Sensitivity Analysis for Treatment Effects with Endogenously Censored Duration Outcome*

Qi Xu†

November 12, 2022

Click here for the latest version

Abstract

With a non-randomly censored duration outcome, we perform sensitivity analyses on various treatment effect parameters when the dependence between the event time and the censoring variable is modeled by a family of Archimedean copula. Bounds of policy effects are characterized as smooth functionals of the copula graphic estimands that satisfy an index sufficiency condition. We then provide an estimation procedure and establish uniform inference theories for the proposed semiparametric estimators. Confidence bands are constructed using multiplier bootstrap. The estimators demonstrate good finite sample properties in Monte Carlo simulations. The methodology is applied to study the efficacy of treatment protocols for acute lymphoblastic leukaemia.

Keywords: Observational studies, dependent censoring, copula graphic estimator, single index model.

JEL Classification: C14, C21, C34, C41

*I am grateful for the guidance of my advisors, Atsushi Inoue and Tong Li, as well as for the insights and helpful discussions from the members of my committee, Atsushi Inoue, Tong Li, Tatsushi Oka and Pedro Sant'Anna.

†Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819, USA. Email: qi.xu.1@vanderbilt.edu.

1 Introduction

Many program evaluation problems involve censored outcomes. A few classical examples include, survival time of patients in clinical trials, duration of unemployment, length of marriage, the lifetimes of firms, and so forth. In addition to the usual problems with counterfactual analyses, censoring poses additional challenges to researchers, as it is well known that the marginal and joint distributions of the latent outcome and the censoring variable are not identifiable if the censoring mechanism is left entirely unrestricted (Tsiatis, 1975). One popular approach to restoring identification is by assuming that the two variables are independent, possibly conditional on observed covariates. Although prevalent, such an assumption can easily be violated in many practically relevant applications. Subject attribution due to unobserved factors being correlated with the latent outcome, and the presence of competing events are among the most frequently encountered reasons for the failure of the proposed assumption. As a concrete example, consider clinical patients who receive poor prognosis. They may decide to withdraw from trials based on such result, causing a positive correlation between the survival and abandonment times. Ignoring such dependence would lead to biased assessments of the treatment.

Methodologies under dependent censoring have received less attention relative to their independent counterpart, partly due to a lack of consensus on how the dependence should be modeled. The current literature is divided between imposing a known censoring mechanism and making no assumptions about it at all. On the one extreme, assuming that the dependence structure is fully characterized by a known copula, distributional information of the latent duration can be recovered from observed durations (see Zheng and Klein (1995)), but such results are sensitive to the specification of the true copula. On the other extreme, robust approaches such as the one proposed by Khan and Tamer (2009), utilize minimal theoretical restrictions and typically generate uninformative identified sets.

In reality, researchers often possess some prior information on the censoring mechanism. Such information may come from auxiliary data, scientific theory, or expert opinion. For instance, in the clinical trial example, we may assume the latent and censoring times are positively correlated based on prior research findings. It is crucial that identification and inference procedures built by researchers allow one to flexibly incorporate partial information as such when addressing policy relevant questions.

With this goal in mind, we follow the partial identification approach proposed by Fan and Liu (2018), and assume that the true copula of latent outcome and censoring time belongs to the well-known Archimedean family. We do not, however, directly specify the true copula. The Archimedean copulas serve two purposes here. First, it allows the distribution of potential outcome be explicitly expressed in the form of copula-graphic-type estimands (see Rivest and Wells (2001)) that we denominate as *bound generating functions* (BGF). Such functions are smooth functionals of the observed (sub-) distributions and are indexed by the level of dependence censoring. Second, many one-parameter Archimedean families are endowed with a concordance ordering (Nelsen, 2007). The BGFs inherit such a property, and as a result, are ordered in terms of the *first-order stochastic dominance* (FOSD) relations. This natural ordering then allow us to explicitly derive the bounds of identified sets for various treatment effects. To the best of our knowledge, this copula-based partial identification approach to program evaluation is new to the literature.

Analytical forms of the bounds render sensitivity analysis with respect to the level of dependent censoring especially convenient. With endogenously censored data, such analyses are crucial for obtaining convincing policy assessments, because the assumptions on the censoring mechanism are intrinsically untestable.

As a second contribution of the paper, we incorporate the single-index structure into the aforementioned copula-based approach, based on which, we propose estimation procedures for the BGFs, as well as the bounds of the treatment effects, using a novel single-index-copula-graphic (SICG) estimator. The dimension-reduction feature of this new estimator is particularly attractive in the current context, where fully nonparametric methods such as those adopted by Braekers and Veraverbeke (2005), Lopez (2011), and Fan and Liu (2018), tend to be plagued by the "curse of dimensionality", due to the multitude of baseline covariates needed for justifying the unconfoundedness setting. We provide comprehensive large sample results for the proposed estimators, including a uniform linear expansion for the new SICG estimator, based on which, we establish functional central limit theorems for the BGFs as well as the bounds of the treatment effects. To conduct uniformly valid inference, we propose easy-to-implement multiplier bootstrap procedures, and show the bootstrap uniform confidence sets are asymptotically accurate.

We illustrate the proposed methodology by Monte Carlo studies and an empirical application where we compare the relative efficacy of two treatment protocols (GHS-2000 and AHOPCA ALL-2008) for acute lymphoblastic leukaemia (ALL). Using data from a series of clinical studies conducted in Honduras, prior work by Bernasconi, Antolini *et al.* (2022) found that the more recent treatment plan leads to better survival prospects for patients in the first three years post treatment. Their results depend crucially on the potential survival time and abandonment-of-treatment being independent conditionally on observed covariates. When we depart from this assumption, however, this conclusion may not continue to hold according to our sensitivity analysis.

Related literature: This article contribute to an extensive literature on program evaluation with censored data. The majority of works in this literature relies on the random censoring assumption. See, e.g. Anstrom and Tsiatis (2001), Hubbard, Laan, and Robins (2000), Lee and Lee (2005), Frandsen (2015), Sant'Anna (2016), Sant'Anna (2021), and so on. Models that accommodate dependent censoring is gaining attention. For instance, Beyhum, Florens, and Van Keilegom

(2021), and Crommen, Beyhum, and Van Keilegom (2022) both study inference problems with endogenous treatment models. Our paper differs from these two in that we do not impose strong functional assumptions on the *data generating process* (DGP) and seek point identification.

This article is also related to the literature on dependent censoring and competing risk models. Early contribution by Tsiatis (1975) shows that the joint distribution of the competing risks is not identified, and the best obtainable bounds are the worst-case bounds derived by Peterson (1976). These results allude to the difficulty of accounting for endogenous censoring without extra constraints or external information. To model the dependence between the potential outcome and the censoring variable, we follow the copula-based approach. With a fully known copula, Zheng and Klein (1995) propose a nonparametric estimator, which extends the one by Kaplan and Meier (1958), and call it the copula-graphic estimator. Rivest and Wells (2001) show that the estimator has a closed-form expression when attention is restricted to the Archimedean copulas. Braekers and Veraverbeke (2005), Huang and Zhang (2008), and Chen (2010) further extend it by incorporating covariates. The known copula assumption is imposed in all of the above works. In a linear quantile regression setting, Fan and Liu (2018) propose a partial identification approach that allows copula to vary within a prespecified class. This paper extends their approach to the program evaluation framework. More recently, Czado and Van Keilegom (2021), Deresa and Van Keilegom (2020), and Deresa, Van Keilegom, and Antonio (2022) also allow an unknown copula, but they establish its identifiability via strong distributional restrictions.

We also build on the literature of single-index estimation with censored data. As a powerful dimension-reduction device, single-index models are widely popular in semiparametric duration analysis, cf. Lopez (2011), Lopez, Patilea, and Van Keilegom (2013), and Bouaziz and Lopez (2010). The available results all rely on the random censoring assumption, and are not directly applicable to the copula-based setting. The novel SICG estimator proposed in this paper fills this gap. It is worthwhile to mention that it is not restricted to the scope of this article and can be used in many other settings. Additionally, since we follow Li and Patilea (2018) and impose the single-index structure directly on the potential laws, rather than on the observed distributions, our estimation procedure for the index parameter is greatly simplified relative to the aforementioned papers.

Organization of the article: Section 2 introduces the endogenous censoring framework and the treatment effect parameters. Section 3 presents the single-index model, and in addition, introduce Archimedean copulas, and the bound generating functions. We also provide identification results on the aforementioned quantities in this section. Next, in Section 4, we propose a multi-step estimation procedure for various treatment effects, using the identification results derived in Section 3. We also establish large sample theories for the proposed estimators in this section. Section 5 establishes the validity of multiplier bootstrap procedures, and provide practical guidelines for constructing uniform bootstrap confidence bands. In Section 6, we illustrate the finite sample performance of our proposed estimators and the bootstrap confidence sets, via Monte Carlo simulations. Section 7 presents an empirical application, and Section 8 concludes. Proofs and auxiliary results are collected in the Supplementary Appendix.

2 Setup and Parameter of Interest

2.1 Model Framework

Consider a program where the outcome of interest is measured by the amount of time until a target event occurs. We let $T \in \mathcal{T} \subset [0, \infty)$ denote such an outcome. We also observe an indicator D for binary treatment: $D = 1$ if the unit is treated and $D = 0$ otherwise. Following Neyman-Rubin potential outcome framework (see e.g. Rubin (1974)), we denote by T_1 and T_0 the values that T would have taken if D is equal to one or zero, respectively. As a result, $T = DT_1 + (1 - D)T_0$. A vector $X \in \mathcal{X} \subset \mathbb{R}^k$ of baseline covariates is recorded prior to the program. In an ideal setting, we would observe (T_1, T_0, D, X), and make inference thereof. However, the ideal data is coarsened in two ways.

First, we only observe the realized event time T but not the potential outcomes T_1 and T_0. The common approach is to assume that all the potential confounders are captured by the covariates, and therefore, we may recover Moreover, the realized T is subject to right censoring, by a random variable, $C \in \mathcal{Y} \subset \mathbb{R}_+$. As a result, one only have access to $Y = \min\{T, C\}$ and a no-censoring indicator R where $R = 1$ if $T \leqslant C$, and $R = 0$, otherwise. Same as the outcome of interest, C, Y and R are also functions of D, i.e. $U = DU_1 + (1 - D)U_0$, where $U \in \{C, Y, R\}$ and U_d stands for the potential realization of U under treatment d. Thus, the available data W consists of (Y, R, D, X). We let $S_{U|V}(\cdot|v) = \mathbb{P}(U > \cdot|V = v)$ denote the *survival function* of a random variable U given $V = v$. Our goal is to make inference on functionals of $S_{T_d|X}$, using information from observed samples of W.

In observational studies, treatment is not randomly assigned. Therefore, the treatment, the event time and the censoring variable are all likely confounded. To manage the relationship between the treatment and latent duration outcomes, we focus on the unconfoundedness setup. That is, we impose the following assumption on the underlying data generating process,

Assumption 1 $(T_1, T_0, C_1, C_0) \perp\!\!\!\perp D|X$.

Assumption 1 implies that, selection into treatment is solely based on observable characteristics. The assumption is akin to the standard unconfoundedness condition in program evaluation literature, not only with complete observations, cf. Rosenbaum and Rubin (1983), Hirano, Imbens, and Ridder

(2003), and Firpo (2007), but also with censored outcomes, cf. Lee and Lee (2005), and Sant'Anna (2016, 2021). It differs from the latter two, however, by requiring independence of the joint law, rather than on the potential event time only. The strengthened condition is necessary since the event and censoring times may remain dependent even after adjusting for the covariates.

Since the observed duration Y and the censoring indicator R are deterministic functions of T and C, the above assumption immediately implies that $(Y_1, Y_0, R_1, R_0) \perp\!\!\!\perp D|X$. We also note that since the experimental setting can be viewed as a special case of Assumption 1, all of our theories presented below will automatically carry over to the randomized-controlled-trial setting.

2.2 Parameters of Interest

We will work mainly with the following four types of treatment effects under the uncounfoundedness setup:

$$\begin{aligned}
&\text{Average Treatment Effect: } ATE \equiv \mathbb{E}[T_1 - T_0],\\
&\text{Distributional Treatment Effect: } DTE(t) \equiv F_{T_1}(t) - F_{T_0}(t),\\
&\text{Quantile Treatment Effect: } QTE(\tau) \equiv F_{T_1}^{-1}(\tau) - F_{T_0}^{-1}(\tau),\\
&\text{Cumulative Hazard Treatment Effect: } CHTE(t) \equiv \Lambda_{T_1}(t) - \Lambda_{T_0}(t),
\end{aligned}$$

where the quantile function $F_{T_d}^{-1}(\cdot)$ and the cumulative hazard function $\Lambda_{T_d}(\cdot)$ associated with treatment type $d \in \{0, 1\}$ are defined by $F_{T_d}^{-1}(\tau) \equiv \inf\{y : F_{T_{d(y)}} \geqslant \tau\}$ and $\Lambda_{T_d} : F_{T_d} \mapsto \int_{[0,\cdot]} \frac{1}{1-F_{T_d}^{-}} dF_{T_d}$, with $F^{-}(x) \equiv \lim_{s\uparrow t} F(s)$, respectively. Note that each of these policy effects can be represented as the difference between smooth functionals of F_{T_d}. These functionals, denoted by $F_{T_d} \mapsto \Upsilon(F_{T_d}(\cdot))(\cdot)$, are usually called the *treatment responses*. It can be shown that each of the treatment responses introduced here respects the FOSD relations of F_{T_d}. That is, either $\Upsilon(F(\cdot))(u) \geqslant \Upsilon(G(\cdot))(u)$ or $\Upsilon(G(\cdot))(u) \geqslant \Upsilon(F(\cdot))(u)$ for all u, whenever $F(t) \geqslant G(t)$, for all t. Such a property will be exploited for characterizing identified set for the treatment effects. There are examples of treatment responses that violate the FOSD relation. For instance, the Gini coefficients and Lorenz curves respect second-order stochastic dominance relations but not the first-order one. Consequently, our identification analysis do not apply in these cases.

