mathbb{P}_n-\mathrm{P}\right)\mathrm{P}^2\Lambda_t+3\left(\mathbb{P}_n-\mathrm{P}\right)^2\mathrm{P}\Lambda_t+\left(\mathbb{P}_n-\mathrm{P}\right)^3\Lambda_t\right|\\
&\quad+O_P\left(\hbar^2+\sqrt{\frac{\log n}{n\hbar}}\right)+O\left(\hbar^2+h^q+b^2+\frac{\max\{b,h\}^4}{h}\right)\\
&=\sup_{t\in\mathcal{T}}\left|3\left(\mathbb{P}_n-\mathrm{P}\right)\mathrm{P}^2\Lambda_t+3\left(\mathbb{P}_n-\mathrm{P}\right)^2\mathrm{P}\Lambda_t+\left(\mathbb{P}_n-\mathrm{P}\right)^3\Lambda_t\right|\\
&\quad+O_P\left(\hbar^2+\sqrt{\frac{\log n}{n\hbar}}\right)+O(\hbar^2+h^2).
\end{aligned}
\tag{38}
$$

According to the above calculations, we can also compute that $\mathrm{P}\Lambda_t$ is equal to the function

$$
\begin{aligned}
(\boldsymbol{U}_{i_1},\boldsymbol{U}_{i_2})\mapsto&\frac{1}{6}\int_{T_{i_1}}^t\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_2}}\left(Y_{i_3},T_{i_3},\boldsymbol{S}_{i_3}\right)\right]}{h^2b^d\cdot p(\widetilde{t},\boldsymbol{S}_{i_2})\cdot p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right)d\widetilde{t}\\
&+\frac{1}{6}\int_{T_{i_2}}^t\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_1}}\left(Y_{i_3},T_{i_3},\boldsymbol{S}_{i_3}\right)\right]}{h^2b^d\cdot p(\widetilde{t},\boldsymbol{S}_{i_2})\cdot p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_1}}{\hbar}\right)d\widetilde{t}\\
&+\frac{1}{6}\int_{T_{i_1}}^t\mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_3}}\left(Y_{i_2},T_{i_2},\boldsymbol{S}_{i_2}\right)}{h^2b^d\cdot p(\widetilde{t},\boldsymbol{S}_{i_3})\cdot p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_3}}{\hbar}\right)\right]d\widetilde{t}\\
&+\frac{1}{6}\int_{T_{i_2}}^t\mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_3}}\left(Y_{i_1},T_{i_1},\boldsymbol{S}_{i_1}\right)}{h^2b^d\cdot p(\widetilde{t},\boldsymbol{S}_{i_3})\cdot p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_3}}{\hbar}\right)\right]d\widetilde{t}\\
&+\frac{1}{6}\cdot\mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\int_{T_{i_3}}^t\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_1}}\left(Y_{i_2},T_{i_2},\boldsymbol{S}_{i_2}\right)}{h^2b^d\cdot p(\widetilde{t},\boldsymbol{S}_{i_1})\cdot p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_1}}{\hbar}\right)d\widetilde{t}\right]\\
&+\frac{1}{6}\cdot\mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\int_{T_{i_3}}^t\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_2}}\left(Y_{i_1},T_{i_1},\boldsymbol{S}_{i_1}\right)}{h^2b^d\cdot p(\widetilde{t},\boldsymbol{S}_{i_2})\cdot p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right)d\widetilde{t}\right]\\
=&\frac{1}{6}\int_{T_{i_1}}^t\frac{\beta_2(\widetilde{t},\boldsymbol{S}_{i_2})}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right)d\widetilde{t}+\frac{1}{6}\int_{T_{i_2}}^t\frac{\beta_2(\widetilde{t},\boldsymbol{S}_{i_1})}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_1}}{\hbar}\right)d\widetilde{t}\\
&+O_P\left(h^q+b^2+\frac{\max\{b,h\}^4}{h}\right)\\
&+\frac{1}{6}\int_{T_{i_1}}^t\mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_3}}\left(Y_{i_2},T_{i_2},\boldsymbol{S}_{i_2}\right)}{h^2b^d\cdot p(\widetilde{t},\boldsymbol{S}_{i_3})\cdot p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_3}}{\hbar}\right)\right]d\widetilde{t}\\
&+\frac{1}{6}\int_{T_{i_2}}^t\mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_3}}\left(Y_{i_1},T_{i_1},\boldsymbol{S}_{i_1}\right)}{h^2b^d\cdot p(\widetilde{t},\boldsymbol{S}_{i_3})\cdot p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_3}}{\hbar}\right)\right]d\widetilde{t}
\end{aligned}
$$

$$+\frac{1}{6}\cdot\mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\int_{T_{i_3}}^{t}\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_1}}(Y_{i_2},T_{i_2},\boldsymbol{S}_{i_2})}{h^2b^d\cdot p(\widetilde{t},\boldsymbol{S}_{i_1})\cdot p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_1}}{\hbar}\right)d\widetilde{t}\right]$$
$$+\frac{1}{6}\cdot\mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\int_{T_{i_3}}^{t}\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_2}}(Y_{i_1},T_{i_1},\boldsymbol{S}_{i_1})}{h^2b^d\cdot p(\widetilde{t},\boldsymbol{S}_{i_2})\cdot p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right)d\widetilde{t}\right],$$

and $\mathrm{P}^2\Lambda_t$ is equal to the function

$$\begin{aligned}
\boldsymbol{U}_{i_1}\mapsto\ &\frac{1}{3}\int_{T_{i_1}}^{t}\mathbb{E}_{\boldsymbol{U}_{i_2}}\left[\frac{\beta_2(\widetilde{t},\boldsymbol{S}_{i_2})}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right)\right]d\widetilde{t}+\frac{1}{3}\cdot\mathbb{E}_{\boldsymbol{U}_{i_2}}\left[\int_{T_{i_2}}^{t}\frac{\beta_2(\widetilde{t},\boldsymbol{S}_{i_1})}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_1}}{\hbar}\right)d\widetilde{t}\right]\\
&+\frac{1}{3}\cdot\mathbb{E}_{\boldsymbol{U}_{i_2}}\left\{\int_{T_{i_2}}^{t}\mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_3}}(Y_{i_1},T_{i_1},\boldsymbol{S}_{i_1})}{h^2b^d\cdot p(\widetilde{t},\boldsymbol{S}_{i_3})\cdot p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_3}}{\hbar}\right)\right]d\widetilde{t}\right\}\\
=\ &\frac{1}{3}\int_{T_{i_1}}^{t}\mathbb{E}\left[\beta_2(\widetilde{t},\boldsymbol{S})|T=\widetilde{t}\right]d\widetilde{t}+\frac{1}{3}\cdot\mathbb{E}_{\boldsymbol{U}_{i_2}}\left[\int_{T_{i_2}}^{t}\frac{\beta_2(\widetilde{t},\boldsymbol{S}_{i_1})}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_1}}{\hbar}\right)d\widetilde{t}\right]\\
&+\frac{1}{3}\cdot\mathbb{E}_{\boldsymbol{U}_{i_2}}\left\{\int_{T_{i_2}}^{t}\mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_3}}(Y_{i_1},T_{i_1},\boldsymbol{S}_{i_1})}{h^2b^d\cdot p(\widetilde{t},\boldsymbol{S}_{i_3})\cdot p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_3}}{\hbar}\right)\right]d\widetilde{t}\right\}\\
&+O_P\left(\hbar^2+h^q+b^2+\frac{\max\{b,h\}^4}{h}\right)\\
=\ &\frac{1}{3}\int_{T_{i_1}}^{t}\theta_C(\widetilde{t})\,d\widetilde{t}+\frac{1}{3}\cdot\mathbb{E}_{\boldsymbol{U}_{i_2}}\left[\int_{T_{i_2}}^{t}\frac{\beta_2(\widetilde{t},\boldsymbol{S}_{i_1})}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_1}}{\hbar}\right)d\widetilde{t}\right]\\
&+\frac{1}{3}\cdot\mathbb{E}_{\boldsymbol{U}_{i_2}}\left\{\int_{T_{i_2}}^{t}\mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_3}}(Y_{i_1},T_{i_1},\boldsymbol{S}_{i_1})}{h^2b^d\cdot p(\widetilde{t},\boldsymbol{S}_{i_3})\cdot p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_3}}{\hbar}\right)\right]d\widetilde{t}\right\}\\
&+O_P\left(\hbar^2+h^q+b^2+\frac{\max\{b,h\}^4}{h}\right).
\end{aligned}$$

Now, focusing on $\mathrm{P}^2\Lambda_t$ above, we know from Assumptions A2 and A3 that $\mathrm{Var}\left[\int_T^t\theta_C(\widetilde{t})\,d\widetilde{t}\right]=\mathrm{Var}\left[m(T)\right]<\infty$ and hence,

$$\begin{aligned}
\sqrt{nh^3b^d}\,(\mathbb{P}_n-\mathrm{P})\int_{T_{i_1}}^{t}\theta_C(\widetilde{t})\,d\widetilde{t}&=\sqrt{nh^3b^d}\left[\frac{1}{n}\sum_{i=1}^{n}\int_{T_i}^{t}\theta_C(\widetilde{t})\,d\widetilde{t}-\mathbb{E}\left(\int_{T_i}^{t}\theta_C(\widetilde{t})\,d\widetilde{t}\right)\right]\\
&=O_P\left(\sqrt{h^3b^d}\right)=o_P(1).
\end{aligned}\tag{39}$$

Furthermore, under Assumptions A3 and A6(c), we also have that

$$\left|\mathbb{E}_{\boldsymbol{U}_{i_2}}\left[\int_{T_{i_2}}^{t}\frac{\beta_2(\widetilde{t},\boldsymbol{S}_{i_1})}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_1}}{\hbar}\right)d\widetilde{t}\right]\right|\lesssim\frac{1}{\hbar}\cdot\mathbb{E}|t-T_{i_2}|=O\left(\frac{1}{\hbar}\right)$$

so that

$$\sqrt{nh^3b^d}\,(\mathbb{P}_n-\mathrm{P})\left\{\mathbb{E}_{\boldsymbol{U}_{i_2}}\left[\int_{T_{i_2}}^{t}\frac{\beta_2(\widetilde{t},\cdot)}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-\cdot}{\hbar}\right)d\widetilde{t}\right]\right\}=O_P\left(\frac{\sqrt{h^3b^d}}{\hbar}\right)=o_P(1).\tag{40}$$

Let

$$\varphi_t\left(Y_{i_1}, T_{i_1}, \boldsymbol{S}_{i_1}\right) = \mathbb{E}_{\boldsymbol{U}_{i_2}}\left\{\int_{T_{i_2}}^{t} \mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\Psi}_{\tilde{t},\boldsymbol{S}_{i_3}}\left(Y_{i_1}, T_{i_1}, \boldsymbol{S}_{i_1}\right)}{\sqrt{hb^d}\cdot p(\tilde{t}, \boldsymbol{S}_{i_3})\cdot p_T(\tilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\tilde{t}-T_{i_3}}{\hbar}\right)\right] d\tilde{t}\right\}.$$

By (39) and (40), we know that the linear term $3\left(\mathbb{P}_n - \mathrm{P}\right)\mathrm{P}^2\Lambda_t$ is dominated by $\left(\mathbb{P}_n - \mathrm{P}\right)\left(\frac{\varphi_t}{\sqrt{hb^d}}\right)$. Then, we can proceed (38) as:

$$\begin{aligned}
&\widehat{m}_\theta(t) - m(t)\\
&= 3\left(\mathbb{P}_n - \mathrm{P}\right)\mathrm{P}^2\Lambda_t + 3\left(\mathbb{P}_n - \mathrm{P}\right)^2\mathrm{P}\Lambda_t + \left(\mathbb{P}_n - \mathrm{P}\right)^3\Lambda_t + O_P(\hbar^2) + O_P\left(\sqrt{\frac{\log n}{n\hbar}}\right)\\
&\quad + O_P(h^2) + O_P(\hbar^2) + O_P\left(\frac{1}{\sqrt{n}}\right) + O_P\left(\frac{1}{\hbar\sqrt{n}}\right)\\
&= \left(\mathbb{P}_n - \mathrm{P}\right)\left(\frac{\varphi_t}{\sqrt{h^3b^d}}\right) + 3\left(\mathbb{P}_n - \mathrm{P}\right)^2\mathrm{P}\Lambda_t + \left(\mathbb{P}_n - \mathrm{P}\right)^3\Lambda_t + O_P(\hbar^2) + O_P\left(\sqrt{\frac{\log n}{n\hbar}}\right)\\
&\quad + O_P(h^2) + O_P(\hbar^2) + O_P\left(\frac{1}{\hbar\sqrt{n}}\right)
\end{aligned}$$

and

(41)

$$\begin{aligned}
&\sup_{t\in\mathcal{T}'}\left|\sqrt{nh^3b^d}\left[\widehat{m}_\theta(t) - m(t)\right]\right|\\
&= \sup_{t\in\mathcal{T}'}\left|3\sqrt{nh^3b^d}\left(\mathbb{P}_n - \mathrm{P}\right)\mathrm{P}^2\Lambda_t + 3\sqrt{nh^3b^d}\left(\mathbb{P}_n - \mathrm{P}\right)^2\mathrm{P}\Lambda_t + \sqrt{nh^3b^d}\left(\mathbb{P}_n - \mathrm{P}\right)^3\Lambda_t\right|\\
&\quad + \sqrt{nh^3b^d}\cdot\left[O_P(\hbar^2) + O_P\left(\sqrt{\frac{\log n}{n\hbar}}\right)\right] + O\left(\sqrt{nh^7b^d}\right) + O\left(\sqrt{nh^3b^d\hbar^4}\right) + O_P\left(\sqrt{\frac{h^3b^d}{\hbar^2}}\right)\\
&\le \sup_{t\in\mathcal{T}'}\left|\mathbb{G}_n\varphi_t\right| + \sup_{t\in\mathcal{T}'}\left|3\sqrt{nh^3b^d}\left(\mathbb{P}_n - \mathrm{P}\right)^2\mathrm{P}\Lambda_t\right| + \sup_{t\in\mathcal{T}'}\left|\sqrt{nh^3b^d}\left(\mathbb{P}_n - \mathrm{P}\right)^3\Lambda_t\right|\\
&\quad + O_P\left(\sqrt{\frac{h^{d+3}\log n}{\hbar}}\right) + O\left(\sqrt{nh^{d+7}}\right) + O\left(\sqrt{nh^{d+3}\hbar^4}\right) + O_P\left(\sqrt{\frac{h^{d+3}}{\hbar^2}}\right).
\end{aligned}$$

To complete **Step 1**, it remains to show that $3\left(\mathbb{P}_n - \mathrm{P}\right)^2\mathrm{P}\Lambda_t + \left(\mathbb{P}_n - \mathrm{P}\right)^3\Lambda_t = o_P\left(\frac{1}{\sqrt{nh^3b^d}}\right)$. This can be achieved in **Step 2** below, where we derive the rates of convergence for $\sup_{t\in\mathcal{T}'}\left|3\sqrt{nh^3b^d}\left(\mathbb{P}_n - \mathrm{P}\right)^2\mathrm{P}\Lambda_t\right|$ and $\sup_{t\in\mathcal{T}'}\left|\sqrt{nh^3b^d}\left(\mathbb{P}_n - \mathrm{P}\right)^3\Lambda_t\right|$, respectively.

- **Step 2:** Based on (34) and the symmetry of Λ_t, we know that

$$\begin{aligned}
&\sup_{t\in\mathcal{T}'}\left|\sqrt{nh^3b^d}\left(\mathbb{P}_n - \mathrm{P}\right)^3\Lambda_t\right|\\
&= \sup_{t\in\mathcal{T}'}\left|\left(\mathbb{P}_n - \mathrm{P}\right)^2\left\{\int_{T_{i_1}}^{t}\frac{\sqrt{hb^d}}{p_T(\tilde{t})}\cdot\mathbb{G}_n\left[\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\tilde{t},\boldsymbol{S}_{i_2}}}{p(\tilde{t},\boldsymbol{S}_{i_2})\cdot hb^d}\right]\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\tilde{t}-T_{i_2}}{\hbar}\right)d\tilde{t}\right\}\right|\\
&\le \sup_{t\in\mathcal{T}'}\left|\left(\mathbb{P}_n - \mathrm{P}\right)^2\int_{T_{i_1}}^{t}\frac{C_{h,b}}{p_T(\tilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\tilde{t}-T_{i_2}}{\hbar}\right)d\tilde{t}\right|,
\end{aligned}$$

where $C_{h,b} = O_P\left(\sqrt{|\log(hb^d)|}\right) = O_P\left(\sqrt{\log n}\right)$ under our choices of h, b and we keep the symbols of those random variables $T_{i_1}, T_{i_2}, \boldsymbol{S}_{i_2}$ to which $(\mathbb{P}_n - \mathrm{P})^2$ applies for clarity. Furthermore, by Assumption A6(d) and Theorem 2 in Einmahl and Mason (2005), $\mathbb{G}_n\left[\frac{1}{\sqrt{\hbar}}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right)\right] = O_P\left(\sqrt{|\log \hbar|}\right)$, and we proceed the above display as:

$$
\begin{aligned}
&\sup_{t\in\mathcal{T}'}\left|\sqrt{nh^3b^d}\,(\mathbb{P}_n - \mathrm{P})^3\,\Lambda_t\right| \\
&\leq \sup_{t\in\mathcal{T}'}\left|(\mathbb{P}_n - \mathrm{P})\left\{\int_{T_{i_1}}^t \frac{1}{p_T(\widetilde{t})\sqrt{n\hbar}}\cdot \mathbb{G}_n\left[\frac{C_{h,b}}{\sqrt{\hbar}}\cdot \bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right)\right] d\widetilde{t}\right\}\right| \\
&\leq \sup_{t\in\mathcal{T}'}\left|(\mathbb{P}_n - \mathrm{P})\int_{T_{i_1}}^t \frac{C'_{h,b}}{p_T(\widetilde{t})\sqrt{n\hbar}} d\widetilde{t}\right|,
\end{aligned}
$$

where $C'_{h,b} = O_P\left(\sqrt{|\log(hb^d)|\cdot|\log \hbar|}\right) = O_P(\log n)$ under our choices of $h, b, \hbar$. Thus, we obtain that

$$
\sup_{t\in\mathcal{T}'}\left|\sqrt{nh^3b^d}\,(\mathbb{P}_n - \mathrm{P})^3\,\Lambda_t\right| = O_P\left(\frac{\log n}{n\sqrt{\hbar}}\right).
$$

Following the similar arguments, we also have that

$$
\begin{aligned}
&\sup_{t\in\mathcal{T}'}\left|3\sqrt{nh^3b^d}\,(\mathbb{P}_n - \mathrm{P})^2\,\mathrm{P}\Lambda_t\right| \\
&\overset{\text{(vi)}}{=} \sup_{t\in\mathcal{T}'}\left|\sqrt{h^3b^d}\,(\mathbb{P}_n - \mathrm{P})\int_{T_{i_1}}^t \frac{1}{p_T(\widetilde{t})}\cdot \mathbb{G}_n\left[\frac{\beta_2(\widetilde{t},\boldsymbol{S}_{i_2})}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_2}}{\hbar}\right)\right] d\widetilde{t}\right| \\
&\quad + \sup_{t\in\mathcal{T}'}\left|(\mathbb{P}_n - \mathrm{P})\left\{\int_{T_{i_1}}^t \mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\frac{\sqrt{hb^d}}{p_T(\widetilde{t})}\cdot \mathbb{G}_n\left[\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_3}}}{p(\widetilde{t},\boldsymbol{S}_{i_3})\cdot hb^d}\right]\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_3}}{\hbar}\right)\right] d\widetilde{t}\right\}\right| \\
&\quad + \sup_{t\in\mathcal{T}'}\left|(\mathbb{P}_n - \mathrm{P})\left\{\mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\int_{T_{i_3}}^t \frac{\sqrt{hb^d}}{p_T(\widetilde{t})}\cdot \mathbb{G}_n\left[\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_1}}}{p(\widetilde{t},\boldsymbol{S}_{i_1})\cdot hb^d}\right]\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_1}}{\hbar}\right) d\widetilde{t}\right]\right\}\right| \\
&\quad + \sqrt{nh^3b^d}\left[O_P\left(h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right)\right] \\
&= \sup_{t\in\mathcal{T}'}\left|\sqrt{h^3b^d}\,(\mathbb{P}_n - \mathrm{P})\int_{T_{i_1}}^t \frac{C_{\hbar}}{p_T(\widetilde{t})\cdot\sqrt{\hbar}} d\widetilde{t}\right| \\
&\quad + \sup_{t\in\mathcal{T}'}\left|(\mathbb{P}_n - \mathrm{P})\left\{\int_{T_{i_1}}^t \mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\frac{C_{h,b}}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_3}}{\hbar}\right)\right] d\widetilde{t}\right\}\right| \\
&\quad + \sup_{t\in\mathcal{T}'}\left|(\mathbb{P}_n - \mathrm{P})\left\{\mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\int_{T_{i_3}}^t \frac{C_{h,b}}{p_T(\widetilde{t})}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_1}}{\hbar}\right) d\widetilde{t}\right]\right\}\right| \\
&\quad + O_P\left(\sqrt{nh^7b^d} + \sqrt{nhb^d\cdot\max\{b,h\}^8}\right) \\
&= O_P\left(\sqrt{\frac{h^3b^d\log n}{n\hbar}} + \sqrt{\frac{\log n}{n\hbar}} + \sqrt{\frac{\log n}{n\hbar^2}} + \sqrt{nh^{d+7}}\right)
\end{aligned}
$$

$$= O_P\left(\sqrt{\frac{\log n}{n\hbar}} + \sqrt{\frac{\log n}{n\hbar^2}} + \sqrt{nh^{d+7}}\right),$$

where $C_\hbar = O_P\left(\sqrt{|\log \hbar|}\right) = O_P\left(\sqrt{\log n}\right)$ and $h \asymp b \asymp n^{-\frac{1}{\gamma}}$. Here, the equality (vi) utilizes Fubini's theorem to argue the interchangeability of $(\mathbb{P}_n - \mathrm{P})$ with respect to $\boldsymbol{U}_{i_2}$ and the expectation $\mathbb{E}_{\boldsymbol{U}_{i_3}}$. Thus, together with (41), we conclude that

$$\begin{aligned}
&\sqrt{nh^3b^d} \cdot \sup_{t\in\mathcal{T}'} |\widehat{m}_\theta(t) - m(t)| \\
&\leq \sup_{t\in\mathcal{T}'} |\mathbb{G}_n\varphi_t| + \sup_{t\in\mathcal{T}'} \left|3\sqrt{nh^3b^d}\,(\mathbb{P}_n - \mathrm{P})^2\,\mathrm{P}\Lambda_t\right| + \sup_{t\in\mathcal{T}} \left|\sqrt{nh^3b^d}\,(\mathbb{P}_n - \mathrm{P})^3\,\Lambda_t\right| \\
&\quad + O_P\left(\sqrt{\frac{h^{d+3}\log n}{\hbar}} + \sqrt{nh^{d+7}} + \sqrt{nh^{d+3}\hbar^4} + \sqrt{\frac{h^{d+3}}{\hbar^2}}\right) \\
&= \sup_{t\in\mathcal{T}'} |\mathbb{G}_n\varphi_t| + O_P\left(\sqrt{\frac{\log n}{n\hbar}} + \sqrt{\frac{\log n}{n\hbar^2}} + \sqrt{nh^{d+7}} + \frac{\log n}{n\sqrt{\hbar}}\right) \\
&\quad + O_P\left(\sqrt{\frac{h^{d+3}\log n}{\hbar}} + \sqrt{nh^{d+7}} + \sqrt{nh^{d+3}\hbar^4} + \sqrt{\frac{h^{d+3}}{\hbar^2}}\right) \\
&= \sup_{t\in\mathcal{T}'} |\mathbb{G}_n\varphi_t| + O_P\left(\sqrt{nh^{d+7}} + \sqrt{\frac{\log n}{n\hbar^2}} + \sqrt{\frac{h^{d+3}\log n}{\hbar}} + \sqrt{\frac{h^{d+3}}{\hbar^2}}\right),
\end{aligned}$$

where $O_P\left(\sqrt{nh^{d+7}}\right)$ asymptotically dominates $O_P\left(\sqrt{nh^{d+3}\hbar^4}\right)$, as well as $O_P\left(\sqrt{\frac{\log n}{n\hbar^2}}\right)$ dominates $O_P\left(\sqrt{\frac{\log n}{n\hbar}}\right)$ and $O_P\left(\frac{\log n}{n\sqrt{\hbar}}\right)$ when $h \asymp b \asymp n^{-\frac{1}{\gamma}}$ and $\hbar \asymp n^{-\frac{1}{\varpi}}$ for some $\gamma \geq \varpi > 0$ such that $\frac{nh^{d+5}}{\log n} \to c_1$ and $\frac{n\hbar^5}{\log n} \to c_2$ for some finite number $c_1, c_2 \geq 0$ and $\frac{\log n}{n\hbar^2}, \frac{h^{d+3}\log n}{\hbar}, \frac{h^{d+3}}{\hbar^2} \to 0$ as $n \to \infty$. This concludes the proof of the asymptotic linearity of $\sqrt{nh^3b^d}\,[\widehat{m}_\theta(t) - m(t)]$.

• **Step 3:** Following the analogous but simpler arguments in **Step 1**, we know that

$$\begin{aligned}
&\widehat{\theta}_C(t) - \theta_C(t) \\
&= \sum_{i=1}^n \frac{\mathbb{P}_n\left(\frac{1}{hb^d}\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{S}_i}\right) \cdot \bar{K}_T\left(\frac{t-T_i}{\hbar}\right)}{h \cdot p(t,\boldsymbol{S}_i) \cdot \sum_{j=1}^n \bar{K}_T\left(\frac{t-T_j}{\hbar}\right)} - \mathbb{E}\left[\beta_2(t,\boldsymbol{s})\middle|T=t\right] \\
&= \frac{1}{n}\sum_{i=1}^n \frac{\mathbb{P}_n\left(\frac{1}{hb^d}\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{S}_i}\right) \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{t-T_i}{\hbar}\right)}{h \cdot p(t,\boldsymbol{S}_i) \cdot p_T(t)} - \mathbb{E}\left[\beta_2(t,\boldsymbol{s})\middle|T=t\right] + O_P\left(\hbar^2 + \sqrt{\frac{\log n}{n\hbar}}\right) \\
&= \frac{1}{n^2}\sum_{i_1=1}^n\sum_{i_2=1}^n \frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{S}_{i_1}}(Y_{i_2},T_{i_2},\boldsymbol{S}_{i_2})}{h^2b^d \cdot p(t,\boldsymbol{S}_{i_1}) \cdot p_T(t)} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{t-T_{i_1}}{\hbar}\right) - \mathbb{E}\left[\beta_2(t,\boldsymbol{s})\middle|T=t\right] \\
&\quad + O_P\left(\hbar^2 + \sqrt{\frac{\log n}{n\hbar}}\right).
\end{aligned}$$

Recall that we denote the data sample by $\boldsymbol{U}_i = (Y_i, T_i, \boldsymbol{S}_i), i = 1, ..., n$ with $\mathbb{P}_n$ as the empirical measure. We again use the V-statistic to handle the double summation term in the above display. Define a symmetric "kernel" function as:

$$\bar{\Lambda}_t(\boldsymbol{U}_{i_1}, \boldsymbol{U}_{i_2}) = \frac{1}{2}\left[\frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\Psi}_{t,\boldsymbol{S}_{i_1}}(Y_{i_2}, T_{i_2}, \boldsymbol{S}_{i_2})}{h^2 b^d \cdot p(t, \boldsymbol{S}_{i_1}) \cdot p_T(t)} \cdot \frac{1}{\hbar} \bar{K}_T\left(\frac{t - T_{i_1}}{\hbar}\right) \right. \\ \left. + \frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\Psi}_{t,\boldsymbol{S}_{i_2}}(Y_{i_1}, T_{i_1}, \boldsymbol{S}_{i_1})}{h^2 b^d \cdot p(t, \boldsymbol{S}_{i_2}) \cdot p_T(t)} \cdot \frac{1}{\hbar} \bar{K}_T\left(\frac{t - T_{i_2}}{\hbar}\right)\right].$$