Under independence censoring mechanism, the policy parameters are known to be point identified from the observed data. See e.g. Hubbard *et al.* (2000), Lee and Lee (2005), and Sant'Anna (2016). However, when the censoring mechanism is entirely unrestricted, the best attainable result is the worse-case bounds by Peterson (1976). We aim to take the middle ground in this paper, and try to address the following type of question: if the level of dependence can be restricted to a given range, what values of the treatment effects are consistent with this information? The answer depends on two factors: (i) the quantification of the level of dependence censoring and (ii) a link from the censoring mechanism to the policy parameters. These two ingredients are discussed in detail in the next section.

3 Identification

We describe our identification strategy in this section. Our main result can be divided into two parts. We first introduce a single-index model, and discuss the identification of its index parameters. Then, in the second part, we provide the identification results on the distributions of potential durations and the treatment effects, all through the lens of copula theory. Despite the order of our exposition, the majority of results in the second part are relatively independent, and can be established without the embedment of the single-index structure.

3.1 Single-Index Model

Semiparametric models offer a good compromise between the parametric approach, which relies on strong assumptions on the functional form assumption that may not hold in practice, and the fully nonparametric one, which suffers from the curse of dimensionality. One famous example of such dimension reduction device is the single-index model. It has been widely adopted in duration analysis. See, for instance, Xia, Zhang, and Xu (2010), Bouaziz and Lopez (2010), Lopez *et al.* (2013), Li and Patilea (2018), and Bücher, El Ghouch, and Van Keilegom (2021). For a generic conditional distribution of Y given X, the single-index model assumes that $F_{Y|X}(y|x) = G(y, x'\gamma^{\dagger})$, where G is an unknown bivariate function, and $\gamma^{\dagger}$ is the vector of index parameters. In general, the coefficients are only identified up to a scale, and thus, normalization is required to get point identification. Towards this end, we arrange the covariates so that the first k_1 variables are absolutely continuously distributed, and the remaining k_2 variables are binary. We set the coefficient associated with the first element, $x_{[1]}$ to 1, and let $x\gamma = x_{[1]} + x'_{[-1]}\gamma$, where $x_{[-1]}$ collects all the other covariates and γ is the corresponding subvector of $\gamma^{\dagger}$. Note that this normalization is not entirely innocuous, as it imposes a positive effect on the first component of the covariates.

Assumption 2 (Single index structure) $(T_d, C_d) \perp\!\!\!\perp X | X\gamma_d$, *where* γ_d *is an interior point of a compact set* $\Gamma \subset \mathbb{R}^{k-1}$*, for* $d \in \{0, 1\}$.

Assumption 2 is a index sufficiency condition on the joint law of the event time and the censoring variable. A similar restriction appears in Li and Patilea (2018) under the random censoring mechanism. Note that the true index coefficient may vary across treatment groups, reflecting potential differences in the treatment response heterogeneity. However, the indeces are restricted to be the same across the marginal laws of T_d and C_d, for each $d \in \{0, 1\}$. An immediate consequence of the index sufficiency condition is that $X\gamma_d$ can be viewed as a balancing score, meaning that the potential outcomes are independent of the treatment choices conditional on this index. The result is formally stated in Lemma 3.1.

Lemma 3.1 *Under Assumptions 1 and 2,* $(T_d, C_d, Y_d, R_d) \perp\!\!\!\perp D|X\gamma_d$*, for* $d \in \{0,1\}$.

The lemma essentially states that the unconfoundedness property as induced by the conditioning set X is maintained under a coarse partition of X generated by the index $X\gamma_d$. This matching condition is crucial for establishing identification of $\boldsymbol{\gamma} \equiv (\gamma_1', \gamma_0')'$ from observed data.

With slight abuse of notation, we write $G_{d,r}(\cdot, x\gamma) = F_{Y,R|D,X\gamma}(\cdot, r|d, x\gamma)$, and $f_d(\cdot) \equiv \partial_{x\gamma} F_{X\gamma,D}(\cdot, d)$, for $(d,r) \in \{0,1\}^2$, where the functional form of $G_{d,r}$ and f_d depends on γ. Furthermore, we define

$$
\begin{aligned}
\mathcal{E}_{d,r,\gamma}(t) &\equiv \mathbb{1}\{D = d\}\{\mathbb{1}\{R = r, Y \leqslant t\} - G_{d,r}(t, X\gamma)\} \\
U_{d,\gamma}(t,d,r) &\equiv \mathcal{E}_{d,r,\gamma}(t) f_d(X\gamma),
\end{aligned}
$$

and let $\mathcal{E}_{d,r,\gamma,\ell}$ and $U_{d,\gamma,\ell}$ be the same functions defined with observation W_ℓ.

Exploiting the balancing property of $X\gamma_d$, we will show in Theorem 3.1 that, under the index sufficiency condition, $\mathbb{E}[U_{d,\gamma_d,\ell}(t,r)|X] = 0$ almost surely, for each $t, d,$ and r. This conditional moment restriction will serve as the basis for the identification of the index parameters. To fully exploit the informational content of such a conditional restriction, we will follow the "integrated conditional moment approach" common in the specification testing literature. See, e.g. González-Manteiga and Crujeiras (2013) for a review. The idea is to characterize the conditional moment restriction as an infinite number of unconditional moment equations via some well-chosen family of weight functions $\{\vartheta(X; z) : z \in \mathcal{Z}\}$. That is,

$$
\mathbb{E}[U_{d,\gamma_d,1}(t,r)|X] = 0 \text{ a.s.} \Leftrightarrow \mathbb{E}[U_{d,\gamma_d}(t,r)\vartheta(X;z)] = 0 \text{ a.e. in } z \in \mathcal{Z}, \tag{3.1}
$$

Lemma 1 of Escanciano (2006b) provides primitive conditions on the family of weights for the equivalence in the preceding display to hold. Here, we list a few examples that satisfy the equivalence condition: (i) $\vartheta(X;z) = \mathbb{1}\{X \leqslant z\}$, with $z \in \mathbb{R}^k$, see e.g., Stute (1997) and Domínguez and Lobato (2004); (ii) $\vartheta(X;z) = \mathbb{1}\{X'z_1 \leqslant z_2\}$, with $z = (z_1, z_2) \in \mathbb{S} \times \mathbb{R}$, where $\mathbb{S}^k$ is the $(k-1)$-dimensional unit sphere, see e.g. Escanciano (2006a); (iii) $\vartheta(X;z) = \exp(iz'X)$, with $z \in \mathbb{R}^k$ and $i = \sqrt{-1}$, see e.g., Bierens (1982) and Lavergne and Patilea (2013).

Now, we define

$$
\mathcal{J}_d(\gamma;\vartheta) \equiv \int_{\mathcal{T}\times\{0,1\}} \int_{z\in\mathcal{Z}} \|\mathbb{E}[U_{d,\gamma}(t,r)\vartheta(X;z)]\|^2 \, d\Pi_Z(z) d\Pi_{T,R}(t,r), \tag{3.2}
$$

where Π_Z is an integrating measure that is absolutely continuous with respect to the dominant measure of z. Likewise, $\Pi_{T,R}$ is an integrating measure for (t, r) that is specified by the researcher. It is not necessarily related to the unobserved law of T. In Theorem 3.1, we will show that $\gamma_d = \arg\min_{\gamma \in \Gamma} \mathcal{J}_d(\gamma; \vartheta)$, $d \in \{0, 1\}$, and the minimization yields a unique solution, if the conditions given in Assumption 3 are fulfilled.

Assumption 3 (Identification of index)

1. *(i)* $\mathcal{X} = \mathcal{X}^c \times \mathcal{X}^b \equiv \Pi_{\ell_1=1}^{k_1}[\underline{x}_{\ell_1}, \bar{x}_{\ell_1}] \times \{0,1\}^{k_2}$*; (ii)* $\inf_{x\gamma \in \mathcal{X}_\Gamma} f_d(X\gamma) > 0$*, where* $\mathcal{X}_\Gamma \equiv \{x\gamma : x \in \mathcal{X}, \gamma \in \Gamma\}$*; (ii)* $\mathcal{T} = [0, \bar{y}]$*, where* $\bar{y} = \inf\{y : \inf_{(r,d,x)\in\{0,1\}^2\times\mathcal{X}} F_{Y_d,R_d|X\gamma_d}(y, r|x\gamma_d) = 1\}$*.*

2. $\mathbb{P}(D = d|X) > 0$*, almost surely.*

3. *There exist sets* $\mathcal{T}_0 \subset \mathcal{T}$*, such that for each* $t \in \mathcal{T}_0$*, (i) the function* $z \mapsto F_{Y_d,R_d|X\gamma_d}(t, r|z)$ *is differentiable in* z*; (ii) there exists a set* $\mathcal{X}_0 \subset \mathcal{X}$*, such that* $\mathbb{P}(X \in \mathcal{X}_0) > 0$*, and* $\partial_{x\gamma} F_{Y_d,R_d|X\gamma_d}(t, r|x\gamma_d) \neq 0$*, for all* $x \in \mathcal{X}_0$*.*

4. *For each* $\gamma \in \Gamma$*, there exists an open interval* $\mathcal{V}_0$ *satisfying (i)* $\mathcal{V}_0 \subset \cap_{\ell=0}^{k_2-1}\{\mathcal{X}_\gamma^c + \gamma_{k_1+\ell}\} \cap \mathcal{X}_\gamma^c$*, where* $\mathcal{X}_\gamma^c = \{x_{[1]} + \gamma_1 x_{[2]} + \ldots + \gamma_{k_1-1}x_{[k_1]} : (x_{[1]}, \ldots, x_{[k_1]}) \in \mathcal{X}^c\}$*, and (ii) for each* $t \in \mathcal{T}_0$*, if* $F_{Y_d,R_d|X\gamma_d}(t, r|v + u) = F_{Y_d,R_d|X\gamma_d}(t, r|v)$ *for all* $v \in \mathcal{V}_0$*, then* $u = 0$*.*

Assumptions 3.1 and 3.2 are standard. We allow for continuous covariates as well as discrete ones. Here, the discrete variables are all assumed to be binary, but the restriction can be easily relaxed. Note that Assumption 2.1(ii) implies that $\bar{t} \leqslant \bar{c}$, where $\bar{t}$ and $\bar{c}$ are the upper bounds in the support of the event time and the censoring variable, respectively. Outcome beyond $\bar{y}$ will never be observed, thus the entire distribution $F_{T_d|X\gamma_d}$, and thus the ATE, can be identified only if $\bar{t} \leqslant \bar{c}$. When the interest lies in functionals that do not involve the entire distribution, this assumption is not needed. Assumption 3.2 is the usual overlapping condition on the treatment assignment mechanism, imposed to guarantee that the conditional distribution $G_{d,r}$ are well defined on $\mathcal{X}_\Gamma$. The next two conditions are adapted from Assumptions 4.1 and 4.2 in Ichimura (1993). They are imposed to ensure the identifiability of the index parameter. Together with the normalization restriction, Assumption 3.3 secures identification of index coefficients corresponding to the continuous covariates. Assumption 3.4 restricts the shape of $\mathcal{X}_\Gamma$, and when it is assumed in addition, the coefficients for binary covariates are point identified as well.

Theorem 3.1 *Under Assumptions 1, 2, 3.1, and 3.2 it holds that (i)*

$$\mathbb{E}[U_{d,\gamma_d}(t, r)|X] = 0, \textit{ almost surely, } \forall (d, r, t) \in \{0, 1\}^2 \times \mathcal{T}. \tag{3.3}$$

(ii) If in addition, Assumptions 3.3 and 3.4 hold, $\gamma \neq \gamma_d$ *implies* $\mathbb{E}[U_{d,\gamma}(t,r)|X] \neq 0$*, almost surely, for all* $(d,r,t) \in \{0,1\}^2 \times \mathcal{T}_0$*. (iii) If in addition,* ϑ *belongs to any of the classes of functions in Lemma 1 of Escanciano (2006b), and* $\int_{\mathcal{T}_0 \times \{0,1\}} d\Pi_{T,R}(t,r) > 0$*, we have that, for* $d \in \{0,1\}$*,* $\mathcal{J}_d(\gamma;\vartheta) \geqslant 0$*,* $\forall \gamma \in \Gamma$*, and the equality holds if and only if* $\gamma = \gamma_d$*.*

Theorem 3 is a global identification result. It shows that the index parameters can be recovered as the unique minimizer of the minimum distance type criterion, (3.2). Compared with similar approaches by Bouaziz and Lopez (2010), Strzalkowska-Kominiak and Cao (2014), and Li and Patilea (2018), we do not directly impose the uniqueness of single-index structure, but rather derive it from primitive and mild conditions on the underlying DGP.

As an implication of Theorem 3, we show how $\boldsymbol{\gamma}$ can be estimated based on a reformulation of $\mathcal{J}_d(\gamma;\vartheta)$. Towards this end, we note that, by means of the law of iterated expectations, $\mathcal{J}_d(\gamma,\vartheta)$ can be written as

$$\mathcal{J}_d(\gamma;\rho) = \int_{\mathcal{T} \times \{0,1\}} \mathbb{E}\left[\rho(X_1,X_2) U_{d,\gamma,1}(t,r) U_{d,\gamma,2}(t,r)\right] d\Pi_{T,R}(t,r), \tag{3.4}$$

where $\rho(x_1,x_2) \equiv \int_{z \in \mathcal{Z}} \vartheta(x_1,z)^c \vartheta(x_2,z) d\Pi_Z(z)$, and A^c is the conjugate transpose of A. The function ρ might appear complicated at first, but a convenient closed-form usually follows once an appropriate weight function and integrating measure combination is chosen. For instance, when $\vartheta(X;z) = \exp(iz'X)$ and $\Pi_Z(z) = \Phi(z)$, where $\Phi(\cdot)$ is the CDF of k-variate standard normal distribution, $\rho(x_1,x_2) = \exp(-\|x_1 - x_2\|/2)$.[1] Other examples can be found in Escanciano (2006b) and Sant'Anna, Song, and Xu (2022). The new criterion (3.4) follows the minimum distance function of Li and Patilea (2018) closely, inheriting many attractive properties of theirs. For one, since the dimension reduction hypothesis is imposed on joint laws of T_d and C_d, (3.2) does not involve complicated conditional Kaplan-Meier type integrals as is common in the literature. See, e.g. Xia *et al.* (2010), Bouaziz and Lopez (2010), and Strzalkowska-Kominiak and Cao (2014). Moreover, no trimming is required, due to the inclusion of f_d in $U_{d,\gamma}$. As such, we may avoid dealing with convoluted multi-step procedures as appeared in Delecroix, Hristache, and Patilea (2006), and Bouaziz and Lopez (2010).

We propose to estimate the index parameters $\boldsymbol{\gamma}$ by minimizing the sample analogue of (3.4). The integration with respect to $\Pi_{T,R}$ may also be avoided when the it is replaced by a suitable empirical measure. Details of the estimation procedure are provided in Section 4.

1 Here we have used the fact that

$$\int_{\mathbb{R}^k} \exp(iu't) \cdot \frac{\exp(-u'u/2)}{(2\pi)^{k/2}} du = \mathbb{E}_U[\exp(iU't)] = \exp(-t't/2),$$

where the first equality follows by the definition of characteristic function for random variable U, and the second is due to the assumption that U follows the k-variate standard normal distribution

Remark 1 *The index sufficiency condition can be tested based on (3.4), following the approach proposed by Maistre and Patilea (2019). Let $Q(\gamma)$ be a $(k-1)\times(k-1)$-invertible matrix with the first column given by γ. Consider the following function*

$$\mathcal{J}_d(\gamma,\rho,g) \equiv \int_{\tilde{\mathcal{T}}\times\{0,1\}} \mathbb{E}\left[\rho(X_1'Q(\gamma), X_2'Q(\gamma))U_{d,\gamma,1}(t,r)U_{d,\gamma,2}(t,r)J_g(X_2\gamma, X_1\gamma)\right] d\Pi_{T,R}(t,r),$$

where $J(\cdot)$ is a symmetric kernel function, g is a bandwidth, and $J_g(u,v) \equiv g^{-1}J(g^{-1}(v-u))$. If Assumption 2 does not hold, the dependence on X after adjusting for $X\gamma$ would drive $\mathcal{J}_d(\gamma,g)$ away from zero, uniformly for $\gamma \in \Gamma$ and a suitably chosen bandwidth sequence. In turn, we may construct the test statistic based on sample analogues of $\mathcal{J}_d(\gamma,\rho,g)$, and use multiplier bootstrap to generate the critical values.

Remark 2 *We would like to emphasize here that point identification of the index parameters does not imply that of the marginal distribution of T_d, as well as the joint distribution of T_d and C_d. For the latter, Peterson (1976)'s worse case bounds is applicable here and they are equivalent to the fully nonparametric case, provided that Assumption 2 indeed holds.*

3.2 Partial Identification through Copula

From Sklar (1959)'s theorem, we know that, conditionally on $X = x$, there exists a conditional survival copula, $\mathcal{C}_x(\cdot,\cdot) : [0,1]^2 \mapsto [0,1]$, such that

$$\mathbb{P}\left(T_d > t, C_d > c | X = x\right) = \mathcal{C}_x(S_{T_d|X}(t|x), S_{C_d|X}(c|x)),$$

for $t, c \in \mathcal{T}$. Moreover, if the conditional survival functions are absolutely continuous, then $\mathcal{C}_x$ is unique; otherwise it is only uniquely determined on the range of the survival functions. Sklar's results allow us to separate the analysis of the marginal laws and the dependence structure. As is discussed in Section 1, we restrict our attention to the one-parameter Archimedean family. Introduced by Genest and MacKay (1986a; 1986b), Archimedean copulas are widely used in economic applications for modeling a variety of dependence structures. The family is characterized by a generator function $\phi_\theta(\cdot) : [0,1] \mapsto [0,\infty)$ that is usually indexed by a parameter $\theta \in \Theta$:

$$\left\{\mathcal{C}\left(u,v;\theta\right) = \phi_\theta^{[-1]}\left(\phi_\theta\left(u\right) + \phi_\theta\left(v\right)\right) : \theta \in \Theta\right\}.$$

For each θ, ϕ_θ is a known continuous, convex, strictly decreasing function with $\phi_\theta(1) = 0$. In the above definition, $\phi_\theta^{[-1]}$ stands for the pseudo-inverse of ϕ_θ, as defined by

$$\phi_\theta^{[-1]}(s) = \begin{cases} \phi_\theta^{-1}(s), & 0 \leqslant s \leqslant \phi_\theta(0) \\ 0 & \phi_\theta(0) \leqslant s \leqslant \infty. \end{cases}$$

If $\phi_\theta(0) = \infty$, $\phi_\theta^{[-1]} = \phi_\theta^{-1}$, and the copula is said to be *strict*.

We do not seek to identify parameter θ in this article. Instead, we treat it as a sensitivity parameter that is varied to trace out a family of identified sets. Prior information on the dependence structure, such as model restrictions and expert opinions, will be translated as constraints on θ, which will serve to restrict the size of the identified set.

In place of a random censoring condition, the censoring mechanism of this paper are defined through mild restrictions on the copula function as seen below.

Assumption 4 (Copula)

1. *(i) The conditional distribution of* (T_1, T_0, C_1, C_0) *is absolutely continuous respect to the Lebesgue measure. Conditional distributions* $F_{Y_d,R_d|D,X}(y, r|d, x)$ *and* $F_{T_d|X}(y|x)$ *are differentiable with respect to* $y \in \mathcal{T}$.

2. *The true conditional survival copula of* (T_d, C_d)*,* $d \in \{0, 1\}$*, is strict and belongs to the one parameter Archimedean family indexed by* $\theta \in \Theta \equiv [\underline{\theta}, \bar{\theta}]$.

3. *The true copula is invariant to* x.

4. *Let* $\phi_\theta'(\cdot) = \partial_u \phi_\theta(u)$*. It holds that* $\phi_{\theta_1}'(\cdot)/\phi_{\theta_2}'(\cdot)$ *is strictly increasing for all* $\theta_1 < \theta_2$.

Assumptions 4.1 is a smoothness condition on the duration outcomes, requiring that the event and censoring times admit densities on the support of Y. Discretely-measured times pose an additional challenge for identification. Accommodating discrete outcome along the lines of Kim (2021) will be left for future work. Assumptions 4.2 can be justified in the context of mixed proportional hazards model with common frailty term, but it is not restricted to such models. See e.g. Joe (1997) and Nelsen (2007) for detailed expositions. Many Archimedean families include the independence copula, $\phi_\theta(u) = \log u^{-1}$, either as a special case or as a limiting one. This feature is particularly convenient for sensitivity analysis using Archimedean copulas. Assumptions 4.3 implies that $\mathcal{C}_x = \mathcal{C}$, for all x. This restriction is strong, yet necessary for generating meaningful bounds when the unconditional policy analysis is considered. It is also imposed by Fan and Liu (2018). We let $\theta_0(d)$ denote the true copula parameter under treatment d. Assumption 4.4 and Corollary 4.4.6 in Nelsen (2007) imply the family of copulas is endowed with a *concordance* ordering, meaning that $\mathcal{C}(u, v; \theta_1) \leqslant \mathcal{C}(u, v; \theta) \leqslant \mathcal{C}(u, v; \theta_2)$, for all $u, v \in [0, 1]^2$ and $\theta \in [\theta_1, \theta_2]$. As a result, θ sufficiently characterizes the level of dependence between T_d and C_d. This property plays a major part in generating analytical bounds for the treatment effects.

Generator functions satisfying Assumption 4.4 are common. A few well-known examples include: (i) *Clayton* copula: $\max\{u^{-\theta} + v^{-\theta} - 1, 0\}^{-1/\theta}$, with the generator $\phi_\theta(u) = \frac{1}{\theta}(u^{-\theta} - 1)$; (ii) *Gumbel* copula: $\exp\left(-\left[(-\log u)^\theta + (-\log v)^\theta\right]^{1/\theta}\right)$, with the generator $\phi_\theta(u) = (\log u^{-1})^\theta$;

(iii) *Gumbel-Hougaard* copula: $uv\exp(-\theta\log u\log v)$, with the generator $\phi_\theta(u) = \log(1-\theta\log u)$. For a comprehensive list, we refer readers to Table 4.1 in Nelsen (2007) and Table 1 in Fan and Liu (2018).

Provided that the true copula belongs to the Archimedean family, the distribution of T_d can be explicitly expressed, in terms of the generator function ϕ_θ, $G_{d,1}$, and $s_d \equiv 1 - G_d$, $d \in \{0,1\}$. We let the linking function be denominated by the conditional *bound generating function* (BGF), which is defined as follows

$$s_{T_d}(t, x\gamma_d, \theta) \equiv \phi_\theta^{-1}\left(-\int_0^t \phi_x'\left(s_d(y, x\gamma_d)\right) G_{d,1}(dy, x\gamma_d)\right), \tag{3.5}$$

for each (t, x, θ). The unconditional BGF, $s_{T_d}(\cdot,\cdot)$, is defined by taking expectation of $s_{T_d}(\cdot, X\gamma_d, \cdot)$ with respect to X, i.e. $s_{T_d}(t,\theta) = \mathbb{E}[s_{T_d}(t, X\gamma_d, \theta)]$.

The next theorem formally states how the conditional and marginal distributions of T_0 and T_1 can be recovered from the observed (conditional) distributions, via the BGFs.

Theorem 3.2 *For $d \in \{0,1\}$, (i) under Assumptions 1 - 3, 4.1 - 4.2, $S_{T_d|X}(\cdot|x) \in \{s_{T_d}(\cdot, x\gamma_d, \theta) : \theta \in \Theta\}$, a.e. for $x \in \mathcal{X}$. If in addition, Assumption 4.3 holds, then $S_{T_d}(\cdot) \in \{s_{T_d}(\cdot,\theta) : \theta \in \Theta\}$. The identified sets are uniformly sharp across t for each $d \in \{0,1\}$.*

(ii) Suppose Assumptions 1 - 3, and 4 hold. If in addition, for $(\theta_1,\theta_2) \in \Theta^2$ such that $\theta_1 \leqslant \theta_0 \leqslant \theta_2$, we have $s_{T_d}(t, x\gamma_d, \theta_2) \leqslant S_{T_d|X}(t|x) \leqslant s_{T_d}(t, x\gamma_d, \theta_1)$, and $s_{T_d}(t,\theta_2) \leqslant S_{T_d}(t) \leqslant s_{T_d}(t,\theta_1)$, a.e. for $(t,x) \in \mathcal{T} \times \mathcal{X}$. The identified sets are uniformly sharp across t and x for each $d \in \{0,1\}$.

As a direct implication of the proposition, when the true copula is known, or equivalently when $\underline{\theta} = \bar{\theta}$, the distribution of potential event time can be point identified from data. The result thus extends Rivest and Wells (2001) and Braekers and Veraverbeke (2005)'s findings to the program evaluation framework. Furthermore, it implies that, when the outcome is randomly censored, the potential survival distribution can be recovered with the famous Kaplan-Meier estimator as proposed by Beran (1981) and Dabrowska (1989). We remark that when the index sufficiency condition fails, conclusions of Theorem 3.2 will continue to hold with (3.5) replaced by its nonparametric counterpart.