Under this notation, we obtain from Pascal's rule that

(42)

$$\widehat{\theta}_C(t) - \theta_C(t) \\ = \mathrm{P}^2 \bar{\Lambda}_t + 2\left(\mathbb{P}_n - \mathrm{P}\right) \mathrm{P} \bar{\Lambda}_t + \left(\mathbb{P}_n - \mathrm{P}\right)^2 \bar{\Lambda}_t - \mathbb{E}\left[\beta_2(t, \boldsymbol{s}) \middle| T = t\right] + O_P\left(\hbar^2 + \sqrt{\frac{\log n}{n\hbar}}\right),$$

where we use the shorthand notation $\mathrm{P}\bar{\Lambda}_t$ to refer to the function $\boldsymbol{U}_{i_1} \mapsto \int \bar{\Lambda}_t(\boldsymbol{U}_{i_1}, \boldsymbol{u}_{i_2})\, d\mathrm{P}(\boldsymbol{u}_{i_2})$. In addition, based on the bias term in the proof of Lemma 3 and our arguments before (38), we know that

$$\mathrm{P}^2 \bar{\Lambda}_t - \mathbb{E}\left[\beta_2(t, \boldsymbol{s}) \middle| T = t\right] \\ = \mathbb{E}_{(T_{i_1}, \boldsymbol{S}_{i_1})}\left[\frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \mathbb{E}_{\boldsymbol{U}_{i_2}}\left[\boldsymbol{\Psi}_{t,\boldsymbol{S}_{i_1}}(Y_{i_2}, T_{i_2}, \boldsymbol{S}_{i_2})\right]}{h^2 b^d \cdot p(t, \boldsymbol{S}_{i_1}) \cdot p_T(t)} \cdot \frac{1}{\hbar} \bar{K}_T\left(\frac{t - T_{i_1}}{\hbar}\right)\right] - \mathbb{E}\left[\beta_2(t, \boldsymbol{s}) \middle| T = t\right] \\ = \mathbb{E}_{(T_{i_1}, \boldsymbol{S}_{i_1})}\left[\frac{\beta_2(t, \boldsymbol{S}_{i_1}) + O(h^q) + O(b^2) + O\left(\frac{\max\{h,b\}^4}{h}\right)}{p_T(t)} \cdot \frac{1}{\hbar} \bar{K}_T\left(\frac{t - T_{i_1}}{\hbar}\right)\right] - \mathbb{E}\left[\beta_2(t, \boldsymbol{s}) \middle| T = t\right] \\ = \mathbb{E}_{(T_{i_1}, \boldsymbol{S}_{i_1})}\left[\frac{\beta_2(t, \boldsymbol{S}_{i_1})}{p_T(t)} \cdot \frac{1}{\hbar} \bar{K}_T\left(\frac{t - T_{i_1}}{\hbar}\right)\right] - \mathbb{E}\left[\beta_2(t, \boldsymbol{s}) \middle| T = t\right] + O_P\left(h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right) \\ = O(\hbar^2) + O_P\left(h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right).$$

According to the above calculation, we can also compute that $\mathrm{P}\bar{\Lambda}_t$ is equal to the function

$$\boldsymbol{U}_{i_1} \mapsto \frac{1}{2} \cdot \frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \mathbb{E}_{\boldsymbol{U}_{i_2}}\left[\boldsymbol{\Psi}_{t,\boldsymbol{S}_{i_1}}(Y_{i_2}, T_{i_2}, \boldsymbol{S}_{i_2})\right]}{h^2 b^d \cdot p(t, \boldsymbol{S}_{i_1}) \cdot p_T(t)} \cdot \frac{1}{\hbar} \bar{K}_T\left(\frac{t - T_{i_1}}{\hbar}\right) \\ + \frac{1}{2} \cdot \mathbb{E}_{(T_{i_2}, \boldsymbol{S}_{i_2})}\left[\frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\Psi}_{t,\boldsymbol{S}_{i_2}}(Y_{i_1}, T_{i_1}, \boldsymbol{S}_{i_1})}{h^2 b^d \cdot p(t, \boldsymbol{S}_{i_2}) \cdot p_T(t)} \cdot \frac{1}{\hbar} \bar{K}_T\left(\frac{t - T_{i_2}}{\hbar}\right)\right] \\ = \frac{\beta_2(t, \boldsymbol{S}_{i_1})}{2 p_T(t)} \cdot \frac{1}{\hbar} \bar{K}_T\left(\frac{t - T_{i_1}}{\hbar}\right) + O_P\left(h^q + b^2 + \frac{\max\{b,h\}^4}{h}\right) \\ + \frac{1}{2} \cdot \mathbb{E}_{(T_{i_2}, \boldsymbol{S}_{i_2})}\left[\frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\Psi}_{t,\boldsymbol{S}_{i_2}}(Y_{i_1}, T_{i_1}, \boldsymbol{S}_{i_1})}{h^2 b^d \cdot p(t, \boldsymbol{S}_{i_2}) \cdot p_T(t)} \cdot \frac{1}{\hbar} \bar{K}_T\left(\frac{t - T_{i_2}}{\hbar}\right)\right].$$

Now, let

$$\bar{\varphi}_t(Y_{i_1}, T_{i_1}, \boldsymbol{S}_{i_1}) = \mathbb{E}_{(T_{i_2}, \boldsymbol{S}_{i_2})}\left[\frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\Psi}_{t,\boldsymbol{S}_{i_2}}(Y_{i_1}, T_{i_1}, \boldsymbol{S}_{i_1})}{\sqrt{h b^d} \cdot p(t, \boldsymbol{S}_{i_2}) \cdot p_T(t)} \cdot \frac{1}{\hbar} \bar{K}_T\left(\frac{t - T_{i_2}}{\hbar}\right)\right].$$

By Assumptions A3, A4, and A6(d), we know from Theorem 2 in Einmahl and Mason (2005) that $\mathbb{G}_n\left[\frac{\beta_2(t,\boldsymbol{S}_2)}{\sqrt{\hbar}}\bar{K}_T\left(\frac{t-T_{i_1}}{\hbar}\right)\right] = O_P\left(\sqrt{\log \hbar}\right)$ and thus,

$$\sqrt{nh^3b^d}\,(\mathbb{P}_n - \mathrm{P})\left[\frac{\beta_2(t,\boldsymbol{S}_{i_1})}{p_T(t)}\cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{t-T_{i_1}}{\hbar}\right)\right] = O_P\left(\sqrt{\frac{h^3b^d}{\hbar}}\right).$$

Then, we can proceed (42) as:

$$\begin{aligned}
&\sqrt{nh^3b^d}\left[\widehat{\theta}_C(t) - \theta_C(t)\right] \\
&= \mathbb{G}_n\bar{\varphi}_t + \sqrt{nh^3b^d}\,(\mathbb{P}_n - \mathrm{P})^2\,\bar{\Lambda}_t + O_P\left(\sqrt{\frac{h^3b^d}{\hbar}}\right) \\
&\quad + \sqrt{nh^3b^d}\left[O_P\left(\hbar^2 + \sqrt{\frac{\log n}{n\hbar}} + h^q + b^2 + \frac{\max\{h,b\}^4}{h}\right)\right] \\
&= \mathbb{G}_n\bar{\varphi}_t + \sqrt{nh^3b^d}\,(\mathbb{P}_n - \mathrm{P})^2\,\bar{\Lambda}_t + O_P\left(\sqrt{nh^{d+7}} + \sqrt{\frac{h^{d+3}\log n}{\hbar}}\right)
\end{aligned} \tag{43}$$

when $h \asymp b \asymp n^{-\frac{1}{\gamma}}$ and $\hbar \asymp n^{-\frac{1}{\varpi}}$ for some $\gamma \geq \varpi > 0$. Based on (34) and the symmetry of $\bar{\Lambda}_t$, we know that

$$\begin{aligned}
\sup_{t\in\mathcal{T}'}\left|\sqrt{nh^3b^d}\,(\mathbb{P}_n - \mathrm{P})^2\,\bar{\Lambda}_t\right| &= \sup_{t\in\mathcal{T}'}\left|(\mathbb{P}_n - \mathrm{P})\left[\frac{\sqrt{hb^d}}{p_T(t)}\cdot\mathbb{G}_n\left(\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{S}_{i_1}}}{p(t,\boldsymbol{S}_{i_1})\cdot hb^d}\right)\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{t-T_{i_1}}{\hbar}\right)\right]\right| \\
&\leq \sup_{t\in\mathcal{T}'}\left|(\mathbb{P}_n - \mathrm{P})\left[\frac{C_{h,b}}{p_T(t)}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{t-T_{i_1}}{\hbar}\right)\right]\right| \\
&= \sup_{t\in\mathcal{T}'}\left|\frac{C_{h,b}}{p_T(t)\cdot\sqrt{n\hbar}}\cdot\mathbb{G}_n\left[\frac{1}{\sqrt{\hbar}}\bar{K}_T\left(\frac{t-T_{i_1}}{\hbar}\right)\right]\right| \\
&\leq \sup_{t\in\mathcal{T}'}\left|\frac{C'_{h,b}}{p_T(t)\cdot\sqrt{n\hbar}}\right| \\
&= O_P\left(\frac{\log n}{\sqrt{n\hbar}}\right),
\end{aligned}$$

where $C_{h,b} = O_P\left(\sqrt{|\log(hb^d)|}\right)$ and $C'_{h,b} = O_P\left(\sqrt{|\log(hb^d)|\cdot|\log\hbar|}\right) = O_P(\log n)$ under our choice of $h, b, \hbar$. Therefore, together with (43), we conclude that

$$\left|\sqrt{nh^3b^d}\sup_{t\in\mathcal{T}'}\left|\widehat{\theta}_C(t) - \theta_C(t)\right| - \sup_{t\in\mathcal{T}'}\left|\mathbb{G}_n\bar{\varphi}_t\right|\right| = O_P\left(\frac{\log n}{\sqrt{n\hbar}} + \sqrt{nh^{d+7}} + \sqrt{\frac{h^{d+3}\log n}{\hbar}}\right).$$

The proof is thus completed. □

B.6. Proof of Theorem 6. Before proving Theorem 6, we first study some auxiliary results. Recall that the class $\mathcal{G}$ of measurable functions on $\mathbb{R}^{d+1}$ is VC-type if there exist constants $A_2, \upsilon_2 > 0$ such that for any $0 < \epsilon < 1$,

$$\sup_Q N\left(\mathcal{G}, L_2(Q), \epsilon\,||G||_{L_2(Q)}\right) \leq \left(\frac{A_2}{\epsilon}\right)^{\upsilon_2},$$

where $N\left(\mathcal{G}, L_2(Q), \epsilon \left|\left|G\right|\right|_{L_2(Q)}\right)$ is the $\epsilon \left|\left|G\right|\right|_{L_2(Q)}$-covering number of the (semi-)metric space $\left(\mathcal{G}, \left|\left|\cdot\right|\right|_{L_2(Q)}\right)$, Q is any probability measure on $\mathbb{R}^{d+1}$, G is an envelope function of $\mathcal{G}$, and $\left|\left|G\right|\right|_{L_2(Q)}$ is defined as $\left[\int_{\mathbb{R}^{d+1}} \left[G(x)\right]^2 dQ(x)\right]^{\frac{1}{2}}$.

LEMMA 9 (VC-type result related to the influence function φ_t). *Let $q \geq 2$ in the local polynomial regression for estimating $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ and $\mathcal{T}' \subset \mathcal{T}$ be a compact set so that $p_T(t)$ is uniformly bounded away from 0 within $\mathcal{T}'$. Suppose that Assumptions A3, A4, and A6 hold. Furthermore, we assume that $p(t,\boldsymbol{s}) < \infty$ for all $(t,\boldsymbol{s}) \in \mathcal{T} \times \mathcal{S}$ and $\boldsymbol{M}_q^{-1}$ exists. Then, the class of scaled influence functions*

$$\widetilde{\mathcal{F}} = \left\{(v,x,\boldsymbol{z}) \mapsto \sqrt{h^3 b^d} \cdot \varphi_t(v,x,\boldsymbol{z}) : t \in \mathcal{T}'\right\}$$

has its covering number $N\left(\widetilde{\mathcal{F}}, L_2(Q), \epsilon\right)$ and bracketing number $N_{[]}\left(\widetilde{\mathcal{F}}, L_2(Q), 2\epsilon\right)$ as:

$$\sup_Q N\left(\widetilde{\mathcal{F}}, L_2(Q), \epsilon\right) \leq \sup_Q N_{[]}\left(\widetilde{\mathcal{F}}, L_2(Q), 2\epsilon\right) \leq \frac{C_4}{\epsilon}$$

for some constant $C_4 > 0$, where the supremum is taken over all probability measures Q for which the class $\widetilde{\mathcal{F}}$ is not identically 0. In other words, $\widetilde{\mathcal{F}}$ is a bounded VC-type class of functions with an envelope function $(v,x,\boldsymbol{z}) \mapsto F_1(v,x,\boldsymbol{z}) = C_5 \cdot |v|$ for some constant $C_5 > 0$. Analogously,

$$\widetilde{\mathcal{F}}_\theta = \left\{(v,x,\boldsymbol{z}) \mapsto \sqrt{h^3 b^d} \cdot \bar{\varphi}_t(v,x,\boldsymbol{z}) : t \in \mathcal{T}'\right\}$$

is also a bounded VC-type class of functions with an envelope function $(v,x,\boldsymbol{z}) \mapsto F_1(v,x,\boldsymbol{z}) = C_6 \cdot |v|$ for some constant $C_6 > 0$. Here, the constants $C_4, C_5, C_6 > 0$ are independent of $n, h, b, \hbar$.