Remark 3 *There is no way to learn about the true copula family from data. Prior work including Zheng and Klein (1995), Huang and Zhang (2008), Lo and Wilke (2010), and Fan and Liu (2018) finds that the choice of generating functions are less important than that of the level of dependence θ. Extensive numerical evidence suggests biases caused by misspecification of the copula family is negligible compared to that caused by θ. As such, the choice of copula family itself does not accord much identification power.*

Remark 4 *Due to the convexity of the generator function and by construction, the function* $t \mapsto s_{T_d}(t, x\gamma_d, \theta)$ *is monotonically decreasing and bounded between* $[0,1]$ *for each* $(d,\theta) \in \{0,1\} \times \Theta$. *These constraints may not be respected when* s_{T_d} *is replaced by its estimator. We discuss a remedy in remark 5.*

The name BGF is motivated by the fact that the ordering of such functions, as induced by the concordance ordering of the copula, yields a convenient characterization for the bounds of the treatment effects. Each type of the treatment effects introduced in Section 2 consists of treatment responses that respect the FOSD relations of S_{T_d}. As a result, bounds of BGFs can be translated to that of the treatment responses. Exploiting this insight, we derive closed-form bounds for various treatment effects in the next proposition.

Let us denote $q_{d,\theta}(\tau) \equiv \inf\{y : s_{T_d}(y,\theta) \leqslant 1-\tau\}$ as the τ-th quantile of $1 - s_{T_d}(\cdot,\theta)$, for $\tau \in (0,1)$.

Proposition 1 *Suppose that Assumptions 1 - 3, 4 hold, and that* $\theta_1 \leqslant \theta_0(d) \leqslant \theta_2$, *for* $d \in \{0,1\}$ *and* $\theta_1, \theta_2 \in \Theta$. *Then, we have*

$$\nu_{ATE}(\boldsymbol{\theta}) \in \left[\int_{\mathcal{T}} \left(s_{T_1}(y,\theta_2) - s_{T_0}(t,\theta_1)\right) dy,\ \int_{\mathcal{T}} \left(s_{T_1}(y,\theta_1) - s_{T_0}(t,\theta_2)\right) dy\right], \tag{3.6}$$

$$\nu_{DTE}(t,\boldsymbol{\theta}) \in [s_{T_0}(t,\theta_2) - s_{T_1}(t,\theta_1),\ s_{T_0}(t,\theta_1) - s_{T_1}(t,\theta_2)], \tag{3.7}$$

$$\nu_{QTE}(\tau,\boldsymbol{\theta}) \in [q_{1,\theta_2}(\tau) - q_{0,\theta_1}(\tau),\ q_{1,\theta_1}(\tau) - q_{0,\theta_2}(\tau)], \tag{3.8}$$

$$\nu_{CHTE}(t,\boldsymbol{\theta}) \in [\log(s_{T_0}(t,\theta_2)) - \log(s_{T_1}(t,\theta_1)),\ \log(s_{T_0}(t,\theta_1)) - \log(s_{T_1}(t,\theta_2))], \tag{3.9}$$

where $\boldsymbol{\theta} = (\theta_1,\theta_2)$, $t \in \mathcal{T}$ *and* $\tau \in (0,\bar{\tau})$. *The identified sets are uniformly sharp across* t *or* τ, *depending on* j.

Bounds for the conditional treatment effects for a fixed x *can be found by replacing* $s_{T_d}(t,\theta)$ *and* $q_{d,\theta}(\tau)$, *in the above displays, with* $s_{T_d}(t,x\gamma_d,\theta)$ *and* $q^x_{d,\theta}(\tau) \equiv \inf\{y : s_{T_d}(y,x\gamma_d,\theta) \leqslant 1-\tau\}$, *respectively.*

For each $j \in \{ATE, DTE, QTE, CHTE\}$, we let the lower and upper bound be denoted by $\nu_{lb,j}(u,\boldsymbol{\theta})$ and $\nu_{ub,j}(u,\boldsymbol{\theta})$, and let them by denominated by the lower and upper overall *treatment effect bound functions* (TEBF) for type j, respectively. We use the vector $\boldsymbol{\nu}_j = (\nu_{lb,j}, \nu_{ub,j})'$ to collect the bounds.[2] With slight abuse of notation, the index variable u, is allowed to vary depending on the type of treatment effect under consideration. In particular, $u = \varnothing$ if $j = ATE$, $u = \tau$ if $j = QTE$, and $u = t$ if $j \in \{DTE, CHTE\}$.

2 Analogous remarks apply to the conditional TEBFs, $\boldsymbol{\nu}^x_j = (\nu^x_{lb,j}, \nu^x_{ub,j})'$, where the definitions for $\nu^x_{lb,j}$ and $\nu^x_{ub,j}$ should be apparent.

Proposition 1 is the first contribution of the paper. It implies that, if the true copula parameter $\{\theta_0(d)\}_{d\in\{0,1\}}$ lies between θ_1 and θ_2, the bounds of treatment effects can be expressed as differences of smooth functionals of $s_{T_d}(\cdot,\theta_1)$ and $s_{T_d}(\cdot,\theta_2)$. Moreover, the lower and upper bounds are related by $\nu_{lb,j}(\cdot,\boldsymbol{\theta}) = \nu_{ub,j}(\cdot,\breve{\boldsymbol{\theta}})$, where $\breve{\boldsymbol{\theta}} = (\theta_2,\theta_1)$. The size of the identified set is determined by the strength of prior information. As the interval $[\theta_1,\theta_2]$ narrows, the identified sets become smaller. In the limit, the true copula is known, and the treatment effects can be point identified. This result generalizes those found in Lee and Lee (2005) and Sant'Anna (2016), to the case where independent censoring is no longer maintained. Again, even when the index sufficiency condition fails, results of Proposition 1 can be preserved, upon appropriate modifications to the BGFs.

4 Estimation and Large Sample Theory

The estimation of TEBFs consists of three steps. We sketch the steps here, in an informal way to illustrate the main idea, and the details are provided in subsequent subsections. In the first step, we estimate the index parameters γ by minimizing an estimator of (3.4). Next, we construct an consistent estimator for the conditional and average BGFs. For this purpose, we propose a new single-index copula graphic estimator. In the last step, estimates of TEBFs are constructed with the estimated BGFs from the preceding step.

4.1 Single-Index Parameters

As we mentioned earlier, when $\Pi_{T,R}$ is chosen to be $F_{Y,R}$, empirical analogue of (3.4) admits an analytical expression. Hence, we keep this choice fixed for the remainder of the paper. Other integrating measures that only involve the observed laws can be used instead. For instance, we may set $\Pi_{T,R} = F_{Y,R|D=d}$ or $\Pi_{T,R} = F_{Y,R=1|D=d}$, correspondingly for $U_{d,\gamma}$ and $d \in \{0,1\}$. Given a univariate kernel function $L(\cdot)$ and a bandwidth b that changes with sample size n, we define the sample analogue of U_γ and $\mathcal{J}$ as

$$\hat{U}_{d,\gamma,i}(y,r) = \frac{1}{n-1}\sum_{j=1}\{I_{d,y,r,i} - I_{d,y,r,j}\}\, L_b(X_i\gamma, X_j\gamma),$$
$$\hat{\mathcal{J}}_d(\gamma;\rho) = \frac{1}{n^2}\sum_{\ell=1}^{n}\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}\rho(X_i,X_j)\left(\hat{U}_{d,\gamma,i}(Y_\ell,R_\ell)\hat{U}_{d,\gamma,j}(Y_\ell,R_\ell)\right)\right\}, \tag{4.1}$$

where $I_{d,y,r,\ell} = D_\ell \mathbb{1}\{R_\ell = r, Y_\ell \leqslant y\}$, and $L_b(x,y) = b^{-1}L\left(b^{-1}(y-x)\right)$. Then, for a user-specified weighting function $\rho(\cdot)$, we estimate γ_d by minimizing $\hat{\mathcal{J}}_d$. That is,

$$\hat{\gamma}_d = \arg\min_{\gamma\in\Gamma} \hat{\mathcal{J}}_d(\gamma;\rho), \tag{4.2}$$

for $d \in \{0, 1\}$.

Under the regularity conditions to be specified in Section 4.3, we can show that the proposed index estimator is consistent, converges at the parametric rate, admits an asymptotic linear representation, and converges to a normal distribution. These results are established in Appendix B. Among these results, the consistency and convergence rate are useful for establishing uniform expansion and other properties of the conditional BGF estimator. Similar results for the unconditional case would further hinge on the existence of the linear representation.

4.2 BGF Estimators

Exploiting Proposition 3.2, estimators of BGFs can be constructed from estimators of the index coefficient γ and the observed distributions $\{G_{d,r}\}_{d,r\in\{0,1\}}$. For the latter, we propose to use the Nadaraya-Watson-type kernel estimator. Specifically, for any $\gamma \in \Gamma$, we let

$$\hat{G}_{d,r}(y, x\gamma) = \frac{\frac{1}{n}\sum_{i=1}^{n} I_{d,y,r,i} K_h(x\gamma, X_i\gamma)}{\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}\{D_i = d\} K_h(x\gamma, X_i\gamma)} \equiv \frac{\hat{\kappa}_{d,r,y}(x\gamma)}{\hat{f}_d(x\gamma)}, \tag{4.3}$$

$$\hat{G}_d(y, x\gamma) = \hat{G}_{d,1}(y, x\gamma) + \hat{G}_{d,0}(y, x\gamma), \tag{4.4}$$

where $K(\cdot)$ is a univariate kernel function, potentially different from $L(\cdot)$, and h is a bandwidth parameter.[3] The observed survival function estimator is given by $\hat{s}_d = 1 - \hat{G}_d$, . Replacing s_d and $G_{d,1}$ in (3.5) with these estimators, we get

$$\hat{s}_{T_d}(t, x\gamma, \theta) \equiv \phi_\theta^{-1}\left\{-\frac{1}{\hat{f}_d(x\gamma)n}\sum_{i=1}^{n} \phi_\theta'(\hat{s}_d(Y_i, x\gamma)) I_{d,t,r,i} K_h(x\gamma, X_i\gamma)\right\}, \tag{4.5}$$

and $\hat{s}_{T_d}(t, \theta) = n^{-1}\sum_{i=1}^{n} \hat{s}_{T,d}(t, X_i\hat{\gamma}_d, \theta)$, for all t, d, x, and θ. We note that (4.5) is an extension of the plug-in estimator of Fan and Liu (2018) to single-index models. Although their estimator can also accommodate multivariate (continuous) covariates, it is, however, not generally recommended due to the curse of dimensionality. We call $\hat{s}_{T_d}$ the *single index copula graphic* (SICG) estimator,

3 It is well known that kernel estimators exhibit large bias around the boundary points. In practice, we may modify estimators of $\hat{\kappa}_{d,r,y}$ and $\hat{f}_d$ as follows to avoid the boundary issue

$$\tilde{\kappa}_{d,r,y}(x\gamma) = \hat{\kappa}_{d,r,y}(z), \text{ and } \tilde{f}_d(x\gamma) = \hat{f}_d(z),$$

where $z = \min\{\mathcal{X}_\gamma\} + h$ if $x\gamma \in [\min\{\mathcal{X}_\gamma\}, \min\{\mathcal{X}_\gamma\} + h]$, $z = \max\{\mathcal{X}_\gamma\} - h$ if $x\gamma \in [\max\{\mathcal{X}_\gamma\} - h, \max\{\mathcal{X}_\gamma\}]$; otherwise, $z = x\gamma$. We keep the untransformed estimator to simplify technical analysis. Results can be extended to the modified estimators with relative ease.

because it is first-order asymptotically equivalent to the following estimator

$$\tilde{s}_{T_d}(t, x\gamma, \theta) \equiv \phi_\theta^{-1}\left\{-\sum_{Y_i \leqslant t}^{n} R_i\left(\phi_\theta(\hat{s}_d(Y_i, x\gamma)) - \phi_\theta(\hat{s}_d(Y_i, x\gamma) - w_{i,n}(x, \gamma))\right)\right\},$$

where $w_{i,n}(x, \gamma) \equiv \mathbb{1}\{D_i = d\} K_h(x\gamma, X_i\gamma)/(n\hat{f}_d(x\gamma))$. This estimator directly adapts the nonparametric copula graphic estimator in Braekers and Veraverbeke (2005) to the single-index models. Due to the first-order equivalence, inference results in Section 4 and 5 automatically apply to $\tilde{s}_{T,d}$ as well.

When the independence copula is assumed, i.e. $\phi_\theta(u) = -\log u$, (4.5) becomes

$$\hat{s}_{T_d,ind}(t, x\gamma) = \exp\left(-\sum_{i=1}^{n} \frac{w_{i,n}(x, \gamma)R_i}{\sum_{j=1}^{n} w_{j,n}(x, \gamma)\mathbb{1}\{Y_j > Y_i\}}\right) \tag{4.6}$$

$$\approx \prod_{Y_i \leqslant t}\left(1 - \frac{w_{i,n}(x, \gamma)}{\sum_{j=1}^{n} w_{j,n}(x, \gamma)\mathbb{1}\{Y_j > Y_i\}}\right)^{R_i}, \tag{4.7}$$

where the asymptotic equivalence follows roughly from Taylor expanding the exponential function. In the absence of treatment, (4.7) coincides with the conditional single-index Kaplan-Meier estimator as proposed by Strzalkowska-Kominiak and Cao (2014) or Li and Patilea (2018). On the other hand, under random censoring, the average BGF can be directly estimated using the standard Kaplan-Meier estimator of Kaplan and Meier (1958), without going through the conditioning step.

Remark 5 *Note that the random function $t \mapsto \hat{s}_{T_d}(t, x\hat{\gamma}_d, \theta)$ necessarily lies between $[0, 1]$ due to the range constraint on $\phi_\theta^{-1}(\cdot)$. However, the estimator may not be monotonically decreasing in finite samples, even though its population counterpart is constrained to be, as we have discussed in the previous section. To enforce the shape constraint, we will rearrange the initial estimator using the procedure proposed by Chernozhukov, Fernández-Val, and Galichon (2010), and the resulting estimator would be a proper conditional survival function. From Chernozhukov* et al. *(2010), we find that the initial and rearranged estimators are asymptotically equivalent if $s_{T_d}(t, x\gamma_d, \theta)$ is indeed monotone. As a result, we will focus on the asymptotic results for the initial estimator in what follows.*

4.3 Uniform Linear Expansion

In this section, we will provide a linear expansion for the conditional SICG estimator $\hat{s}_{T_d}(t, x\hat{\gamma}_d, \theta)$ that is valid uniformly across t, x, and θ, based on which, a uniform linear representation for the unconditional SICG estimator $\hat{s}_{T_d}(t, \theta)$ is also derived. These uniform representations are crucial for establishing results on weak convergence and bootstrap validity. Here, we first introduce and discuss several assumptions necessary for deriving the claimed results.

Assumption 5 (Data)

1. *The data $\{Y_i, R_i, D_i, X_i\}_{i=1}^n$ are independently and identically distributed.*

2. *There exists $y_o \in (0, \bar{y}]$ and $\upsilon_o > 0$ such that, for each $d \in \{0,1\}$, $F_{Y_d,R_d|X\gamma_d}(y_o, 1|x\gamma_d) \leqslant 1 - \upsilon_o$ almost surely in $x \in \mathcal{X}$. Let $\tilde{\mathcal{T}} \equiv [0, y_o]$.*

Assumption 5.1 is standard. Assumption 5.2, which is also imposed by Rivest and Wells (2001), and Fan and Liu (2018), strengthens Assumption 2.1 by further restricting the support of the event time. Since many generator functions are not finite at 0, the condition is imposed to avoid dealing with a divergent $\phi^{-1}(\cdot)$, in a neighborhood of the origin.

For the following set of assumptions, we define a shrinking neighborhood of γ_d by $\Gamma_{d,n} \equiv \{\|\gamma - \gamma_d\| \leqslant Cn^{-1/2}\}$, for some positive constant C.

Assumption 6 (Smoothness)

1. *For a positive integer $s \geqslant 2$ and $d = 0,1$, (i) The function $v \mapsto f_{d,\gamma_d}(v)$, is $(s+1)$-times continuously differentiable, and the derivatives up to order s are bounded; (ii) $\partial_v^{(s+1)} f_{d,\gamma_d}(v)$ is Lipschitz continuous in v with the Lipschitz constant being independent of v.*

2. *For $d, r \in \{0,1\}$, (i) the function $v \mapsto F_{Y,R|D,X\gamma_d}(y, r|d, v)$ is $(s+1)$-times continuously differentiable and the derivatives up to order s are bounded uniformly on $\tilde{\mathcal{T}}$; (ii) $\partial_v^{(s+1)} F_{Y,R|D,X\gamma_d}(y, r|d, v)$ is Lipschitz continuous in v with the Lipschitz constant being independent of y and v; (iii) $y \mapsto F_{Y,R|D,X\gamma_d}(y, r|d, v)$ is continuously differentiable and the first-order derivative is uniformly bounded with respect to v; (iv) $\partial_y F_{Y,R|D,X\gamma_d}(y, r|d, v)$ is Lipschitz continuous in both y and v, where the Lipschitz constants are independent of y, and v; (v) $\partial_v F_{Y,R|D,X\gamma_d}(y, r|d, v)$ is Lipschitz continuous in y with the Lipschitz constant being independent of v and y;*

3. *(i) The functions $v \mapsto \mathbb{E}[X_{[\ell]}|X\gamma_d = v]$, and $v \mapsto \mathbb{E}[X_{[\ell]}X_{[j]}|X\gamma_d = v]$, $\ell, j = 2, ..., k$, are four times continuously differentiable, and the derivatives up to fourth order are all bounded; (ii) the fourth order derivatives are Lipschitz continuous in v. For $r = 0,1$, $\ell_1 = 0,1$, and $\ell_2 = 0,1$, (iii) $v \mapsto \varrho^\gamma_{\ell_1,\ell_2}(y, v)$[4] is continuously differentiable and the derivatives are bounded uniformly on $\tilde{\mathcal{T}} \times \mathcal{X} \times \Gamma_{d,n}$; (iv) $\partial_v \varrho^\gamma_{\ell_1,\ell_2}(y, v)$ is Lipschitz continuous in v with the Lipschitz constant being independent of y, x, and $\gamma \in \Gamma_{d,n}$.*

4 For a matrix X, $X^{\otimes \ell}$ with $\ell = 0,1,2$ denote $1, X$, and XX', respectively. Define

$$\varrho^\gamma_{\ell_1,\ell_2}(y, x\gamma) \equiv \partial^{\ell_2}_{x\gamma} \left\{ f_d(x\gamma)\, \mathbb{E}\left[G_d^{\ell_1}(y, X\gamma_d)(x_{[-1]} - X_{[-1]})^{\otimes \ell_2} | X\gamma = x\gamma \right] \right\}. \tag{4.8}$$

4. *(i)* $u \mapsto \phi_\theta(u)$ *is three times continuously differentiable with the third order derivative* $\phi_\theta'''(u) \leqslant 0$ *and* $\phi_\theta'''(u)$ *being bounded uniformly for* $(u, \theta) \in [\upsilon_o, 1] \times \Theta$*; (ii)* $1/\dot{\phi}_\theta^{-1}(z)$ *and* $\ddot{\phi}_\theta^{-1}(z)$ *are bounded away from 0 for* $(z, \theta) \in [0, y_o^*] \times \Theta$*, where* $\dot{\phi}_\theta^{-1}(z) \equiv \phi_\theta'(\phi_\theta^{-1}(z))$*,* $\ddot{\phi}_\theta^{-1}(z) \equiv \phi_\theta''(\phi_\theta^{-1}(z))$*, and* $y_o^* = (1 - \upsilon_o)\sup_{(u,\theta)\in[\upsilon_o,1]\times\Theta} \left|\phi_\theta'(u)\right|$*; (iii)* $\phi_\theta'(u)$ *and* $\phi_\theta''(u)$ *are Lipschitz continuous in* θ *with Lipschitz constant being independent of* $u \in [\upsilon_o, 1]$.

Assumption 6 gathers a set of smoothness conditions on various functions. Assumptions 6.1 - 6.2 are assumed in most of prior works, including Delecroix *et al.* (2006), Bouaziz and Lopez (2010), Xia *et al.* (2010), and Chiang and Huang (2012). Assumption 6.3 serves to bound the bias and to control the rate of first-order remainder terms. Assumption 6.4 stipulates that the generator functions exhibit enough smoothness with respect to both u and θ. These requirements, akin to Assumption (C8) in Braekers and Veraverbeke (2005) and Assumption G in Fan and Liu (2018), are necessary when establishing uniformity of the linear expansion of the SICG estimator with respect to θ. Now, we define $\psi_{d,r}^a(t,x) \equiv -\left(x_{[-1]} - \mathbb{E}[X_{[-1]}|x\gamma_d]\right)\partial_{x\gamma} G_{d,r}(t, x\gamma_d)$, and $V_d(t,r) \equiv \mathbb{E}\left[\psi_{d,r}^a(t,X)\psi_{d,r}^a(t,X)' f_d(X\gamma_d)^2\right]$.

Assumption 7 (Index Estimation)

1. *(i) The class of functions* $\{v \mapsto g_{d,r,\gamma}(v;t) : (d,r,t,\gamma) \in \{0,1\}^2 \times \mathcal{T} \times \Gamma\}$ *is of the VC type with bounded envelop function,*[5] *where* $g_{d,r,\gamma}(v;t)$ *is either of the following functions and their derivatives up to the second order:*

   $$v \mapsto f_{d,\gamma}(v), \quad v \mapsto F_{Y,R|D,X\gamma}(t,r|d,v), \quad v \mapsto \mathbb{E}[X_{[\ell]}|X\gamma = v], \quad v \mapsto \mathbb{E}[X_{[\ell]}X_{[j]}|X\gamma = v],$$

   for $\ell, j = 2, \ldots, k$*; (ii) for* $d \in \{0,1\}$ *and each sequence* $\delta_n \to 0$*,*

   $$\sup_{\|\gamma-\gamma_d\|\leqslant\delta_n} \sup_{(t,r,v)\in\mathcal{T}\times\{0,1\}\times\mathcal{X}_\Gamma} |g_{d,r,\gamma}(v;t) - g_{d,r,\gamma}(v;t)| \to 0.$$

2. *There exists a set* $\mathcal{T}_v \subset \mathcal{T}$*, such that* $\mathbb{P}\left((Y,R) \in \mathcal{T}_v \times \{0,1\}\right) > 0$ *and* $V_d(t,r)$ *is positive definite for each* $(t,r,d) \in \mathcal{T} \times \{0,1\}^2$.

This assumption collects several regularity conditions needed for showing asymptotic behavior of the index estimator $\hat{\gamma}$. The first condition provides uniform control for the local difference of second order derivatives of (4.1), while the second condition is imposed to guarantee that the Hessian matrix is positive definite, and thus, the asymptotic variance matrix is invertible.

Assumption 8 (Kernel)

---

5 Precise definition of VC (Vapnik-Cěrvonenkis) type class is recalled in Appendix C

1. *The kernel function, $L(\cdot)$ is symmetric, supported on $[-1, 1]$, and of bounded variation; (ii) it is twice continuously differentiable on $(-1, 1)$ and the derivatives up to the second order are continuous and of bounded variation.*

2. *The kernel function, $K(\cdot)$ is symmetric, supported on $[-1, 1]$ and of bounded variation; (ii) it is twice continuously differentiable on $(-1, 1)$ with uniformly continuous and bounded derivatives; (iii) $\int K(u)du = 1$, $\int u^\ell K(u)du = 0$ for nonnegative integers $\ell < s$, and $\int u^s K(u)du < \infty$.*

Assumption 9 (Bandwidth)

1. *The bandwidth b satisfies: $b \to 0$, $\log n/(nb^3) \to \infty$, $nb^4 \to 0$, as $n \to \infty$.*

2. *The bandwidth h satisfies: $h \to 0$, $\log n/(nh^3) \to 0$, and $nh^{2s} \to 0$, as $n \to \infty$.*

The restrictions on the kernel and the bandwidth are relatively mild. Assumptions 8.1 and 9.1 are imposed to ensure the estimation error from estimating $\boldsymbol{\gamma}$ is of the order less than $n^{-1/2}$. Meanwhile, Assumptions 8.2 and 9.2 provide rates control for the conditional SICG estimator. The smoothness conditions on the kernel functions serve two purposes: (1) it guarantees that $L(b^{-1}(\cdot\gamma_d - x\gamma_d))$ and $K(h^{-1}(\cdot\gamma_d - x\gamma_d))$ belong to the VC type class, which is necessary for establishing uniform convergence of several U-processes arising from the expansion of the kernel estimators; (2) it also allows us to control the rate of bias terms by means of Taylor expansions. Assumption 8 is satisfied by frequently used kernel functions, such as uniform, triangular, biweight, triweight, Epanechnikov kernels, etc. The Gaussian kernel, however, is ruled out due to the compact support condition.[6]

Due to the single-index structure, bandwidth conditions are independent of the dimension of X, k, meaning our estimator is not subject to the "curse of dimensionality". As a result, higher order kernels are not necessary when the covariate is multivariate. We require $nb^3/\log n, nh^3/\log n$ diverge to infinity, so that the first-order expansion of the kernel function with respect to γ is uniformly convergent. By imposing $nh^{2s} \to 0$, we undersmooth to make the bias disappear asymptotically.

Remark 6 *As is the case for all semiparametric estimators, the smoothing parameters play a crucial role in the trade-off between reducing bias and variance. It is therefore desirable to have a data-adaptive way of choosing the parameter. One possibility is to estimate b and γ_d simultaneously via minimizing (4.1) with respect to (b, γ). In practice, we may follow a simple grid search procedure: (i) pick a finite grid $\{b_\ell\}_{\ell=1}^m$ from the set $[b_l n^{-\iota}, b_u n^{-\iota}]$, for some positive constants $b_l < b_u$ and some ι that fulfills Assumption 9.1. (ii) Minimize (4.1) with respect to γ, and record the minimum*

6 Compactness of the kernel function is not essential, and can be relaxed at the expanse of longer proof. See, e.g. Maistre and Patilea (2019) for a detailed treatment.