PROOF OF LEMMA 9. For any $f_{t_1}, f_{t_2} \in \widetilde{\mathcal{F}}$, we know from (17) that

$$\begin{aligned}
&\left|f_{t_1}(v,x,\boldsymbol{z}) - f_{t_2}(v,x,\boldsymbol{z})\right| \\
&= \left|\sqrt{h^3 b^d} \cdot \left[\varphi_{t_1}(v,x,\boldsymbol{z}) - \varphi_{t_2}(v,x,\boldsymbol{z})\right]\right| \\
&= \Bigg|\mathbb{E}_{\boldsymbol{U}_{i_2}}\left[\int_{T_{i_2}}^{t_1} h \cdot \mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\Psi}_{\tilde{t},\boldsymbol{S}_{i_3}}(v,x,\boldsymbol{z})}{p(\tilde{t},\boldsymbol{S}_{i_3}) \cdot p_T(\tilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\tilde{t} - T_{i_3}}{\hbar}\right)\right] d\tilde{t}\right] \\
&\quad - \mathbb{E}_{\boldsymbol{U}_{i_2}}\left[\int_{T_{i_2}}^{t_2} h \cdot \mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\Psi}_{\tilde{t},\boldsymbol{S}_{i_3}}(v,x,\boldsymbol{z})}{p(\tilde{t},\boldsymbol{S}_{i_3}) \cdot p_T(\tilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\tilde{t} - T_{i_3}}{\hbar}\right)\right] d\tilde{t}\right]\Bigg| \\
&= \left|\int_{t_2}^{t_1} h \cdot \mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\Psi}_{\tilde{t},\boldsymbol{S}_{i_3}}(v,x,\boldsymbol{z})}{p(\tilde{t},\boldsymbol{S}_{i_3}) \cdot p_T(\tilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\tilde{t} - T_{i_3}}{\hbar}\right)\right] d\tilde{t}\right|.
\end{aligned}$$

By the definition of $\boldsymbol{M}_q$ in (14) when $q = 2$, we know that $\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} = \left(0, C_K, 0, \underbrace{0, ..., 0}_{d}\right)$ for some constant $C_K \neq 0$ that only depends on the kernels K_T, K_S. Together with (15), we know that

$$
\begin{aligned}
&\left|f_{t_1}(v,x,\boldsymbol{z}) - f_{t_2}(v,x,\boldsymbol{z})\right| \\
&= \left| \int_{t_2}^{t_1} h \cdot \mathbb{E}_{\boldsymbol{U}_{i_3}} \left[\frac{C_K v \left(\frac{x-\widetilde{t}}{h}\right) K_T\left(\frac{x-\widetilde{t}}{h}\right) K_S\left(\frac{\boldsymbol{z}-\boldsymbol{S}_{i_3}}{b}\right)}{p(\widetilde{t},\boldsymbol{S}_{i_3}) \cdot p_T(\widetilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_3}}{\hbar}\right)\right] d\widetilde{t} \right| \qquad (44)\\
&= \left| \int_{t_2}^{t_1} \mathbb{E}_{\boldsymbol{U}_{i_3}} \left[\frac{C_K v \left(x-\widetilde{t}\right) K_T\left(\frac{x-\widetilde{t}}{h}\right) K_S\left(\frac{\boldsymbol{z}-\boldsymbol{S}_{i_3}}{b}\right)}{p(\widetilde{t},\boldsymbol{S}_{i_3}) \cdot p_T(\widetilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_3}}{\hbar}\right)\right] d\widetilde{t} \right|.
\end{aligned}
$$

For any $\widetilde{t} \in \mathcal{T}'$, we consider bounding the (absolute) integrand in the above display when $\hbar$ is sufficiently small as:

$$
\begin{aligned}
&\left| \mathbb{E}_{\boldsymbol{U}_{i_3}} \left[\frac{C_K v \left(x-\widetilde{t}\right) K_T\left(\frac{x-\widetilde{t}}{h}\right) K_S\left(\frac{\boldsymbol{z}-\boldsymbol{S}_{i_3}}{b}\right)}{p(\widetilde{t},\boldsymbol{S}_{i_3}) \cdot p_T(\widetilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_3}}{\hbar}\right)\right] \right| \\
&= \left| \int_{\mathcal{T}\times\mathcal{S}} \frac{C_K v \left(x-\widetilde{t}\right) K_T\left(\frac{x-\widetilde{t}}{h}\right) K_S\left(\frac{\boldsymbol{z}-\boldsymbol{s}_3}{b}\right)}{p(\widetilde{t},\boldsymbol{s}_3) \cdot p_T(\widetilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-t_3}{\hbar}\right) \cdot p(t_3,\boldsymbol{s}_3)\, dt_3 d\boldsymbol{s}_3 \right| \\
&= C_K|v| \cdot \left| \int_{\mathcal{T}\times\mathcal{S}} \frac{\left(x-\widetilde{t}\right) K_T\left(\frac{x-\widetilde{t}}{h}\right) K_S\left(\frac{\boldsymbol{z}-\boldsymbol{s}_3}{b}\right)}{p(\widetilde{t},\boldsymbol{s}_3) \cdot p_T(\widetilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-t_3}{\hbar}\right) \cdot p(t_3,\boldsymbol{s}_3)\, dt_3 d\boldsymbol{s}_3 \right| \\
&= C_K|v| \cdot \left| \int_{\mathbb{R}\times\mathcal{S}} \frac{\left(x-\widetilde{t}\right) K_T\left(\frac{x-\widetilde{t}}{h}\right) K_S\left(\frac{\boldsymbol{z}-\boldsymbol{s}_3}{b}\right)}{p(\widetilde{t},\boldsymbol{s}_3) \cdot p_T(\widetilde{t})} \cdot \bar{K}_T(u) \cdot p(\widetilde{t}+u\hbar,\boldsymbol{s}_3)\, du d\boldsymbol{s}_3 \right| \\
&\stackrel{\text{(i)}}{\leq} |v| \left\{ \int_{\mathbb{R}\times\mathcal{S}} \left(x-\widetilde{t}\right)^2 K_T^2\left(\frac{x-\widetilde{t}}{h}\right) K_S^2\left(\frac{\boldsymbol{z}-\boldsymbol{s}_3}{b}\right) p(\widetilde{t}+u\hbar,\boldsymbol{s}_3)\, du d\boldsymbol{s}_3 \right\}^{\frac{1}{2}} \\
&\quad \times \left[\int_{\mathbb{R}\times\mathcal{S}} \frac{\bar{K}_T^2(u)}{p(\widetilde{t},\boldsymbol{s}_3)^2 p_T(\widetilde{t})^2} \cdot p(\widetilde{t}+u\hbar,\boldsymbol{s}_3)\, d\boldsymbol{s}_3 du \right]^{\frac{1}{2}} \\
&\stackrel{\text{(ii)}}{\leq} C_1|v| \left\{ \int_{\mathcal{S}} \left(x-\widetilde{t}\right)^2 K_T^2\left(\frac{x-\widetilde{t}}{h}\right) K_S^2\left(\frac{\boldsymbol{z}-\boldsymbol{s}_3}{b}\right) p_S(\boldsymbol{s}_3)\, d\boldsymbol{s}_3 \right\}^{\frac{1}{2}} \left[\int_{\mathcal{S}(\widetilde{t})} \frac{1}{p(\widetilde{t},\boldsymbol{s}_3)^2}\, d\boldsymbol{s}_3 \right]^{\frac{1}{2}} \\
&\stackrel{\text{(iii)}}{\leq} C_2|v| \left\{ \int_{\mathcal{S}} \left(x-\widetilde{t}\right)^2 K_T^2\left(\frac{x-\widetilde{t}}{h}\right) K_S^2\left(\frac{\boldsymbol{z}-\boldsymbol{s}_3}{b}\right) p_S(\boldsymbol{s}_3)\, d\boldsymbol{s}_3 \right\}^{\frac{1}{2}},
\end{aligned}
$$

where $C_1, C_2 > 0$ are some absolute constants under our Assumption A4. In addition, the inequality (i) follows from Cauchy-Schwarz inequality, the inequality (ii) utilizes the facts that $\bar{K}_T, p(t,\boldsymbol{s})$ is bounded and $p_T(\widetilde{t})$ is uniformly bounded away from 0 in $\mathcal{T}'$, and (iii) leverages the fact that the support $\mathcal{S}(t)$ of the conditional density $p(\boldsymbol{s}|t)$ is compact and $p(t,\boldsymbol{s})$ is uniformly bounded away from 0 within $\mathcal{S}(t)$ for any $t \in \mathcal{T}'$. Now, since the supports of K_T, K_S are compact (or K_T, K_S are square integrable) under Assumptions A6, we know that $\int_{\mathcal{S}} \left(x-\widetilde{t}\right)^2 K_T^2\left(\frac{x-\widetilde{t}}{h}\right) K_S^2\left(\frac{\boldsymbol{z}-\boldsymbol{s}_3}{b}\right) p_S(\boldsymbol{s}_3)\, d\boldsymbol{s}_3$ is again bounded, which implies that

$$
\left| \mathbb{E}_{\boldsymbol{U}_{i_3}} \left[\frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\Psi}_{\widetilde{t},\boldsymbol{S}_{i_3}}(v,x,\boldsymbol{z})}{p(\widetilde{t},\boldsymbol{S}_{i_3}) \cdot p_T(\widetilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\widetilde{t}-T_{i_3}}{\hbar}\right)\right] \right|
$$

$$
\leq C_2|v| \left\{ \int_{\mathcal{S}} \left[\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\psi}_{\tilde{t},\boldsymbol{s}_3}(x,\boldsymbol{z}) \right]^2 p_S(\boldsymbol{s}_3)\, d\boldsymbol{s}_3 \right\}^{\frac{1}{2}}
$$
$$
\leq C_2'|v|
$$

for some absolute constant $C_2' > 0$. Finally, plugging this result into our original display (44) yields that

$$
|f_{t_1}(v,x,\boldsymbol{z}) - f_{t_2}(v,x,\boldsymbol{z})| \leq C_3 \cdot |t_1 - t_2| \cdot |v|
$$

for some absolute constant $C_3 > 0$. Since $\mathbb{E}_Q|Y|^2 = \mathbb{E}_Q\left[(\mu(T,\boldsymbol{S})+\epsilon)^2\right] \leq 2\mathbb{E}_Q\left[\mu(T,\boldsymbol{S})^2\right] + 2\sigma^2 < \infty$ for any probability measure Q under Assumption A3 and the diameter of $\mathcal{T}'$ is a finite constant, we conclude from Example 19.7 in van der Vaart (1998) that

$$
\sup_Q N_{[]}\left(\widetilde{\mathcal{F}}, L_2(Q), \epsilon\right) \leq \frac{C_3'}{\epsilon}
$$

for some constant $C_3' > 0$. Additionally, since any 2ϵ bracket $[f_{t_1}, f_{t_2}]$ of $\mathcal{F}$ is contained in a ball of radius ϵ centered at $\frac{f_{t_1}+f_{t_2}}{2}$, we know that

$$
\sup_Q N\left(\widetilde{\mathcal{F}}, L_2(Q), \epsilon\right) \leq \sup_Q N_{[]}\left(\widetilde{\mathcal{F}}, L_2(Q), 2\epsilon\right) \leq \frac{C_4}{\epsilon}
$$

for some constant $C_4 > 0$. Thus, $\widetilde{\mathcal{F}}$ is a bounded VC-type class of functions. Finally, the above calculations tell us that for any $f_t \in \widetilde{\mathcal{F}}$,

$$
\begin{aligned}
|f_t(v,x,\boldsymbol{z})| &= \left| \mathbb{E}_{\boldsymbol{U}_{i_2}}\left[\int_{T_{i_2}}^t \mathbb{E}_{\boldsymbol{U}_{i_3}}\left[\frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\Psi}_{\tilde{t},\boldsymbol{S}_{i_3}}(v,x,\boldsymbol{z})}{p(\tilde{t},\boldsymbol{S}_{i_3}) \cdot p_T(\tilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\tilde{t}-T_{i_3}}{\hbar}\right) \right] d\tilde{t} \right] \right| \\
&\leq C_3 \cdot \mathbb{E}_{\boldsymbol{U}_{i_2}}\left[|t - T_{i_2}| \cdot |v|\right] \\
&\leq C_5 \cdot |v|,
\end{aligned}
$$

where $C_5 > 0$ is some absolute constant. Thus, an envelope function of $\widetilde{\mathcal{F}}$ can be given by $(v,x,\boldsymbol{z}) \mapsto F_1(v,x,\boldsymbol{z}) = C_5 \cdot |v|$. The similar argument applies to the case where $q > 2$ in the local polynomial regression for estimating $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$, because $\mathcal{K}_{q,d}$ in Assumption A6(b) is a bounded class of functions. The result thus follows for $\widetilde{\mathcal{F}}$.

As for $\widetilde{\mathcal{F}}_\theta$, we note that when $q = 2$,

$$
\begin{aligned}
g_t(v,x,\boldsymbol{z}) &= \sqrt{h^3 b^d} \cdot \bar{\varphi}_t(v,x,\boldsymbol{z}) \\
&= \mathbb{E}_{(T_{i_3},\boldsymbol{S}_{i_3})}\left[\frac{h \cdot \boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\Psi}_{t,\boldsymbol{S}_{i_3}}(Y,T,\boldsymbol{S})}{p(t,\boldsymbol{S}_{i_3}) \cdot p_T(t)} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{t-T_{i_3}}{\hbar}\right) \right] \\
&= \mathbb{E}_{(T_{i_3},\boldsymbol{S}_{i_3})}\left[\frac{C_K v\,(x-t)\,K_T\left(\frac{x-t}{h}\right) K_S\left(\frac{\boldsymbol{z}-\boldsymbol{S}_{i_3}}{b}\right)}{p(t,\boldsymbol{S}_{i_3}) \cdot p_T(t)} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{t-T_{i_3}}{\hbar}\right) \right]
\end{aligned}
$$

for each $g_t \in \widetilde{\mathcal{F}}_\theta$. Furthermore, by Assumptions A3 and A6, we know that the function inside the above expectation is Lipschitz with respect to $t \in \mathcal{T}'$, because $\frac{1}{p(t,\boldsymbol{S}_{i_3})\cdot p_T(t)}$ is differentiable with a bounded partial derivative with respect to t and the kernel functions are Lipschitz. Hence, $|g_{t_1}(v,x,\boldsymbol{z}) - g_{t_2}(v,x,\boldsymbol{z})| \leq C_6|v|$ for some constant $C_6 > 0$. Again, the analogous argument applies to the case where $q > 2$ in the local polynomial regression for estimating $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$. The result follows for $\widetilde{\mathcal{F}}_\theta$. $\square$

LEMMA 10 (Corollary 2.2 in Chernozhukov et al. 2014). *Let $\mathcal{G}$ be a collection of functions that is pointwise measurable and of VC-type (see Assumption A6) with an envelop function $\widetilde{G}$ and constants $A \geq e$ and $\upsilon \geq 1$. Suppose also that for some $\widetilde{A} \geq \widetilde{\sigma} > 0$ and $q' \in [4, \infty]$, we have $\sup_{\widetilde{g} \in \mathcal{G}} \mathbb{E}\left[\widetilde{g}(\boldsymbol{U})^k\right] \leq \widetilde{\sigma}^2 \cdot \widetilde{A}^{k-2}$ for $k = 2, 3, 4$ and $||G||_{P,q'} \leq \widetilde{A}$. Let $\mathbb{B}$ be a centered Gaussian process defined on $\mathcal{G}$ with covariance function*

$$\mathrm{Cov}\left(\mathbb{B}(\widetilde{g}_1), \mathbb{B}(\widetilde{g}_2)\right) = \mathbb{E}\left[\widetilde{g}_1(\boldsymbol{U}) \cdot \widetilde{g}_2(\boldsymbol{U})\right],$$

where $\widetilde{g}_1, \widetilde{g}_2 \in \mathcal{G}$ and $\boldsymbol{U} = (Y, T, \boldsymbol{S})$. Then, for every $\widetilde{\gamma} \in (0,1)$ and sufficiently large, there exists a random variable $\widetilde{B} \stackrel{d}{=} \sup_{\widetilde{g} \in \mathcal{G}} \mathbb{B}(\widetilde{g})$ such that

$$\mathrm{P}\left(\left|\sup_{\widetilde{g} \in \mathcal{G}} |\mathbb{G}_n(\widetilde{g})| - \widetilde{B}\right| > \frac{C_1 \cdot \widetilde{A}^{\frac{1}{3}} \widetilde{\sigma}^{\frac{2}{3}} \log^{\frac{2}{3}} n}{\widetilde{\gamma}^{\frac{1}{3}} n^{\frac{1}{6}}}\right) \leq C_2 \cdot \widetilde{\gamma},$$

where $C_1, C_2 > 0$ are two constants that only depend on q'. Here, $\widetilde{B}_1 \stackrel{d}{=} \widetilde{B}_2$ for two random variables $\widetilde{B}_1, \widetilde{B}_2$ means that they have the same distribution.

LEMMA 11 (Lemma 2.3 in Chernozhukov et al. 2014). *Under the same setup for $\mathcal{G}$ as in Lemma 10, we assume that there exist constants $\underline{\sigma}, \bar{\sigma} > 0$ such that $\underline{\sigma}^2 \leq \mathbb{E}\left[\widetilde{g}^2(\boldsymbol{U})\right] \leq \bar{\sigma}^2$ for all $\widetilde{g} \in \mathcal{G}$. Moreover, suppose that there exist constants $r_1, r_2 > 0$ such that*

$$\mathrm{P}\left(\left|\sup_{\widetilde{g} \in \mathcal{G}} |\mathbb{G}_n(f)| - \sup_{\widetilde{g} \in \mathcal{G}} |\mathbb{B}(f)|\right| > r_1\right) \leq r_2.$$

Then,

$$\begin{aligned} &\sup_{u \geq 0}\left|\mathrm{P}\left(\sup_{\widetilde{g} \in \mathcal{G}} |\mathbb{G}_n(f)| \leq u\right) - \mathrm{P}\left(\sup_{\widetilde{g} \in \mathcal{G}} |\mathbb{B}(f)| \geq u\right)\right| \\ &\leq C_{\sigma} r_1 \left\{\mathbb{E}\left[\sup_{\widetilde{g} \in \mathcal{G}} |\mathbb{B}(f)|\right] + \sqrt{\max\left\{1, \log\left(\frac{\underline{\sigma}}{r_1}\right)\right\}}\right\} + r_2, \end{aligned}$$

where $C_{\sigma} > 0$ is a constant depending only on $\underline{\sigma}$ and $\bar{\sigma}$.

LEMMA 12. *Let $d \geq 1$. If $h \asymp b \asymp n^{-\frac{1}{\gamma}}$ and $\hbar \asymp n^{-\frac{1}{\varpi}}$ for some $\gamma \geq \varpi > 0$ such that $\frac{nh^{d+5}}{\log n} \to c_1$ and $\frac{n\hbar^5}{\log n} \to c_2$ for some finite number $c_1, c_2 \geq 0$ and $\frac{n\hbar^2}{\log n}, \frac{\hbar}{h^{d+3} \log n}, \hbar n^{\frac{1}{4}}, \frac{\hbar^2}{h^{d+3}} \to \infty$ as $n \to \infty$, then*

$$\max\left\{\sqrt{nh^{d+7}}, \sqrt{\frac{\log n}{n\hbar^2}}, \sqrt{\frac{h^{d+3} \log n}{\hbar}}, \sqrt{\frac{h^{d+3}}{\hbar^2}}\right\} \lesssim \frac{\log^{\frac{2}{3}} n}{(nh^{d+3})^{\frac{1}{6}}} \tag{45}$$

when n is sufficiently large.

PROOF OF LEMMA 12. We consider controlling the four quantities on the left hand side of (45) by the right hand side separately.

- **Quantity I:** Note that $\sqrt{nh^{d+7}} \lesssim \frac{\log^{\frac{2}{3}} n}{(nh^{d+3})^{\frac{1}{6}}}$ is equivalent to $nh^{d+6} \lesssim \log n$. Under our conditions, we know that the slowest rate of convergence for h is $O\left(n^{-\frac{1}{d+5}}\right)$ up to some logarithmic factors of n. Thus, $nh^{d+6} = O\left(n^{-\frac{1}{d+5}}\right)$ will be dominated by $O(\log n)$.

• **Quantity II:** Note that $\sqrt{\frac{\log n}{n\hbar^2}} \lesssim \frac{\log^{\frac{2}{3}} n}{(nh^{d+3})^{\frac{1}{6}}}$ is equivalent to $\frac{h^{d+3}}{n^2\hbar^6} \lesssim \log n$. Again, under our conditions, the slowest rate of convergence for h is $O\left(n^{-\frac{1}{d+5}}\right)$ up to some logarithmic factors of n. Furthermore, given that $\hbar n^{\frac{1}{4}} \to \infty$, the fastest rate of convergence for $\hbar$ should be smaller than the order $O\left(n^{-\frac{1}{3}}\right)$. Thus, under these rates, $\frac{h^{d+3}}{n^2\hbar^6}$ is dominated by $O\left(n^{-\frac{d+3}{d+5}}\right)$ and thus by $O(\log n)$.

• **Quantity III:** Note that $\sqrt{\frac{h^{d+3}\log n}{\hbar}} \lesssim \frac{\log^{\frac{2}{3}} n}{(nh^{d+3})^{\frac{1}{6}}}$ is equivalent to $\frac{nh^{4(d+6)}}{\hbar^3} \lesssim \log n$. Again, under our conditions, the slowest rates of convergence for h is $O\left(n^{-\frac{1}{d+5}}\right)$, and the fastest rate of convergence for $\hbar$ is smaller than the order $O\left(n^{-\frac{1}{3}}\right)$ up to some logarithmic factors of n. Under these rates, $\frac{nh^{4(d+6)}}{\hbar^3} = O\left(n^{-\frac{2d+2}{d+5}}\right)$ will be dominated by $O(\log n)$.

• **Quantity IV:** Note that $\sqrt{\frac{h^{d+3}}{\hbar^2}} \lesssim \frac{\log^{\frac{2}{3}} n}{(nh^{d+3})^{\frac{1}{6}}}$ is equivalent to $n^{\frac{1}{2}}h^{2d+6}\hbar^{-3} \lesssim \log^2 n$. Again, under our conditions, the slowest rate of convergence for h is $O\left(n^{-\frac{1}{d+5}}\right)$ up to some logarithmic factors of n. Furthermore, given that $\hbar n^{\frac{1}{4}} \to \infty$, the fastest rate of convergence for $\hbar$ should be smaller than the order $O\left(n^{-\frac{1}{4}}\right)$. Thus, $n^{\frac{1}{2}}h^{2d+6}\hbar^{-3} = O\left(n^{\frac{1-3d}{4(d+5)}}\right)$, which is dominated by $O\left(\log^2 n\right)$ when $d \geq 1$.

In summary, the result follows from combining the above four cases. $\square$

LEMMA 13. *Let $\mathcal{T}' \subset \mathcal{T}$ be a compact set so that $p_T(t)$ is uniformly bounded away from 0 within $\mathcal{T}'$. Suppose that Assumptions A3, A4, A5, and A6 hold. Assume that* $\mathrm{Var}(\epsilon) = \sigma^2 > 0$ *under the general confounding model (2). Then, when $h, b, \hbar$ are sufficiently small as $n \to \infty$, there exist constants $\underline{\sigma}, \bar{\sigma} > 0$ such that $\underline{\sigma}^2 \leq \mathbb{E}\left[\varphi_t(Y,T,\boldsymbol{S})^2\right] \leq \bar{\sigma}^2$ for any $t \in \mathcal{T}'$, where φ_t is defined in (17). Here, all the constants $\underline{\sigma}, \bar{\sigma} > 0$ are independent of $h, b, \hbar$ and n. Furthermore, the same upper and lower bounds apply to* $\mathrm{Var}\left[\varphi_t(Y,T,\boldsymbol{S})\right]$. *Finally, the same result holds true when we replace φ_t with $\bar{\varphi}_t$ for any $t \in \mathcal{T}'$.*

PROOF OF LEMMA 13. We first compute the upper bound on $\mathbb{E}\left[\varphi_t(Y,T,\boldsymbol{S})^2\right]$. Recall from (17) that

$$
\begin{aligned}
&\mathbb{E}\left[\varphi_t(Y,T,\boldsymbol{S})^2\right] \\
&= \mathbb{E}\left\{\left[\mathbb{E}_{T_{i_2}}\left(\int_{T_{i_2}}^t \mathbb{E}_{(T_{i_3},\boldsymbol{S}_{i_3})}\left[\frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\Psi}_{\tilde{t},\boldsymbol{S}_{i_3}}(Y,T,\boldsymbol{S})}{\sqrt{hb^d}\cdot p(\tilde{t},\boldsymbol{S}_{i_3})\cdot p_T(\tilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\tilde{t}-T_{i_3}}{\hbar}\right)\right] d\tilde{t}\right)\right]^2\right\} \\
&= \mathbb{E}\left\{\frac{1}{hb^d}\left[\mathbb{E}_{T_{i_2}}\left(\int_{T_{i_2}}^t \int_{\mathcal{T}\times\mathcal{S}} \frac{\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} \boldsymbol{\psi}_{\tilde{t},\boldsymbol{s}_3}(T,\boldsymbol{S})\cdot Y}{p(\tilde{t},\boldsymbol{s}_3)\cdot p_T(\tilde{t})} \cdot \frac{1}{\hbar}\bar{K}_T\left(\frac{\tilde{t}-t_3}{\hbar}\right) p(t_3,\boldsymbol{s}_3)\, dt_3 d\boldsymbol{s}_3 d\tilde{t}\right)\right]^2\right\}.
\end{aligned}
$$

By the definition of $\boldsymbol{M}_q$ in (14) when $q=2$, we know that $\boldsymbol{e}_2^T \boldsymbol{M}_q^{-1} = \left(0, C_K, 0, \underbrace{0,...,0}_{d}\right)$ for some constant $C_K \neq 0$ that only depends on the kernels K_T, K_S. Then, we proceed the above display as:

$$
\begin{aligned}
&\mathbb{E}\left[\varphi_t(Y,T,\boldsymbol{S})^2\right] \\
&= \frac{C_K^2}{hb^d}\cdot\mathbb{E}\left\{\left[\mathbb{E}_{T_{i_2}}\left(\int_{T_{i_2}}^t\int_{\mathbb{R}\times\mathcal{S}} \frac{Y\left(\frac{T-\tilde{t}}{h}\right)K_T\left(\frac{T-\tilde{t}}{h}\right)K_S\left(\frac{\boldsymbol{S}-\boldsymbol{s}_3}{b}\right)}{p(\tilde{t},\boldsymbol{s}_3)\cdot p_T(\tilde{t})}\cdot\bar{K}_T(u)\cdot p(\tilde{t}+u\hbar,\boldsymbol{s}_3)\,dud\boldsymbol{s}_3d\tilde{t}\right)\right]^2\right\} \\
&\le \frac{C_K^2}{hb^d}\cdot\mathbb{E}\left[\mathbb{E}_{T_{i_2}}\left(\int_{T_{i_2}}^t\int_{\mathbb{R}\times\mathcal{S}} \frac{Y^2\left(\frac{T-\tilde{t}}{h}\right)^2K_T^2\left(\frac{T-\tilde{t}}{h}\right)K_S^2\left(\frac{\boldsymbol{S}-\boldsymbol{s}_3}{b}\right)}{p(\tilde{t},\boldsymbol{s}_3)^2\cdot p_T(\tilde{t})^2}\cdot\bar{K}_T^2(u)\cdot p(\tilde{t}+u\hbar,\boldsymbol{s}_3)^2\,dud\boldsymbol{s}_3d\tilde{t}\right)\right] \\
&\le \frac{2C_K^2}{hb^d}\cdot\int_{\mathcal{T}\times\mathcal{S}}\int_{\mathcal{T}}\int_{t_2}^t\int_{\mathbb{R}\times\mathcal{S}} \frac{\left[\mu(t_1,\boldsymbol{s}_1)^2+\sigma^2\right]\left(\frac{t_1-\tilde{t}}{h}\right)^2K_T^2\left(\frac{t_1-\tilde{t}}{h}\right)K_S^2\left(\frac{\boldsymbol{s}_1-\boldsymbol{s}_3}{b}\right)}{p(\tilde{t},\boldsymbol{s}_3)^2\cdot p_T(\tilde{t})^2} \\
&\qquad\times\bar{K}_T^2(u)\cdot p(\tilde{t}+u\hbar,\boldsymbol{s}_3)^2\,dud\boldsymbol{s}_3d\tilde{t}\cdot p_T(t_2)\,dt_2\cdot p(t_1,\boldsymbol{s}_1)\,dt_1d\boldsymbol{s}_1 \\
&= 2C_K^2\int_{\mathcal{T}}\int_{t_2}^t\int_{\mathbb{R}\times\mathcal{S}}\int_{\mathbb{R}\times\mathbb{R}^d} \frac{\left[\mu(\tilde{t}+xh,\boldsymbol{s}_3+b\boldsymbol{z})^2+\sigma^2\right]x^2K_T^2(x)\,K_S^2(\boldsymbol{z})}{p(\tilde{t},\boldsymbol{s}_3)^2\cdot p_T(\tilde{t})^2} \\
&\qquad\times\bar{K}_T^2(u)\cdot p(\tilde{t}+u\hbar,\boldsymbol{s}_3)^2p(\tilde{t}+xh,\boldsymbol{s}_3+b\boldsymbol{z})\,dxd\boldsymbol{z}\,dud\boldsymbol{s}_3d\tilde{t}\cdot p_T(t_2)\,dt_2 \\
&<\infty,
\end{aligned}
$$

because all the terms inside the integrand are bounded under Assumptions A3, A4, and A6. Hence, $\mathbb{E}\left[\varphi_t(Y,T,\boldsymbol{S})^2\right]$ can be upper bounded by a constant $\bar{\sigma}^2>0$ that is independent of the bandwidths $h,b,\hbar$.