$\{\hat{\mathcal{J}}_d(b)\}_{d\in\{0,1\}}$ *for each* b *in the grid, and keep the value of bandwidth such that* $\{\hat{\mathcal{J}}_d(b)\}_{d\in\{0,1\}}$ *attains the minimum value. When a second-order kernel is adopted, i.e.* $s = 2$*, we may set* h *equal to* b*.*

Now, we define a few more quantities related to the influence functions. Let $\mathcal{E}_{d,\gamma} = \sum_{r=0,1} \mathcal{E}_{d,r,\gamma}$, and $\psi_d^a = \sum_{r=0,1} \psi_{d,r}^a$. Moreover, $\psi_d^b(W) \equiv \int_{\tilde{\mathcal{T}}\times\{0,1\}} \mathbb{E}[\psi_{d,r}^a(y, X_1) f_d(X_1\gamma_d)\rho(X_1, X)|X]$ $U_{d,\gamma_d}(y, r)dF_{Y,R}(y, r)$, and

$$V_d \equiv \int_{\tilde{\mathcal{T}}\times\{0,1\}} \mathbb{E}\left[\psi_d^a(y, X_1)\psi_d^a(y, X_2)' f_{d,\gamma_d}(X_1\gamma_d) f_{d,\gamma_d}(X_2\gamma_d)\rho(X_1, X_2)\right] dF_{Y,R}(y, r),$$

$$\Psi_d\left(f_1, f_2\right)(t, x, \theta) \equiv \frac{1}{\phi_\theta'\left(s_{T_d}(t, x\gamma_d, \theta)\right)}\left\{-\int_0^t \phi_\theta''\left(s_d(y, x\gamma_d)\right) f_1(y) G_{d,1}(dy, x\gamma_d)\right.$$
$$\left. +\phi_\theta'\left(s_d(t, x\gamma_d)\right) f_2(t) - \int_0^t \phi_\theta''\left(s_d(y, x\gamma_d)\right) f_2(y) s_d(dy, x\gamma_d)\right\},$$

where $\Psi_d(\cdot,\cdot)$ is a functional mapping from $\ell_\infty(\tilde{\mathcal{T}}) \times \ell_\infty(\tilde{\mathcal{T}})$ to $\ell_\infty(\tilde{\mathcal{T}} \times \mathcal{X} \times \Theta)$, for $d \in \{0, 1\}$.[7]

Theorem 4.1 (Uniform asymptotic linear representation) *Suppose Assumptions 1 - 9 hold,*

$$\hat{s}_{T_d}(t, x\hat{\gamma}_d, \theta) - s_{T_d}(t, x\gamma_d, \theta) = \frac{1}{n}\sum_{i=1}^n \sum_{j\in\{s,b,l\}} \eta_{j,d}(W_i, x, t, \theta) + r_n(x, t, \theta)$$

where $\eta_{s,d}(W, x, t, \theta) \equiv K_h\left(x\gamma_d, X\gamma_d\right)\Psi_d\left(\mathcal{E}_{d,\gamma_d}, \mathcal{E}_{d,1,\gamma_d}\right)(t, x, \theta)/f_d(x\gamma_d)$, $\eta_{b,d}(W, x, t, \theta) \equiv K_h\left(x\gamma_d, X\gamma_d\right)\Psi_d\left(G_d(\cdot, X\gamma_d) - G_d(\cdot, x\gamma_d), G_{d,1}(\cdot, X\gamma_d) - G_{d,1}(\cdot, x\gamma_d)\right)(t, x, \theta)/f_d(x\gamma_d)$, $\eta_{l,d}(W, x, t, \theta) \equiv \psi_d^b(W)'V_d^{-1}\Psi_d\left(\psi_d^a, \psi_{d,1}^a\right)(t, x, \theta)/f_d(x\gamma_d)$, *and* $\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta}$ $|r_n(x, t, \theta)| = O_p\left((\log n)^{1/2} n^{-1}h^{-3/2}\right)$.

This theorem is the second main result of this article. It shows that, the conditional SICG estimator is asymptotically linear, and its influence functions can be split into four parts. The first two term, $\eta_{s,d}$ and $\eta_{b,d}$ are associated with the stochastic part and the bias of the usual kernel expansions. Of the two, the first component dominates in the limit with a uniform rate of $O_p\left((\log n)^{1/2} \cdot n^{-1/2}h^{-1/2}\right)$, free from the curse of dimensionality. The third component, $\eta_{l,d}$, is unique to the SICG estimator. It arises from the estimation of the index parameters and converges at the parametric rate, implying that the estimation error of the index coefficients is asymptotically negligible. Consequently, the main conclusions of the previous theorem remain intact when estimators other than $\hat{\gamma}$ are used, provided such estimators are root-n consistent.

7 For a generic set $\mathcal{S}$, $\ell_\infty(\mathcal{S})$ is the space of all uniformly bounded real functions on $\mathcal{S}$, equipped with the supremum norm, $\|f\|_{\mathcal{S}} \equiv \sup_{s\in\mathcal{S}} |f(s)|$

4.4 Weak Convergence

This uniform linear representation allows us to apply techniques in the empirical process literature to establish weak convergence of the bound generating processes. The weak convergence, denoted by "$\Rightarrow$", is in the sense of Hoffmann–Jørgensen–Dudley, as recalled in the Appendix C. See, also Definition 1.3.3 of Van Der Vaart and Wellner (1996). The convergence takes place in $\ell_\infty(\mathcal{S})$.

Again, we will first need to introduce a few notations before stating the results. For $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \Theta^2$, we collect the conditional and unconditional SICG estimators by $\hat{\mathbf{S}}^x(t, \boldsymbol{\theta}) \equiv (\hat{s}_{T_1}(t, x\hat{\gamma}_d, \theta_1),\ \hat{s}_{T_0}(t, x\hat{\gamma}_d, \theta_2))'$, and by $\hat{\mathbf{S}}(t, \boldsymbol{\theta}) \equiv (\hat{s}_{T_1}(t, \theta_1),\ \hat{s}_{T_0}(t, \theta_2))'$, respectively. Analogously, the BGFs are collected by $\mathbf{S}^x$ and $\mathbf{S}$, respectively.

We refer to $\hat{\mathbb{G}}_n^x(\cdot, \cdot) \equiv \sqrt{nh}\left(\hat{\mathbf{S}}^x(\cdot, \cdot) - \mathbf{S}^x(\cdot, \cdot)\right)$, for a fix $x \in \mathcal{X}$, as the *conditional bound generating process* (CBGP), and let $\hat{\mathbb{G}}_n(\cdot, \cdot) \equiv \sqrt{n}\left(\hat{\mathbf{S}}(\cdot, \cdot) - \mathbf{S}(\cdot, \cdot)\right)$ stand for the *unconditional bound generating process* (UBGP). The main goal of this section is to show that both CBGP and UBGP converge weakly to centered Gaussian processes. For this purpose, we need an additional assumption, which is given as follows.

Assumption 10 *(i) $1/\dot{\phi}_\theta^{-1}(z)$ is Lipschitz continuous in θ with Lipschitz constant being independent of θ and $z \in [0, y_o^*]$; (ii) $u \mapsto \phi_\theta''(u)$ is $(s+1)$ times continuously differentiable with the $(s+1)$-th order derivative being bounded uniformly for $(u, \theta) \in [\upsilon_o, 1] \times \Theta$;*

Assumption 10.(i) allows us to uniformly bound the derivative of $\phi_\theta^{-1}(\cdot)$, and also plays a part in controlling the size of functional space associated with influence functions of the CBGP. Assumption 10.(ii) strengthens Assumption 6.4.(i). It ensures that the bias from approximating the UBGP by an empirical process is uniformly negligible. Most generator functions of the Archimedean family satisfy this stronger smoothness condition.

In the following corollary, we establish weak convergence of the CBGP. Its proof is the combination of Theorem 4.1 and Theorem 10.6 in Pollard (1990), the latter of which provides a set of sufficient conditions for the weak convergence of triangular arrays of non-identically distributed random elements.

Corollary 1 *(i) Under the assumptions of Theorem 4.1, and suppose that Assumption 10.(i) hold, then $\hat{\mathbb{G}}_n^x(\cdot, \cdot) \Rightarrow \mathbb{G}^x(\cdot, \cdot)$, in $\ell_\infty(\tilde{\mathcal{T}} \times \Theta^2) \times \ell_\infty(\tilde{\mathcal{T}} \times \Theta^2)$, where $\mathbb{G}^x$ is a two-dimensional, tight, centered Gaussian process with covariance function,*

$$\Sigma_\eta^x(\mathbf{t}, \boldsymbol{\theta}) = \mathbb{E}\left[\boldsymbol{\varphi}^x(W, t_1, \boldsymbol{\theta}_1)\boldsymbol{\varphi}^x(W, t_2, \boldsymbol{\theta}_2)'\right],$$

for each $\mathbf{t} \equiv (t_1, t_2)' \in \tilde{\mathcal{T}} \times \tilde{\mathcal{T}}$, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')' \in \Theta^2 \times \Theta^2$, and for $j = 1, 2$, $\boldsymbol{\theta}_j = (\theta_j, \tilde{\theta}_j)$, $\boldsymbol{\varphi}^x(w, t, \boldsymbol{\theta}) = (\eta_{s,1}(w, x, t, \theta),\ \eta_{s,0}(w, x, t, \tilde{\theta}))$.

To the best of our knowledge, this result is new to the literature. This result differs from Theorem 2 in Braekers and Veraverbeke (2005) in a number of ways. First, our CBGP is indexed not only by the time t but also by the copula parameter θ. In comparison, they study a similar process indexed by the time only. Consequently, our result generalizes theirs by relaxing the restrictive assumption that copula is completely known. In view of Tsiatis (1975)'s non-identifiability result, such a generalization is a crucial first step towards our sensitivity analysis. Secondly, they consider a univariate fixed design for the covariates, whereas we adopt a single-index model that accommodates multivariate random variables, and allows them to be either discrete or continuous. Despite these differences, the covariance functions share a similar structure in the two papers. See Appendix C.2.1 for formulas.

The following corollary records results on the UBGP that are parallel to Theorem 4.1 and Corollary 1.

Corollary 2 *(i) Suppose the assumptions of Theorem 4.1, and Assumption 10 hold, we have that, for* $(d, t, \theta) \in \{0, 1\} \times \tilde{\mathcal{T}} \times \Theta$*,*

$$\hat{s}_{T_d}(t, \theta) - s_{T_d}(t, \theta) = \frac{1}{n} \sum_{i=1}^{n} \varphi_d(W_i, t, \theta) + R_n(t, \theta),$$

where $\varphi_d = \sum_{j=1}^{2} \varphi_{d,j}$, $\varphi_{d,1}(W, t, \theta) = \Psi_d\left(\mathcal{E}_{d,\gamma_d}, \mathcal{E}_{d,1,\gamma_d}\right)(t, X, \theta) f(X\gamma_d)/f_d(X\gamma_d)$, $\varphi_{d,2}(W, t, \theta) = s_{T_d}(t, X\gamma_d, \theta) - s_{T_d}(t, \theta)$*, and* $\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta} |R_n(t, \theta)| = o_p\left(n^{-1/2}\right)$.

(ii) Furthermore, $\hat{\mathbb{G}}_n(\cdot, \cdot) \Rightarrow \mathbb{G}(\cdot, \cdot)$*, in* $\ell_\infty(\tilde{\mathcal{T}} \times \Theta^2) \times \ell_\infty(\tilde{\mathcal{T}} \times \Theta^2)$*, where* $\mathbb{G}$ *is a two-dimensional, tight, centered Gaussian process with covariance function*

$$\Sigma_\varphi(\mathbf{t}, \boldsymbol{\theta}) = \mathbb{E}\left[\boldsymbol{\varphi}(W, t_1, \boldsymbol{\theta}_1)\boldsymbol{\varphi}(W, t_2, \boldsymbol{\theta}_2)'\right],$$

for each $\mathbf{t} \equiv (t_1, t_2)' \in \tilde{\mathcal{T}} \times \tilde{\mathcal{T}}$ *and* $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')' \in \Theta^2 \times \Theta^2$*, and for* $j = 1, 2$, $\boldsymbol{\theta}_j \equiv (\theta_j, \tilde{\theta}_j)$, $\boldsymbol{\varphi}(W, t, \boldsymbol{\theta}) = (\varphi_1(W, t, \theta),\ \varphi_0(W, t, \tilde{\theta}))$.

As a first result, Corollary 2 provides a uniform linear expansion for the unconditional SICG estimator. The influence function can be decomposed into two parts. The first part, $\varphi_{d,1}$, comes from estimating the conditional BGF. It is the linear representation of the first-order Hoeffding projection of the dominant U statistic. The second component, $\varphi_{d,2}$, arises from the sampling variation of X. Based on this uniform expansion, we show that the UBGP, as a process indexed by both t and θ, converges weakly to a centered Gaussian process. The rates of convergence are, however, different from the CBGP.

With these two corollaries in hand, we are equipped to present inference theories on the estimators of TEBFs. According to our discussion earlier, TEBFs are smooth functionals of the BGFs. It implies that, we can apply the functional delta method (see e.g. Theorem 3.9.5 in

Van Der Vaart and Wellner (1996)), and show that the plug-in estimators of TEBFs will satisfy functional central limit theorems.