For the lower bound on $\mathbb{E}\left[\varphi_t(Y,T,\boldsymbol{S})^2\right]$, we compute that

$$
\begin{aligned}
&\mathbb{E}\left[\varphi_t(Y,T,\boldsymbol{S})^2\right] \\
&= \frac{C_K^2}{hb^d}\cdot\mathbb{E}\left\{\left[\mathbb{E}_{T_{i_2}}\left(\int_{T_{i_2}}^t\int_{\mathbb{R}\times\mathcal{S}} \frac{[\mu(T,\boldsymbol{S})+\epsilon]\left(\frac{T-\tilde{t}}{h}\right)K_T\left(\frac{T-\tilde{t}}{h}\right)K_S\left(\frac{\boldsymbol{S}-\boldsymbol{s}_3}{b}\right)}{p(\tilde{t},\boldsymbol{s}_3)\cdot p_T(\tilde{t})}\cdot\bar{K}_T(u)\cdot p(\tilde{t}+u\hbar,\boldsymbol{s}_3)\,dud\boldsymbol{s}_3d\tilde{t}\right)\right]^2\right\} \\
&= \frac{C_K^2}{hb^d}\cdot\mathbb{E}\left\{\left[\mathbb{E}_{T_{i_2}}\left(\int_{T_{i_2}}^t\int_{\mathbb{R}\times\mathcal{S}} \frac{\mu(T,\boldsymbol{S})\left(\frac{T-\tilde{t}}{h}\right)K_T\left(\frac{T-\tilde{t}}{h}\right)K_S\left(\frac{\boldsymbol{S}-\boldsymbol{s}_3}{b}\right)}{p(\tilde{t},\boldsymbol{s}_3)\cdot p_T(\tilde{t})}\cdot\bar{K}_T(u)\cdot p(\tilde{t}+u\hbar,\boldsymbol{s}_3)\,dud\boldsymbol{s}_3d\tilde{t}\right)\right]^2\right\} \\
&\quad+\frac{C_K^2\sigma^2}{hb^d}\cdot\mathbb{E}\left\{\left[\mathbb{E}_{T_{i_2}}\left(\int_{T_{i_2}}^t\int_{\mathbb{R}\times\mathcal{S}} \frac{\left(\frac{T-\tilde{t}}{h}\right)K_T\left(\frac{T-\tilde{t}}{h}\right)K_S\left(\frac{\boldsymbol{S}-\boldsymbol{s}_3}{b}\right)}{p(\tilde{t},\boldsymbol{s}_3)\cdot p_T(\tilde{t})}\cdot\bar{K}_T(u)\cdot p(\tilde{t}+u\hbar,\boldsymbol{s}_3)\,dud\boldsymbol{s}_3d\tilde{t}\right)\right]^2\right\} \\
&= \frac{C_K^2}{hb^d}\int_{\mathcal{T}\times\mathcal{S}}\left[\mathbb{E}_{T_{i_2}}\left(\int_{T_{i_2}}^t\int_{\mathbb{R}\times\mathcal{S}} \frac{\mu(t_1,\boldsymbol{s}_1)\left(\frac{t_1-\tilde{t}}{h}\right)K_T\left(\frac{t_1-\tilde{t}}{h}\right)K_S\left(\frac{\boldsymbol{s}_1-\boldsymbol{s}_3}{b}\right)}{p(\tilde{t},\boldsymbol{s}_3)\cdot p_T(\tilde{t})}\bar{K}_T(u)\,p(\tilde{t}+u\hbar,\boldsymbol{s}_3)dud\boldsymbol{s}_3d\tilde{t}\right)\right]^2p(t_1,\boldsymbol{s}_1)dt_1d\boldsymbol{s}_1 \\
&\quad+\frac{C_K^2\sigma^2}{hb^d}\int_{\mathcal{T}\times\mathcal{S}}\left[\mathbb{E}_{T_{i_2}}\left(\int_{T_{i_2}}^t\int_{\mathbb{R}\times\mathcal{S}} \frac{\left(\frac{t_1-\tilde{t}}{h}\right)K_T\left(\frac{t_1-\tilde{t}}{h}\right)K_S\left(\frac{\boldsymbol{s}_1-\boldsymbol{s}_3}{b}\right)}{p(\tilde{t},\boldsymbol{s}_3)\cdot p_T(\tilde{t})}\cdot\bar{K}_T(u)\cdot p(\tilde{t}+u\hbar,\boldsymbol{s}_3)\,dud\boldsymbol{s}_3d\tilde{t}\right)\right]^2p(t_1,\boldsymbol{s}_1)dt_1d\boldsymbol{s}_1 \\
&= \frac{C_K^2}{hb^d}\int_{\mathcal{T}\times\mathcal{S}}\left[\mathbb{E}_{T_{i_2}}\left(\int_{T_{i_2}}^t\int_{\mathbb{R}\times\mathcal{S}} \frac{\mu(t_1,\boldsymbol{s}_1)\left(\frac{t_1-\tilde{t}}{h}\right)K_T\left(\frac{t_1-\tilde{t}}{h}\right)K_S\left(\frac{\boldsymbol{s}_1-\boldsymbol{s}_3}{b}\right)}{p(\tilde{t},\boldsymbol{s}_3)\cdot p_T(\tilde{t})}\bar{K}_T(u)\,p(\tilde{t}+u\hbar,\boldsymbol{s}_3)\sqrt{p(t_1,\boldsymbol{s}_1)}\,dud\boldsymbol{s}_3d\tilde{t}\right)\right]^2dt_1d\boldsymbol{s}_1
\end{aligned}
$$

$$+\frac{C_K^2\sigma^2}{hb^d}\int_{\mathcal{T}\times\mathcal{S}}\left[\mathbb{E}_{T_{i_2}}\left(\int_{T_{i_2}}^t\int_{\mathbb{R}\times\mathcal{S}}\frac{\left(\frac{t_1-\widetilde{t}}{h}\right)K_T\left(\frac{t_1-\widetilde{t}}{h}\right)K_S\left(\frac{\boldsymbol{s}_1-\boldsymbol{s}_3}{b}\right)}{p(\widetilde{t},\boldsymbol{s}_3)\cdot p_T(\widetilde{t})}\bar{K}_T(u)\,p(\widetilde{t}+u\hbar,\boldsymbol{s}_3)\sqrt{p(t_1,\boldsymbol{s}_1)}\,dud\boldsymbol{s}_3d\widetilde{t}\right)\right]^2dt_1d\boldsymbol{s}_1$$

$$\overset{\text{(i)}}{\geq}C_K^2\sigma^2\int_{\mathbb{R}\times\mathbb{R}^d}\left[\mathbb{E}_{T_{i_2}}\left(\int_{T_{i_2}}^t\int_{\mathbb{R}\times\mathcal{S}}\frac{xK_T(x)\cdot K_S(\boldsymbol{z})}{p(\widetilde{t},\boldsymbol{s}_3)\cdot p_T(\widetilde{t})}\cdot\bar{K}_T(u)\,p(\widetilde{t}+u\hbar,\boldsymbol{s}_3)\sqrt{p(\widetilde{t}+hx,\boldsymbol{s}_3+b\boldsymbol{z})}\,dud\boldsymbol{s}_3d\widetilde{t}\right)\right]^2dxd\boldsymbol{z},$$

where (i) applies the changes of variables $x=\frac{t_1-\widetilde{t}}{h}$, $\boldsymbol{z}=\frac{\boldsymbol{s}_1-\boldsymbol{s}_3}{b}$. Now, since the variance $\sigma^2=\mathbb{E}(\epsilon)>0$, we have that for any $t\in\mathcal{T}'$,

(46)

$$\begin{aligned}
&\mathbb{E}\left[\varphi_t(Y,T,\boldsymbol{S})^2\right]\\
&\geq C_K^2\sigma^2\int_{\mathbb{R}\times\mathbb{R}^d}x^2K_T^2(x)\,K_S^2(\boldsymbol{z})\left[\mathbb{E}_{T_{i_2}}\left(\int_{T_{i_2}}^t\int_{\mathbb{R}\times\mathcal{S}}\frac{\bar{K}_T(u)\,p(\widetilde{t}+u\hbar,\boldsymbol{s}_3)\sqrt{p(\widetilde{t}+hx,\boldsymbol{s}_3+b\boldsymbol{z})}}{p(\widetilde{t},\boldsymbol{s}_3)\cdot p_T(\widetilde{t})}\,dud\boldsymbol{s}_3d\widetilde{t}\right)\right]^2dxd\boldsymbol{z}\\
&=\widetilde{C}_K^2\sigma^2\int_{\mathbb{R}^d}K_S^2(\boldsymbol{z})\left[\mathbb{E}_{T_{i_2}}\left(\int_{T_{i_2}}^t\int_{\mathbb{R}\times\mathcal{S}}\frac{\bar{K}_T(u)\,p(\widetilde{t}+u\hbar,\boldsymbol{s}_3)\sqrt{p(\widetilde{t},\boldsymbol{s}_3)}}{p(\widetilde{t},\boldsymbol{s}_3)\cdot p_T(\widetilde{t})}\,dud\boldsymbol{s}_3d\widetilde{t}\right)\right]^2d\boldsymbol{z}+O\left(\max\{h,b\}\right)\\
&=\widetilde{C}_K^2\sigma^2\left[\mathbb{E}_{T_{i_2}}\left(\int_{T_{i_2}}^t\int_{\mathcal{S}(\widetilde{t})}\frac{\sqrt{p(\widetilde{t},\boldsymbol{s}_3)}}{p_T(\widetilde{t})}\,d\boldsymbol{s}_3d\widetilde{t}\right)\right]^2+O\left(\max\{h,b\}+\hbar^2\right)
\end{aligned}$$

where $\widetilde{C}_K^2>0$ is again some constant that only depends on the kernels $K_T,K_S,\bar{K}_T$. By Assumption A4, $p(t,\boldsymbol{s})$ is upper bounded and also uniformly bounded away from 0 within it support $\mathcal{E}$. Thus, there exists a constant $C'>0$ such that

$$\int_{\mathcal{S}(\widetilde{t})}\frac{\sqrt{p(\widetilde{t},\boldsymbol{s}_3)}}{p_T(\widetilde{t})}\,d\boldsymbol{s}_3\geq C' \tag{47}$$

for some $\widetilde{t}\in\mathcal{T}'$. Let A_T be an subset of $\mathcal{T}$ on which (47) holds. By the non-degeneracy of p_T under Assumption A4, we can proceed (46) as:

$$\mathbb{E}\left[\varphi_t(Y,T,\boldsymbol{S})^2\right]\geq\widetilde{C}_K^2\sigma^2C'^2\left\{\mathbb{E}_{T_{i_2}}\left[(t-T_{i_2})\cdot\mathbb{1}_{A_T}\right]\right\}^2+O\left(\max\{h,b\}\right)+O(\hbar^2)>\underline{\sigma}^2$$

for some constant $\underline{\sigma}^2>0$.

The similar argument applies to the case where $q>2$ in the local polynomial regression for estimating $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$, because $\mathcal{K}_{q,d}$ in Assumption A6(b) is a bounded class of functions. Furthermore, $\mathbb{E}\left[\varphi_t(Y,T,\boldsymbol{S})\right]\to 0$ as $h,b\to 0$ with $n\to\infty$. The result thus follows for the influence function φ_t of $\widehat{m}_\theta(t)$.

As for the influence function $\bar{\varphi}_t$ of $\widehat{\theta}_C(t)$, we use the similar arguments to obtain its upper bound as:

$$\begin{aligned}
&\mathbb{E}\left[\bar{\varphi}_t(Y,T,\boldsymbol{S})^2\right]\\
&=\mathbb{E}\left\{\left[\mathbb{E}_{(T_{i_3},\boldsymbol{S}_{i_3})}\left(\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\Psi}_{t,\boldsymbol{S}_{i_3}}(Y,T,\boldsymbol{S})}{\sqrt{hb^d}\cdot p(t,\boldsymbol{S}_{i_3})\cdot p_T(t)}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{t-T_{i_3}}{\hbar}\right)\right)\right]^2\right\}\\
&=\frac{1}{hb^d}\cdot\mathbb{E}\left\{\left[\int_{\mathcal{T}\times\mathcal{S}}\frac{\boldsymbol{e}_2^T\boldsymbol{M}_q^{-1}\boldsymbol{\psi}_{t,\boldsymbol{s}_3}(T,\boldsymbol{S})\cdot Y}{p(t,\boldsymbol{s}_3)\cdot p_T(t)}\cdot\frac{1}{\hbar}\bar{K}_T\left(\frac{t-t_3}{\hbar}\right)p(t_3,\boldsymbol{s}_3)\,dt_3d\boldsymbol{s}_3\right]^2\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \frac{C_K^2}{hb^d} \cdot \mathbb{E}\left\{\left[\int_{\mathbb{R}\times\mathcal{S}} \frac{Y\left(\frac{T-t}{h}\right) K_T\left(\frac{T-t}{h}\right) K_S\left(\frac{\boldsymbol{S}-\boldsymbol{s}_3}{b}\right)}{p(t,\boldsymbol{s}_3)\cdot p_T(t)} \cdot \bar{K}_T(u)\cdot p(t+u\hbar,\boldsymbol{s}_3)\,dud\boldsymbol{s}_3\right]^2\right\} \\
&\leq \frac{C_K^2}{hb^d} \cdot \mathbb{E}\left[\int_{\mathbb{R}\times\mathcal{S}} \frac{Y^2\left(\frac{T-t}{h}\right)^2 K_T^2\left(\frac{T-t}{h}\right) K_S^2\left(\frac{\boldsymbol{S}-\boldsymbol{s}_3}{b}\right)}{p(t,\boldsymbol{s}_3)^2\cdot p_T(t)^2} \cdot \bar{K}_T^2(u)\cdot p(t+u\hbar,\boldsymbol{s}_3)^2\,dud\boldsymbol{s}_3\right] \\
&\leq \frac{2C_K^2}{hb^d} \int_{\mathcal{T}\times\mathcal{S}}\int_{\mathbb{R}\times\mathcal{S}} \frac{\left[\mu(t_1,\boldsymbol{s}_1)^2+\sigma^2\right]\left(\frac{t_1-t}{h}\right)^2 K_T^2\left(\frac{t_1-t}{h}\right) K_S^2\left(\frac{\boldsymbol{s}_1-\boldsymbol{s}_3}{b}\right)}{p(t,\boldsymbol{s}_3)^2\cdot p_T(t)^2}\cdot \bar{K}_T^2(u) \\
&\quad \times p(t+u\hbar,\boldsymbol{s}_3)^2\cdot p(t_1,\boldsymbol{s}_1)\,dud\boldsymbol{s}_3dt_1d\boldsymbol{s}_1 \\
&= 2C_K^2 \int_{\mathbb{R}\times\mathcal{S}}\int_{\mathbb{R}\times\mathbb{R}^d} \frac{\left[\mu(t+hx,\boldsymbol{s}_3+b\boldsymbol{z})^2+\sigma^2\right] x^2 K_T^2(x)\, K_S^2(\boldsymbol{z})}{p(t,\boldsymbol{s}_3)^2\cdot p_T(t)^2}\cdot \bar{K}_T^2(u) \\
&\quad \times p(t+u\hbar,\boldsymbol{s}_3)^2\cdot p(t+hx,\boldsymbol{s}_3+b\boldsymbol{z})\,dxd\boldsymbol{z}dud\boldsymbol{s}_3 \\
&< \infty
\end{aligned}
$$

when $q=2$ under Assumptions A3, A4, and A6. Hence, $\mathbb{E}\left[\bar{\varphi}_t(Y,T,\boldsymbol{S})^2\right]$ can be upper bounded by a constant $\bar{\sigma}^2>0$ that is independent of the bandwidths $h,b,\hbar$.

For the lower bound on $\mathbb{E}\left[\bar{\varphi}_t(Y,T,\boldsymbol{S})^2\right]$, we again compute that

$$
\begin{aligned}
&\mathbb{E}\left[\bar{\varphi}_t(Y,T,\boldsymbol{S})^2\right] \\
&= \frac{C_K^2}{hb^d} \cdot \mathbb{E}\left\{\left[\int_{\mathbb{R}\times\mathcal{S}} \frac{\left[\mu(T,\boldsymbol{S})+\epsilon\right]\left(\frac{T-t}{h}\right) K_T\left(\frac{T-t}{h}\right) K_S\left(\frac{\boldsymbol{S}-\boldsymbol{s}_3}{b}\right)}{p(t,\boldsymbol{s}_3)\cdot p_T(t)} \cdot \bar{K}_T(u)\cdot p(t+u\hbar,\boldsymbol{s}_3)\,dud\boldsymbol{s}_3\right]^2\right\} \\
&\geq \frac{C_K^2\sigma^2}{hb^d}\int_{\mathcal{T}\times\mathcal{S}}\left[\int_{\mathbb{R}\times\mathcal{S}} \frac{\left(\frac{t_1-t}{h}\right) K_T\left(\frac{t_1-t}{h}\right) K_S\left(\frac{\boldsymbol{s}_1-\boldsymbol{s}_3}{b}\right)}{p(t,\boldsymbol{s}_3)\cdot p_T(t)}\cdot \bar{K}_T(u)\cdot p(t+u\hbar,\boldsymbol{s}_3)\,dud\boldsymbol{s}_3\right]^2 p(t_1,\boldsymbol{s}_1)\,dt_1d\boldsymbol{s}_1 \\
&= \frac{C_K^2\sigma^2}{hb^d}\int_{\mathcal{T}\times\mathcal{S}}\left[\int_{\mathbb{R}\times\mathcal{S}} \frac{\left(\frac{t_1-t}{h}\right) K_T\left(\frac{t_1-t}{h}\right) K_S\left(\frac{\boldsymbol{s}_1-\boldsymbol{s}_3}{b}\right)}{p(t,\boldsymbol{s}_3)\cdot p_T(t)}\cdot \bar{K}_T(u)\cdot p(t+u\hbar,\boldsymbol{s}_3)\sqrt{p(t_1,\boldsymbol{s}_1)}\,dud\boldsymbol{s}_3\right]^2 dt_1d\boldsymbol{s}_1 \\
&\geq C_K^2\sigma^2\int_{\mathbb{R}\times\mathbb{R}^d}\left[\int_{\mathbb{R}\times\mathcal{S}} \frac{xK_T(x)\,K_S(\boldsymbol{z})}{p(t,\boldsymbol{s}_3)\cdot p_T(t)}\cdot \bar{K}_T(u)\cdot p(t+u\hbar,\boldsymbol{s}_3)\sqrt{p(t+hx,\boldsymbol{s}_3+b\boldsymbol{z})}\,dud\boldsymbol{s}_3\right]^2 dxd\boldsymbol{z} \\
&= \widetilde{C}_K^2\sigma^2\int_{\mathbb{R}\times\mathbb{R}^d} x^2K_T^2(x)K_S^2(\boldsymbol{z})\left[\int_{\mathbb{R}\times\mathcal{S}} \frac{\bar{K}_T(u)\cdot p(t,\boldsymbol{s}_3)\sqrt{p(t,\boldsymbol{s}_3)}}{p(t,\boldsymbol{s}_3)\cdot p_T(t)}\,dud\boldsymbol{s}_3\right]^2 dxd\boldsymbol{z} + O\left(\max\{h,b,\hbar^2\}\right) \\
&= \widetilde{C}_K^2\sigma^2\int_{\mathbb{R}\times\mathbb{R}^d} x^2K_T^2(x)K_S^2(\boldsymbol{z})\left[\int_{\mathcal{S}(t)} \frac{\sqrt{p(t,\boldsymbol{s}_3)}}{p_T(t)}\,d\boldsymbol{s}_3\right]^2 dxd\boldsymbol{z} + O\left(\max\{h,b,\hbar^2\}\right) \\
&\geq \underline{\sigma}^2
\end{aligned}
$$

for some $\underline{\sigma}^2>0$ by (47) when $h,b,\hbar$ are sufficiently small as $n\to\infty$. The result thus follows for the influence function $\bar{\varphi}_t$ of $\widehat{\theta}_C(t)$. ☐

THEOREM 6 (Gaussian approximation). *Let $q\geq 2$ in the local polynomial regression for estimating $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$ and $\mathcal{T}'\subset\mathcal{T}$ be a compact set so that $p_T(t)$ is uniformly bounded away* *from 0 within $\mathcal{T}'$. Suppose that Assumptions A1, A2, A3, A4, A5, and A6 hold. If $h \asymp b \asymp n^{-\frac{1}{\gamma}}$ and $\hbar \asymp n^{-\frac{1}{\varpi}}$ for some $\gamma \geq \varpi > 0$ such that $\frac{nh^{d+5}}{\log n} \to c_1$ and $\frac{n\hbar^5}{\log n} \to c_2$ for some finite number $c_1, c_2 \geq 0$ and $\frac{n\hbar^2}{\log n}, \frac{\hbar}{h^{d+3}\log n}, \hbar n^{\frac{1}{4}}, \frac{\hbar^2}{h^{d+3}} \to \infty$ as $n \to \infty$, then there exist Gaussian processes $\mathbb{B}, \bar{\mathbb{B}}$ such that*

$$\sup_{u\geq 0}\left|\mathrm{P}\left(\sqrt{nh^3b^d}\cdot\sup_{t\in\mathcal{T}'}|\widehat{m}_\theta(t)-m(t)|\leq u\right)-\mathrm{P}\left(\sup_{f\in\mathcal{F}}|\mathbb{B}(f)|\leq u\right)\right| = O\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right),$$

$$\sup_{u\geq 0}\left|\mathrm{P}\left(\sqrt{nh^3b^d}\cdot\sup_{t\in\mathcal{T}'}\left|\widehat{\theta}_C(t)-\theta_C(t)\right|\leq u\right)-\mathrm{P}\left(\sup_{g\in\mathcal{F}_\theta}|\bar{\mathbb{B}}(g)|\leq u\right)\right| = O\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right),$$

where $\mathcal{F}, \mathcal{F}_\theta$ are defined in (18).

PROOF OF THEOREM 6. We only prove the Gaussian approximation for $\widehat{m}_\theta(t)$, since the result for its derivative estimator $\widehat{\theta}_C(t)$ follows from an identical argument. At a high level, given the asymptotic linearity of $\widehat{m}_\theta(t)$ in Lemma 5, we will use Lemma 10 to establish the coupling between $\sup_{f\in\mathcal{F}}|\mathbb{G}_n(f)|$ and $\sup_{f\in\mathcal{F}}|\mathbb{B}(f)|$ for the Gaussian process $\mathbb{B}$ defined in the theorem statement and then utilize Lemma 11 to translate the coupling to a bound on the Kolmogorov distance between $\sqrt{nh^3b^d}\cdot\sup_{t\in\mathcal{T}'}|\widehat{m}_\theta(t)-m(t)|$ and $\sup_{f\in\mathcal{F}}|\mathbb{B}(f)|$.

By Lemma 9, we know that the class of scaled influence functions

$$\widetilde{\mathcal{F}} = \left\{(v,x,\boldsymbol{z})\mapsto\sqrt{h^3b^d}\cdot\varphi_t(v,x,\boldsymbol{z}): t\in\mathcal{T}'\right\} = \left\{\sqrt{h^3b^d}\cdot f: f\in\mathcal{F}\right\}$$

is a VC-type class with an envelope function $(v,x,\boldsymbol{z})\mapsto F_1(v,x,\boldsymbol{z}) = C_5\cdot|v|$ for some constant $C_5>0$ that is independent of $n,h,b,\hbar$. In addition, recalling the definition of φ_t in (17) together with our Assumption A6 on moments of the kernel functions K_T, K_S and Lemma 13, we obtain that

$$\sup_{f\in\widetilde{\mathcal{F}}}\mathbb{E}\left[f(\boldsymbol{U})^2\right]\leq C_6^2\cdot h^3b^d := \widetilde{\sigma}^2<\infty$$

$$\text{and}\quad\left[\mathbb{E}\left(C_5|Y|\right)^4\right]^{\frac{1}{4}}\leq C_5'\left(\mathbb{E}\left|\mu(T,\boldsymbol{S})\right|^4+\mathbb{E}|\epsilon|^4\right)^{\frac{1}{4}} := \widetilde{A}<\infty,$$

where $C_5', C_6>0$ are some constants that are independent of $n,h,b,\hbar$. By Lemma 10, we know that for any $\widetilde{\gamma}\in(0,1)$,

$$\mathrm{P}\left(\left|\sup_{f\in\widetilde{\mathcal{F}}}|\mathbb{G}_n(f)|-\sup_{f\in\widetilde{\mathcal{F}}}|\mathbb{B}(f)|\right|>\frac{C_1\cdot\widetilde{A}^{\frac{1}{3}}\left(h^3b^d\right)^{\frac{1}{3}}\log^{\frac{2}{3}}n}{\widetilde{\gamma}^{\frac{1}{3}}n^{\frac{1}{6}}}\right)\leq\frac{C_2\cdot\widetilde{\gamma}}{2}.$$

Dividing $\sqrt{h^3b^d}$ on both sides of the inequality inside P gives us that

$$\mathrm{P}\left(\left|\sup_{t\in\mathcal{T}'}|\mathbb{G}_n(\varphi_t)|-\sup_{f\in\mathcal{F}}|\mathbb{B}(f)|\right|>\frac{C_1\cdot\widetilde{A}^{\frac{1}{3}}\log^{\frac{2}{3}}n}{\widetilde{\gamma}^{\frac{1}{3}}\left(nh^3b^d\right)^{\frac{1}{6}}}\right)\leq\frac{C_2\cdot\widetilde{\gamma}}{2}.$$

On the other hand, we know from Lemma 5 that for any $\widetilde{\gamma}\in(0,1)$,

$$\mathrm{P}\left(\frac{\left|\sqrt{nh^3b^d}\sup_{t\in\mathcal{T}'}|\widehat{m}_\theta(t)-m(t)|-\sup_{t\in\mathcal{T}'}|\mathbb{G}_n\varphi_t|\right|}{\max\left\{\sqrt{nh^{d+7}},\sqrt{\frac{\log n}{n\hbar^2}},\sqrt{\frac{h^{d+3}\log n}{\hbar}},\sqrt{\frac{h^{d+3}}{\hbar^2}}\right\}}>\frac{C_3}{\widetilde{\gamma}^{\frac{1}{3}}}\right)\leq\frac{C_2\cdot\widetilde{\gamma}}{2},$$

where $C_3 > 0$ is some large constant. Combining the above two inequalities yields that

$$\mathrm{P}\Bigg(\left|\sqrt{nh^3b^d}\sup_{t\in\mathcal{T}'}|\widehat{m}_\theta(t)-m(t)| - \sup_{f\in\mathcal{F}}|\mathbb{B}(f)|\right| > \frac{C_3}{\widetilde{\gamma}^{\frac{1}{3}}}\cdot\max\left\{\sqrt{nh^{d+7}}, \sqrt{\frac{\log n}{n\hbar^2}}, \sqrt{\frac{h^{d+3}\log n}{\hbar}}, \sqrt{\frac{h^{d+3}}{\hbar^2}}\right\} + \frac{C_1'\cdot\log^{\frac{2}{3}} n}{\widetilde{\gamma}^{\frac{1}{3}}\left(nh^3b^d\right)^{\frac{1}{6}}}\Bigg) \leq C_2\cdot\widetilde{\gamma},$$

where $C_1' = C_1\cdot\widetilde{A}^{\frac{1}{3}} > 0$ is again a constant. Now, if $h \asymp b \asymp n^{-\frac{1}{\gamma}}$ and $\hbar \asymp n^{-\frac{1}{\varpi}}$ for some $\gamma \geq \varpi > 0$ such that $\frac{nh^{d+5}}{\log n} \to c_1$ and $\frac{n\hbar^5}{\log n} \to c_2$ for some finite number $c_1, c_2 \geq 0$ and $\frac{n\hbar^2}{\log n}, \frac{\hbar}{h^{d+3}\log n}, \hbar n^{\frac{1}{4}}, \frac{\hbar^2}{h^{d+3}} \to \infty$ as $n\to\infty$, then we know from Lemma 12 that

$$\max\left\{\sqrt{nh^{d+7}}, \sqrt{\frac{\log n}{n\hbar^2}}, \sqrt{\frac{h^{d+3}\log n}{\hbar}}, \sqrt{\frac{h^{d+3}}{\hbar^2}}\right\} \leq \frac{C_1'\cdot\log^{\frac{2}{3}} n}{\left(nh^{d+3}\right)^{\frac{1}{6}}}$$

when n is sufficiently large. Hence, we conclude that when n is sufficiently large,

$$\mathrm{P}\left(\left|\sqrt{nh^3b^d}\sup_{t\in\mathcal{T}'}|\widehat{m}_\theta(t)-m(t)| - \sup_{f\in\mathcal{F}}|\mathbb{B}(f)|\right| > \frac{C_4\cdot\log^{\frac{2}{3}} n}{\widetilde{\gamma}^{\frac{1}{3}}\left(nh^{d+3}\right)^{\frac{1}{6}}}\right) \leq C_2\cdot\widetilde{\gamma},$$

for some large constants $C_2, C_4 > 0$. To upper bound the Kolmogorov distance between $\sqrt{nh^3b^d}\cdot\sup_{t\in\mathcal{T}'}|\widehat{m}_\theta(t)-m(t)|$ and $\sup_{f\in\mathcal{F}}|\mathbb{B}(f)|$, we leverage Lemmas 11 and 13 to obtain that

$$\begin{aligned}&\sup_{u\geq 0}\left|\mathrm{P}\left(\sqrt{nh^3b^d}\cdot\sup_{t\in\mathcal{T}'}|\widehat{m}_\theta(t)-m(t)|\leq u\right) - \mathrm{P}\left(\sup_{f\in\mathcal{F}}|\mathbb{B}(f)|\leq u\right)\right| \\ &\leq \frac{C_5\cdot\log^{\frac{5}{6}} n}{\widetilde{\gamma}^{\frac{1}{3}}\left(nh^{d+3}\right)^{\frac{1}{6}}} + C_2\cdot\widetilde{\gamma},\end{aligned}$$

where $C_5 > 0$ is some constant that depends only on $\bar{\sigma} \geq \underline{\sigma} > 0$ in Lemma 13. Here, we also utilize the fact that $\log\left(\frac{1}{r_1}\right) = \log n$ when $r_1 = \frac{C_4\cdot\log^{\frac{2}{3}} n}{\widetilde{\gamma}^{\frac{1}{3}}\left(nh^{d+3}\right)^{\frac{1}{6}}}$ and use the Dudley's entropy inequality for Gaussian processes (Corollary 2.2.8 in van der Vaart and Wellner 1996) to argue that $\mathbb{E}\left[\sup_{f\in\mathcal{F}}|\mathbb{B}(f)|\right] = \mathbb{E}\left[\sup_{t\in\mathcal{T}}|\mathbb{B}(\varphi_t)|\right] = O\left(\sqrt{\log n}\right)$. We take $\widetilde{\gamma} = O\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right)$ to optimize the right hand side of the above inequality and deduce that

$$\sup_{u\geq 0}\left|\mathrm{P}\left(\sqrt{nh^3b^d}\cdot\sup_{t\in\mathcal{T}'}|\widehat{m}_\theta(t)-m(t)|\leq u\right) - \mathrm{P}\left(\sup_{f\in\mathcal{F}}|\mathbb{B}(f)|\leq u\right)\right| = O\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right).$$

The result follows. □

B.7. Proof of Theorem 7.