Now, let us define plug-in estimators for TEBFs. For $i = 1, ..., n$, denote the i-th order statistics of Y in the sample by $Y_{i:n}$. Let the quantile curve estimator be given by $\hat{q}_{d,\theta}(\tau) \equiv \inf\{y : \hat{s}_{T_d}(y, \theta) \leqslant 1 - \tau\}$, and its conditional version, by $\hat{q}^x_{d,\theta}(\tau) \equiv \inf\{y : \hat{s}_{T_d}(y, x\hat{\gamma}_d, \theta) \leqslant 1 - \tau\}$. With these notations, and in view of (3.6) - (3.9), we consider the following estimators of the lower TEBFs,

$$\begin{aligned}\hat{\nu}_{lb,ATE}(\boldsymbol{\theta}) \equiv \sum_{i=1}^{n-1} \mathbb{1}\left\{Y_{(i+1):n} \leqslant y_o\right\}\left(Y_{(i+1):n} - Y_{i:n}\right)\left(\hat{s}_{T_1}(Y_{i:n}, \theta_2) - \hat{s}_{T_0}(Y_{i:n}, \theta_1)\right)& \\ - y_o\left(\hat{s}_{T_1}(y_o, \theta_2) - \hat{s}_{T_0}(y_o, \theta_1)\right),& \quad (4.9)\end{aligned}$$

$$\hat{\nu}_{lb,DTE}(t, \boldsymbol{\theta}) \equiv \hat{s}_{T,0}(t, \theta_2) - \hat{s}_{T,1}(t, \theta_1), \quad (4.10)$$

$$\hat{\nu}_{lb,QTE}(\tau, \boldsymbol{\theta}) \equiv \hat{q}_{1,\theta_2}(\tau) - \hat{q}_{0,\theta_1}(\tau), \quad (4.11)$$

$$\hat{\nu}_{lb,CHTE}(t, \boldsymbol{\theta}) \equiv \log(\hat{s}_{T,0}(t, \theta_2)) - \log(\hat{s}_{T,1}(t, \theta_1)), \quad (4.12)$$

where $\theta_1 \leqslant \theta_2$. $t \in \tilde{\mathcal{T}}$, and $\tau \in (0, \tau_o)$, where $\tau_o \equiv 1 - \sup_{(x,\theta)\in\mathcal{X}\times\Theta} s_{T_d}(y_o, x\gamma_d, \theta)$. Estimators of the upper TEBFs can be constructed, by swapping the places of θ_1 and θ_2 on the right hand side of preceding equations. Here, we have restricted the upper bound of $\mathcal{T}$ to y_o, in order to avoid entering into the explosive tail area of the generator functions. Consistency of the ATE requires that y_o be sufficiently close to $\bar{y}$. Toward this end, we may set y_o as a large value close to $Y_{n:n}$ in practice.

To understand the formula for $\hat{\nu}_{lb,ATE}$, we note that $\hat{s}_{T_1}(t, \theta)$ and $\hat{s}_{T_0}(t, \theta)$ are step functions in t, with jumps at $\{Y_{i:n}\}_{i=1}^n$ only. The integral over $[0, t]$ can thus be divided into intervals with end points set by the order statistics. In each interval, the integrand is constant, yielding the product form. The second line is added to adjust for the truncation error induced by the artificial upper bound y_o.

In what remains of this section, we will investigate the asymptotic behavior of $\sqrt{n}\left(\hat{\boldsymbol{\nu}}_j - \boldsymbol{\nu}_j\right)$ as well as $\sqrt{nh}\left(\hat{\boldsymbol{\nu}}^x_j - \boldsymbol{\nu}^x_j\right)$, for $j \in \{ATE, DTE, QTE, CHTE\}$. Again, let us introduce a few quantities, which are related to the influence functions of limiting processes. For $\boldsymbol{\theta} = (\theta_1, \theta_2) \in \Theta^2$, with $\theta_1 \leqslant \theta_2$, define $\boldsymbol{\psi}_j(W, u, \boldsymbol{\theta}) = (\psi_{1,j}(W, u, \theta_2) - \psi_{0,j}(W, u, \theta_1), \psi_{1,j}(W, u, \theta_1) - \psi_{0,j}(W, u, \theta_2))'$ and $\boldsymbol{\psi}^x_j(W, u, \boldsymbol{\theta}) = (\psi^x_{1,j}(W, u, \theta_2) - \psi^x_{0,j}(W, u, \theta_1), \psi^x_{1,j}(W, u, \theta_1) - \psi^x_{0,j}(W, u, \theta_2))'$, for $j \in \{ATE, QTE\}$. Meanwhile, we let $\boldsymbol{\psi}_j(W, u, \boldsymbol{\theta}) = (\psi_{1,j}(W, u, \theta_1) - \psi_{0,j}(W, u, \theta_2), \psi_{1,j}(W, u, \theta_2) - \psi_{0,j}(W, u, \theta_1))'$ $\boldsymbol{\psi}^x_j(W, u, \boldsymbol{\theta}) = (\psi^x_{1,j}(W, u, \theta_1) - \psi^x_{0,j}(W, u, \theta_2), \psi^x_{1,j}(W, u, \theta_2) - \psi^x_{0,j}(W, u, \theta_1))'$, for $j \in \{DTE, CHTE\}$, where

$$\psi_{d,ATE}(W, \theta) = \int_{\tilde{\mathcal{T}}} \varphi_d(W, y, \theta)dy - y_o\varphi_d(W, y_o, \theta),$$

$$\psi^x_{d,ATE}(W,\theta) = \int_{\tilde{\mathcal{T}}} \eta_{s,d}(W,x,y,\theta)dy - y_o\eta_{s,d}(W,x,y_o,\theta),$$
$$\psi_{d,DTE}(W,t,\theta) = -\varphi_d(W,t,\theta), \qquad \psi^x_{d,DTE}(W,t,\theta) = -\eta_{s,d}(W,x,t,\theta),$$
$$\psi_{d,QTE}(W,\tau,\theta) = \frac{\varphi_d(W,q_{d,\theta}(\tau),\theta)}{f_{T_d}(q_{d,\theta}(\tau),\theta)}, \qquad \psi^x_{d,QTE}(W,\tau,\theta) = \frac{\eta_{s,d}(W,x,q^x_{d,\theta}(\tau),\theta)}{f_{T_d,x}(q^x_{d,\theta}(\tau),\theta)},$$
$$\psi_{d,CHTE}(W,t,\theta) = -\frac{\varphi_d(W,t,\theta)}{s_{T_d}(t,\theta)}, \qquad \psi^x_{d,CHTE}(W,t,\theta) = -\frac{\eta_{s,d}(W,x,t,\theta)}{s_{T_d}(t,x\gamma_d,\theta)},$$

for $d \in \{0,1\}$.

The next theorem establishes uniform central limit theorems for the conditional and overall TEBF estimators.

Theorem 4.2 *(i) Suppose the assumptions of Corollary 1 hold. Then, for $j = ATE, DTE, CHTE$,*

$$\sqrt{nh}\left(\hat{\boldsymbol{\nu}}^x_j(\cdot,\cdot) - \boldsymbol{\nu}^x_j(\cdot,\cdot)\right) \Rightarrow \nu'_{j,\mathbf{S}^x}(\mathbb{G}^x)(\cdot,\cdot),$$

in $\ell_\infty(\mathcal{U}\times\Theta^2)\times\ell_\infty(\mathcal{U}\times\Theta^2)$,[8] *where $\nu'_{j,\mathbf{S}^x}(\mathbb{G}^x)(\cdot,\cdot)$ is a tight, two-dimensional, centered Gaussian process with covariance kernels $\Sigma^x_j(\mathbf{u},\boldsymbol{\theta}) = \mathbb{E}[\boldsymbol{\psi}^x_j(W,u_1,\boldsymbol{\theta}_1)\boldsymbol{\psi}^x_j(W,u_2,\boldsymbol{\theta}_2)']$, and $\mathbf{u} = (u_1,u_2)$, $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1,\boldsymbol{\theta}'_2)'$. If in addition, for $d \in \{0,1\}$, $0 < \inf_{(\tau,\theta)\in(0,\tau_o)\times\Theta} f_{T_d}(q_{d,\theta}(\tau),\theta) < \sup_{(\tau,\theta)\in(0,\tau_o)\times\Theta} f_{T_d}(q_{d,\theta}(\tau),\theta) < \infty$, we have, in $\ell_\infty(\mathcal{U}\times\Theta^2)\times\ell_\infty(\mathcal{U}\times\Theta^2)$,*

$$\sqrt{nh}\left(\hat{\boldsymbol{\nu}}^x_{QTE}(\cdot,\cdot) - \boldsymbol{\nu}^x_{QTE}(\cdot,\cdot)\right) \Rightarrow \nu'_{QTE,\mathbf{S}^x}(\mathbb{G}^x)(\cdot,\cdot).$$

The tight, two-dimensional process $\nu'_{QTE,\mathbf{S}^x}(\mathbb{G}^x)$ is centered Gaussian with covariance kernel $\Sigma^x_{QTE}(\boldsymbol{\tau},\boldsymbol{\theta}) = \mathbb{E}[\boldsymbol{\psi}^x_{QTE}(W,\tau_1,\boldsymbol{\theta}_1)\boldsymbol{\psi}^x_{QTE}(W,\tau_2,\boldsymbol{\theta}_2)']$.

(ii) Suppose the assumptions of Corollary 2 hold. Then, for $j \in \{ATE, DTE, CHTE\}$,

$$\sqrt{n}\left(\hat{\boldsymbol{\nu}}_j(\cdot,\cdot) - \boldsymbol{\nu}_j(\cdot,\cdot)\right) \Rightarrow \nu'_{j,\mathbf{S}}(\mathbb{G})(\cdot,\cdot),$$

in $\ell_\infty(\mathcal{U}\times\Theta^2)\times\ell_\infty(\mathcal{U}\times\Theta^2)$, where $\nu'_{j,\mathbf{S}}(\mathbb{G})(\cdot,\cdot)$ is a tight, two-dimensional, centered Gaussian process with covariance kernels $\Sigma_j(\mathbf{u},\boldsymbol{\theta}) = \mathbb{E}[\boldsymbol{\psi}_j(W,u_1,\boldsymbol{\theta}_1)\boldsymbol{\psi}_j(W,u_2,\boldsymbol{\theta}_2)']$, and $\mathbf{u} = (u_1,u_2)$, $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1,\boldsymbol{\theta}'_2)'$. If in addition, for $d \in \{0,1\}$, $0 < \inf_{(\tau,x,\theta)\in(0,\tau_o)\times\mathcal{X}\times\Theta} f_{T_d,x}(q^x_{d,\theta}(\tau),\theta) < \sup_{(\tau,x,\theta)\in(0,\tau_o)\times\mathcal{X}\times\Theta} f_{T_d,x}(q^x_{d,\theta}(\tau),\theta) < \infty$, we have, in $\ell_\infty(\mathcal{U}\times\Theta^2)\times\ell_\infty(\mathcal{U}\times\Theta^2)$,

$$\sqrt{n}\left(\hat{\boldsymbol{\nu}}_{QTE}(\cdot,\cdot) - \boldsymbol{\nu}_{QTE}(\cdot,\cdot)\right) \Rightarrow \nu'_{QTE,\mathbf{S}}(\mathbb{G})(\cdot,\cdot).$$

The tight, two-dimensional process $\nu'_{QTE,\mathbf{S}}(\mathbb{G})$ is centered Gaussian with covariance kernel

8 Definition of the set $\mathcal{U}$ depends on the type of treatment effect under consideration. Specifically, $\mathcal{U} = \varnothing$ if $j = ATE$, $\mathcal{U} = (0,\tau_o)$, if $j = QTE$. Otherwise, $\mathcal{U} = \tilde{\mathcal{T}}$.

$\Sigma_{QTE}(\boldsymbol{\tau}, \boldsymbol{\theta}) = \mathbb{E}[\boldsymbol{\psi}_{QTE}(W, \tau_1, \boldsymbol{\theta}_1)\boldsymbol{\psi}_{QTE}(W, \tau_2, \boldsymbol{\theta}_2)'].$

Theorem 4.2 forms the basis for pointwise as well as uniform inference on the TEBFs. Nonetheless, the result cannot be directly used for such a purpose, since the limit processes contain various unknown quantities. The approximation of these quantities can be avoided by means of a standard nonparametric bootstrap procedure. However, its implementation requires recalculating estimators of $\boldsymbol{\gamma}$, the BGFs, and the TEBFs in each bootstrap iteration. Since optimization of (4.1) is computationally intensive, we adopt an alternative multiplier bootstrap procedure that entail approximating the influence functions, but dispense with the need for reoptimizations.

5 Multiplier Bootstrap

In this section, we propose simulation methods, based on the multiplier bootstrap, for approximating the limiting processes introduced in the previous section. We show the bootstrapped processes converge uniformly to the limiting Gaussian processes defined in Corollaries 1, 2, and Theorem 4.2. Given these theoretical results, we then provide practical algorithms for conducting pointwise and uniform inference on the treatment effects.