THEOREM 7 (Bootstrap consistency). *Let $q \geq 2$ in the local polynomial regression for estimating $\frac{\partial}{\partial t}\mu(t,\boldsymbol{s})$, $\mathcal{T}' \subset \mathcal{T}$ be a compact set so that $p_T(t)$ is uniformly bounded away from 0 within $\mathcal{T}'$, and $\mathbb{U}_n = \{(Y_i, T_i, \boldsymbol{S}_i)\}_{i=1}^n$ be the observed data. Suppose that Assumptions A1, A2, A3, A4, A5, and A6 hold. If $h \asymp b \asymp n^{-\frac{1}{\gamma}}$ and $\hbar \asymp n^{-\frac{1}{\varpi}}$ for some $\gamma, \varpi > 0$ such that $\frac{nh^{d+5}}{\log n} \to c_1$ and $\frac{n\hbar^5}{\log n} \to c_2$ for some finite number $c_1, c_2 \geq 0$ and $\frac{n\hbar^2}{\log n}, \frac{\hbar}{h^{d+3}\log n}, \hbar n^{\frac{1}{4}}, \frac{\hbar^2}{h^{d+3}} \to$*

∞ as $n \to \infty$, then

$$\sup_{u\geq 0}\left|\mathrm{P}\left(\sqrt{nh^3b^d}\cdot\sup_{t\in\mathcal{T}'}\left|\widehat{m}_\theta^*(t)-\widehat{m}_\theta(t)\right|\leq u\Big|\mathbb{U}_n\right)-\mathrm{P}\left(\sup_{f\in\mathcal{F}}|\mathbb{B}(f)|\leq u\right)\right|=O_P\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right)$$

$$\sup_{u\geq 0}\left|\mathrm{P}\left(\sqrt{nh^3b^d}\cdot\sup_{t\in\mathcal{T}'}\left|\widehat{\theta}_C^*(t)-\widehat{\theta}_C(t)\right|\leq u\Big|\mathbb{U}_n\right)-\mathrm{P}\left(\sup_{g\in\mathcal{F}_\theta}|\bar{\mathbb{B}}(g)|\leq u\right)\right|=O_P\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right),$$

where $\widehat{m}_\theta^*(t)$ and $\widehat{\theta}_C^*(t)$ are the integral estimator (8) and localized derivative estimator (12) based on a bootstrap sample $\mathbb{U}_n^*=\{(Y_i^*,T_i^*,\boldsymbol{S}_i^*)\}_{i=1}^n$ respectively, and $\mathbb{B},\bar{\mathbb{B}}$ are the same Gaussian processes as in Theorem 6.

PROOF OF THEOREM 7. We only prove the bootstrap consistency for $\widehat{m}_\theta(t)$, since the result for its derivative estimator $\widehat{\theta}_C(t)$ follows from an identical argument. Our proof here is similar to the proof of Theorem 7 in Chen et al. (2015) and the proof of Theorem 4 in Chen et al. (2017). The key difference is that the functional space remains unchanged as $\mathcal{F}$ in our scenario here for both the original Gaussian approximation (Theorem 6) and the bootstrapped version, because the index set $\mathcal{T}'$ of $\mathcal{F}$ is fixed.

Let $\mathbb{U}_n=\{(Y_i,T_i,\boldsymbol{S}_i)\}_{i=1}^n$ be the observed data and $\mathbb{U}_n^*=\{(Y_i^*,T_i^*,\boldsymbol{S}_i^*)\}_{i=1}^n$ be the bootstrap sample. We also denote $\mathbb{G}_n^*(\mathbb{U}_n)=\sqrt{n}\left(\mathbb{P}_n^*-\mathbb{P}_n\right)$, where $\mathbb{P}_n^*$ is the empirical measure defined by the bootstrap sample $\mathbb{U}_n^*$. Assume that $\mathbb{U}_n\subset\mathcal{Y}\times\mathcal{T}\times\mathcal{S}$ is fixed for a moment. Then, we can apply our arguments in Lemma 5 by replacing the probability measure P by $\mathbb{P}_n$ and obtain that

$$\begin{aligned}&\left|\sqrt{nh^3b^d}\sup_{t\in\mathcal{T}'}\left|\widehat{m}_\theta^*(t)-\widehat{m}_\theta(t)\right|-\sup_{t\in\mathcal{T}'}\left|\mathbb{G}_n^*(\mathbb{U}_n)\varphi_t\right|\right|\\&=O_P\left(\sqrt{nh^{d+7}}+\sqrt{\frac{\log n}{n\hbar^2}}+\sqrt{\frac{h^{d+3}\log n}{\hbar}}+\sqrt{\frac{h^{d+3}}{\hbar^2}}\right).\end{aligned}$$

Following the same argument in Theorem 6, we obtain that

$$\sup_{u\geq 0}\left|\mathrm{P}\left(\sqrt{nh^3b^d}\sup_{t\in\mathcal{T}'}\left|\widehat{m}_\theta^*(t)-\widehat{m}_\theta(t)\right|\leq u\Big|\mathbb{U}_n\right)-\mathrm{P}\left(\sup_{f\in\mathcal{F}}|\mathbb{B}_n(f)|\leq u\Big|\mathbb{U}_n\right)\right|=O_P\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right), \tag{48}$$

where $\mathbb{B}_n$ is a Gaussian process on $\mathcal{F}$ such that for any $f_1,f_2\in\mathcal{F}$, it has

$$\mathbb{E}\left[\mathbb{B}_n(f_1)|\mathbb{U}_n\right]=0\quad\text{and}\quad\mathrm{Cov}\left[\mathbb{B}_n(f_1),\mathbb{B}_n(f_2)\big|\mathbb{U}_n\right]=\frac{1}{n}\sum_{i=1}^n f_1(Y_i,T_i,\boldsymbol{S}_i)\cdot f_2(Y_i,T_i,\boldsymbol{S}_i).$$

Notice that the difference between $\sup_{f\in\mathcal{F}}|\mathbb{B}_n(f)|$ and $\sup_{f\in\mathcal{F}}|\mathbb{B}(f)|$ is small, because these two Gaussian processes differ in their covariance and

$$\begin{aligned}\mathrm{Cov}\left[\mathbb{B}_n(f_1),\mathbb{B}_n(f_2)\big|\mathbb{U}_n\right]&=\frac{1}{n}\sum_{i=1}^n f_1(Y_i,T_i,\boldsymbol{S}_i)\cdot f_2(Y_i,T_i,\boldsymbol{S}_i)\\&\to\mathrm{Cov}(\mathbb{B}(f_1),\mathbb{B}(f_2))=\mathbb{E}\left[f_1(Y,T,\boldsymbol{S})\cdot f_2(Y,T,\boldsymbol{S})\right]\end{aligned}$$

as $n \to \infty$. More precisely, by Corollary 9 in Giessing (2023), we know that

$$\begin{aligned}
&\sup_{u\geq 0}\left|\mathrm{P}\left(\sup_{f\in\mathcal{F}}|\mathbb{B}_n(f)|\leq u\middle|\mathbb{U}_n\right)-\mathrm{P}\left(\sup_{f\in\mathcal{F}}|\mathbb{B}(f)|\leq u\right)\right| \\
&\leq C_6\left[\frac{\sup_{f_1,f_2\in\mathcal{F}}\left|\mathrm{Cov}\left[\mathbb{B}_n(f_1),\mathbb{B}_n(f_2)\middle|\mathbb{U}_n\right]-\mathrm{Cov}\left[\mathbb{B}(f_1),\mathbb{B}(f_2)\right]\right|}{\max\left\{\mathrm{Var}\left(\sup_{f\in\mathcal{F}}|\mathbb{B}(f)|\right),\mathrm{Var}\left(\sup_{f\in\mathcal{F}}|\mathbb{B}_n(f)|\right)\right\}}\right]^{\frac{1}{3}} \\
&\leq C_6'\left[\sup_{f_1,f_2\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^n f_1(Y_i,T_i,\boldsymbol{S}_i)\cdot f_2(Y_i,T_i,\boldsymbol{S}_i)-\mathbb{E}\left[f_1(Y,T,\boldsymbol{S})\cdot f_2(Y,T,\boldsymbol{S})\right]\right|\right]^{\frac{1}{3}} \qquad (49)\\
&= C_6'\left[\sup_{f\in\mathcal{F}^2}\left|\frac{1}{n}\sum_{i=1}^n f(Y_i,T_i,\boldsymbol{S}_i)-\mathbb{E}\left[f(Y,T,\boldsymbol{S})\right]\right|\right]^{\frac{1}{3}}
\end{aligned}$$

where the last inequality follows from Lemma 13 and $C_6, C_6' > 0$ are two absolute constants. Here, $\mathcal{F}^2 \equiv \{f_1 \cdot f_2 : f_1, f_2 \in \mathcal{F}\}$. Now, by symmetrization (Lemma 2.3.1 in van der Vaart and Wellner 1996) and maximal inequality (Corollary 2.2.8 in van der Vaart and Wellner 1996), we also know that

$$\begin{aligned}
&\sup_{f\in\mathcal{F}^2}\left|\frac{1}{n}\sum_{i=1}^n f(Y_i,T_i,\boldsymbol{S}_i)-\mathbb{E}\left[f(Y,T,\boldsymbol{S})\right]\right| \\
&\leq 2\mathbb{E}\left[\sup_{f\in\mathcal{F}^2}\left|\frac{1}{n}\sum_{i=1}^n \chi_i f(Y_i,T_i,\boldsymbol{S}_i)\right|\right] \\
&\leq \frac{C_7}{\sqrt{n}}\int_0^\infty \sup_Q\sqrt{\log N\left(\mathcal{F}^2, L_2(Q),\epsilon\right)}\,d\epsilon \\
&\leq \frac{C_7'}{\sqrt{n}}\int_0^\infty \sup_Q\sqrt{\log N\left(\mathcal{F}, L_2(Q),\epsilon\right)}\,d\epsilon \\
&= \frac{C_7'}{\sqrt{n}}\int_0^{\sqrt{h^3b^d}\cdot M_{\mathcal{F}}} \sup_Q\sqrt{\log N\left(\widetilde{\mathcal{F}}, L_2(Q),\sqrt{h^3b^d}\cdot\epsilon\right)}\,d\epsilon \\
&= \frac{C_7'}{\sqrt{nh^3b^d}}\int_0^{M_{\mathcal{F}}} \sup_Q\sqrt{\log N\left(\widetilde{\mathcal{F}}, L_2(Q),u\right)}\,du \\
&\leq \frac{C_8}{\sqrt{nh^3b^d}},
\end{aligned}$$

where $\chi_1,...,\chi_n$ are Rademacher random variables, $\widetilde{\mathcal{F}} = \sqrt{h^3b^d}\cdot\mathcal{F} = \left\{\sqrt{h^3b^d}f : f \in \mathcal{F}\right\}$ by Lemma 9, and $C_7, C_7', C_8 > 0$ are absolute constants. In addition, $M_{\mathcal{F}} = \sup_{f\in\mathcal{F}}\mathbb{E}\left[f(Y_1,T_1,\boldsymbol{S}_1)^2\right] < \infty$ by Lemma 13 and the last inequality follows from the VC-type property of $\widetilde{\mathcal{F}}$ by Lemma 9. Combining the above result with (49), we obtain that

$$\sup_{u\geq 0}\left|\mathrm{P}\left(\sup_{f\in\mathcal{F}}|\mathbb{B}_n(f)|\leq u\middle|\mathbb{U}_n\right)-\mathrm{P}\left(\sup_{f\in\mathcal{F}}|\mathbb{B}(f)|\leq u\right)\right| = O_P\left(\frac{1}{(nh^3b^d)^{\frac{1}{6}}}\right).$$

Together with (48), the result thus follows. □

B.8. Proof of Corollary 8.

COROLLARY 8 (Uniform confidence band). *Under the setup of Theorem 7, we have that*

$$\mathrm{P}\left(\theta(t) \in \left[\widehat{\theta}_C(t) - \bar{\xi}^*_{1-\alpha}, \widehat{\theta}_C(t) + \bar{\xi}^*_{1-\alpha}\right] \textit{ for all } t \in \mathcal{T}'\right) = 1 - \alpha + O\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right),$$

$$\mathrm{P}\left(m(t) \in \left[\widehat{m}_\theta(t) - \xi^*_{1-\alpha}, \widehat{m}_\theta(t) + \xi^*_{1-\alpha}\right] \textit{ for all } t \in \mathcal{T}'\right) = 1 - \alpha + O\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right).$$

PROOF OF COROLLARY 8. By Theorem 6 and Theorem 7, we have the Berry-Esseen bounds related to bootstrap estimates for the distributions of $\sqrt{nh^3b^d} \cdot \sup_{t\in\mathcal{T}'} |\widehat{m}_\theta(t) - m(t)|$ and $\sqrt{nh^3b^d} \cdot \sup_{t\in\mathcal{T}'} \left|\widehat{\theta}_C(t) - \theta_C(t)\right|$ as:

$$\sup_{u\geq 0}\left|\mathrm{P}\left(\sqrt{nh^3b^d} \sup_{t\in\mathcal{T}'} |\widehat{m}^*_\theta(t) - \widehat{m}_\theta(t)| \leq u \middle| \mathbb{U}_n\right) - \mathrm{P}\left(\sqrt{nh^3b^d} \cdot \sup_{t\in\mathcal{T}'} |\widehat{m}_\theta(t) - m(t)| \leq u\right)\right|$$

$$= O_P\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right),$$

and

$$\sup_{u\geq 0}\left|\mathrm{P}\left(\sqrt{nh^3b^d} \cdot \sup_{t\in\mathcal{T}'} \left|\widehat{\theta}^*_C(t) - \widehat{\theta}_C(t)\right| \leq u\right) - \mathrm{P}\left(\sqrt{nh^3b^d} \cdot \sup_{t\in\mathcal{T}'} \left|\widehat{\theta}_C(t) - \theta_C(t)\right| \leq u\right)\right|$$

$$= O_P\left(\left(\frac{\log^5 n}{nh^{d+3}}\right)^{\frac{1}{8}}\right).$$

The results thus follow. $\square$