Let $\{\xi_{i,b}\}_{i=1}^n$ be a sequence of random variables with zero mean and unit variance. We call them the *multiplier weights*. These weights are drawn independently of the main sample $\{W_i\}_{i=1}^n$. Given the weights, we define the following two multiplier processes,

$$\mathbb{G}_{n,\xi}^x(t, \boldsymbol{\theta}) \equiv n^{-1/2} h^{1/2} \sum_{i=1}^n \boldsymbol{\varphi}^{x*}(W_i, \xi_i, t, \boldsymbol{\theta}), \quad \text{and} \quad \mathbb{G}_{n,\xi}(t, \boldsymbol{\theta}) \equiv n^{-1/2} \sum_{i=1}^n \boldsymbol{\varphi}^*(W_i, \xi_i, t, \boldsymbol{\theta}),$$

where $\boldsymbol{\varphi}^{x*} \equiv (\eta_{s,1}^*, \eta_{s,0}^*)'$, and $\eta_{s,d}^*(W, \xi, t, \theta) \equiv \xi \cdot \eta_{s,d}(W, t, \theta)$. We also have $\boldsymbol{\varphi}^* \equiv (\varphi_1^*, \varphi_0^*)'$, and $\varphi_d^*(W, \xi, t, \theta) \equiv \xi \{\varphi_{d,1}(W, t, \theta) + \varphi_{d,2}(W, t, \theta)\}$. Since the influence functions contain unknown quantities, we need to replace them with their estimates in practice. We propose to reuse $\hat{G}_d(y, x\hat{\gamma}_d)$ and $\hat{G}_{d,1}(y, x\hat{\gamma}_d)$ in the construction of the influence function estimators:

$$\begin{aligned}
&\hat{\mathcal{E}}_{d,\hat{\gamma}_d}(y, x) = \mathbb{1}\{D = d\}(\mathbb{1}\{Y \leqslant y\} - \hat{G}_d(y, x\hat{\gamma}_d)),\\
&\hat{\mathcal{E}}_{d,1,\hat{\gamma}_d}(y, x) = \mathbb{1}\{D = d\}(R\mathbb{1}\{Y \leqslant y\} - \hat{G}_{d,1}(y, x\hat{\gamma}_d)),\\
&\hat{\Psi}_d\left(\hat{\mathcal{E}}_{d,\hat{\gamma}_d}, \hat{\mathcal{E}}_{d,1,\hat{\gamma}_d}\right)(t, x, \theta) = \frac{1}{\phi_\theta'\left(\hat{s}_{T,d}(t, x\hat{\gamma}_d, \theta)\right)}\Big\{\phi_\theta'\left(\hat{s}_d(t, x\hat{\gamma}_d)\right)\hat{\mathcal{E}}_{d,1,\hat{\gamma}_d}(t, x)\\
&\qquad + \frac{1}{n\hat{f}(x\hat{\gamma}_d, d)}\sum_{i=1}^n I_{d,t,i}\phi_\theta''\left(\hat{s}_d(Y_i, x\hat{\gamma}_d)\right)(\hat{\mathcal{E}}_{d,1,\hat{\gamma}_d}(Y_i, x) - R_i\hat{\mathcal{E}}_{d,\hat{\gamma}_d}(Y_i, x))K_h(x\hat{\gamma}_d, X_i\hat{\gamma}_d)\Big\}.
\end{aligned}$$

We define the estimated multiplier processes by substituting the above estimators into the multiplier processes. Specifically,

$$\hat{\mathbb{G}}^x_{n,\xi}(t,\boldsymbol{\theta}) = n^{-1/2}h^{1/2}\sum_{i=1}^n \hat{\boldsymbol{\varphi}}^{x*}(W_i,\xi_i,t,\boldsymbol{\theta}), \text{ and } \hat{\mathbb{G}}_{n,\xi}(t,\boldsymbol{\theta}) = n^{-1/2}\sum_{i=1}^n \hat{\boldsymbol{\varphi}}^{*}(W_i,\xi_i,t,\boldsymbol{\theta}),$$

where $\hat{\boldsymbol{\varphi}}^{x*} = (\hat{\varphi}^{x*}_1, \hat{\varphi}^{x*}_0)'$, $\hat{\boldsymbol{\varphi}}^{*} = (\hat{\varphi}^{*}_1, \hat{\varphi}^{*}_0)'$, $\hat{\varphi}^{x*}_d(W,\xi,t,\theta) = \xi \cdot \hat{\eta}_{s,d}(W,x,t,\theta)$, $\hat{\varphi}^{*}_d(W,\xi,t,\theta) = \xi \cdot \{\hat{\varphi}_{d,1}(W,t,\theta) + \hat{\varphi}_{d,2}(W,t,\theta)\}$, and

$$\hat{\eta}_{s,d}(W,\xi,x,t,\theta) = \frac{K_h(x\hat{\gamma}_d, X\hat{\gamma}_d)}{\hat{f}(x\hat{\gamma}_d,d)}\hat{\Psi}_d\left(\hat{\mathcal{E}}_{d,\hat{\gamma}_d},\hat{\mathcal{E}}_{d,1,\hat{\gamma}_d}\right)(t,x,\theta), \tag{5.1}$$

$$\hat{\varphi}_{d,1}(W,t,\theta) = \frac{\hat{f}(X\hat{\gamma}_d)}{\hat{f}(X\hat{\gamma}_d,d)}\hat{\Psi}_d\left(\hat{\mathcal{E}}_{d,\hat{\gamma}_d},\hat{\mathcal{E}}_{d,1,\hat{\gamma}_d}\right)(t,X,\theta), \tag{5.2}$$

$$\hat{\varphi}_{d,2}(W,t,\theta) = \hat{s}_{T,d}(t,X\hat{\gamma}_d,\theta) - \mathbb{E}_n[\hat{s}_{T,d}(t,X\hat{\gamma}_d,\theta)]. \tag{5.3}$$

Assumption 11 (Multiplier weights) *$\{\xi_i\}_{i=1}^n$ is a sequence of $i.i.d.$ random variables, defined on a probability space independent of $\{W_i\}_{i=1}^n$, satisfying $E[\xi_1] = 0$ and $\mathbb{E}[\xi_1^2] = 1$.*

There are several different choices for ξ that are commonly encountered in the literature. For instance, when $\xi = \mathcal{N}$, where $\mathcal{N}$ is a standard normal random variable, it is referred to as the Gaussian multiplier method, see Giné and Zinn (1984). When $\xi = \mathcal{N}_1/\sqrt{2} + (\mathcal{N}_2^2 - 1)/2$, where $\mathcal{N}_1$ and $\mathcal{N}_1$ are mutually independent standard normal random variables, it corresponds to the wild bootstrap mthod, see Mammen (1993).

In the next theorem, we show that the estimated multiplier processes $\hat{\mathbb{G}}^x_{n,\xi}$ and $\hat{\mathbb{G}}_{n,\xi}$ approximate $\mathbb{G}^x$ and $\mathbb{G}$, respectively. The approximation, formally termed as *conditional weak convergence in probability*, where the condition is on the main sample, is in the sense of Section 2.2.3 in Kosorok (2008).

Theorem 5.1 *Under the assumptions of Theorem 4.1, Assumptions 10, and 11, we have that (i) $\hat{\mathbb{G}}^x_{n,\xi} \underset{\xi}{\overset{p}{\rightarrow}} \mathbb{G}^x$, and (ii) $\hat{\mathbb{G}}_{n,\xi} \underset{\xi}{\overset{p}{\rightarrow}} \mathbb{G}$.*

Theorem 5.1, combined with the functional delta method for the bootstrap (see e.g. Theorem 3.9.11 in Van Der Vaart and Wellner (1996)), allows us to establish the validity of plug-in estimators of Hadamard differentiable functionals. Let us first define the estimated multiplier processes for the bound curves, $\hat{\mathbb{G}}_{\xi,j}$ and $\hat{\mathbb{G}}^x_{\xi,j}$, by

$$\hat{\mathbb{G}}^x_{\xi,j}(u,\boldsymbol{\theta}) = n^{-1/2}h^{1/2}\sum_{i=1}^n \hat{\boldsymbol{\psi}}^{x*}_j(W_i,\xi_i,u,\boldsymbol{\theta}), \text{ and } \hat{\mathbb{G}}_{\xi,j}(u,\boldsymbol{\theta}) = n^{-1/2}\sum_{i=1}^n \hat{\boldsymbol{\psi}}^{*}_j(W_i,\xi_i,u,\boldsymbol{\theta}), \tag{5.4}$$

for $j \in \{ATE, DTE, QTE, CHTE\}$, $\boldsymbol{\theta} = (\theta_1,\theta_2)' \in \Theta^2$, with $\theta_1 \leqslant \theta_2$. In the preceding definition, we use the following estimators of the influence functions: $\hat{\boldsymbol{\psi}}^{*}_j(W,\xi,u,\boldsymbol{\theta}) \equiv$

$\xi \cdot (\hat{\psi}_{1,j}(W,u,\theta_2) - \hat{\psi}_{0,j}(W,u,\theta_1), \hat{\psi}_{1,j}(W,u,\theta_1) - \hat{\psi}_{0,j}(W,u,\theta_2))'$ and $\hat{\boldsymbol{\psi}}_j^{x*}(W,\xi,u,\boldsymbol{\theta}) \equiv \xi \cdot (\hat{\psi}^x_{1,j}(W,u,\theta_2) - \hat{\psi}^x_{0,j}(W,u,\theta_1), \hat{\psi}^x_{1,j}(W,u,\theta_1) - \hat{\psi}^x_{0,j}(W,u,\theta_2))'$, for $j \in \{ATE, QTE\}$; moreover, $\hat{\boldsymbol{\psi}}^*_j(W,\xi,u,\boldsymbol{\theta}) \equiv \xi \cdot (\hat{\psi}_{1,j}(W,u,\theta_1) - \hat{\psi}_{0,j}(W,u,\theta_2), \hat{\psi}_{1,j}(W,u,\theta_2) - \hat{\psi}_{0,j}(W,u,\theta_1))'$ $\hat{\boldsymbol{\psi}}_j^{x*}(W,\xi,u,\boldsymbol{\theta}) \equiv \xi \cdot (\hat{\psi}^x_{1,j}(W,u,\theta_1) - \hat{\psi}^x_{0,j}(W,u,\theta_2), \hat{\psi}^x_{1,j}(W,u,\theta_2) - \hat{\psi}^x_{0,j}(W,u,\theta_1))'$, for $j \in \{DTE, CHTE\}$, where

$$\hat{\psi}_{d,ATE}(W,\theta) = \sum_{i=1}^{n-1} \mathbb{1}\left\{Y_{(i+1):n} \leqslant y_o\right\} \left(Y_{(i+1):n} - Y_{i:n}\right) \hat{\varphi}_d(W, Y_{i:n}, \theta) - y_o \hat{\varphi}_d(W, y_o, \theta), \tag{5.5}$$

$$\hat{\psi}^x_{d,ATE}(W,\theta) = \sum_{i=1}^{n-1} \mathbb{1}\left\{Y_{(i+1):n} \leqslant y_o\right\} \left(Y_{(i+1):n} - Y_{i:n}\right) \hat{\eta}_{s,d}(W, x, Y_{i:n}, \theta) - y_o \hat{\eta}_{s,d}(W, x, y_o, \theta), \tag{5.6}$$

$$\hat{\psi}_{d,DTE}(W,t,\theta) = -\hat{\varphi}_d(W,t,\theta), \qquad \hat{\psi}^x_{d,DTE}(W,t,\theta) = -\hat{\eta}_{s,d}(W,x,t,\theta), \tag{5.7}$$

$$\hat{\psi}_{d,QTE}(W,\tau,\theta) = \frac{\hat{\varphi}_d(W, \hat{q}_{d,\theta}(\tau), \theta)}{\hat{f}_{T_d}(\hat{q}_{d,\theta}(\tau), \theta)}, \qquad \hat{\psi}^x_{d,QTE}(W,\tau,\theta) = \frac{\hat{\eta}_{s,d}(W, x, \hat{q}^x_{d,\theta}(\tau), \theta)}{\hat{f}_{T_d,x}(\hat{q}^x_{d,\theta}(\tau), \theta)}, \tag{5.8}$$

$$\hat{\psi}_{d,CHTE}(W,t,\theta) = -\frac{\hat{\varphi}_d(W,t,\theta)}{\hat{s}_{T_d}(t,\theta)}, \qquad \hat{\psi}^x_{d,CHTE}(W,t,\theta) = -\frac{\hat{\eta}_{s,d}(W,x,t,\theta)}{\hat{s}_{T_d}(t, x\hat{\gamma}_d, \theta)}, \tag{5.9}$$

for $d \in \{0,1\}$. In (5.8), $\hat{f}_{T_d}(t,\theta)$ and $\hat{f}_{T_d,x}(t,\theta)$ are any first-stage estimators of $f_{T_d}(t,\theta) \equiv -\partial_y s_{T_d}(t,\theta)$ and $f_{T_d,x}(t,\theta) \equiv -\partial_y s_{T_d}(t, x\gamma_d, \theta)$, respectively, that are uniformly convergent as required in the assumption below.

Assumption 12 (First stage density estimator) *For $d \in \{0,1\}$, there exist first stage estimators $\hat{f}_{T_d}(t,\theta)$ and $\hat{f}_{T_d,x}(t,\theta)$ that are consistent for $f_{T_d}(t,\theta)$ and $f_{T_d,x}(t,\theta)$, respectively, uniformly over $\tilde{\mathcal{T}} \times \mathcal{X} \times \Theta$.*

In Appendix C.3, we describe an estimator of the conditional density and show that it fulfills the preceding assumption.

Corollary 3 *Suppose the assumptions of Theorem 4.2, Assumptions 11, and 12 hold, we get (i) $\hat{\mathbb{G}}^x_{\xi,j} \overset{p}{\underset{\xi}{\rightsquigarrow}} \nu'_{j,\mathbf{S}^x}(\mathbb{G}^x)$, and (ii) $\hat{\mathbb{G}}_{\xi,j} \overset{p}{\underset{\xi}{\rightsquigarrow}} \nu'_{j,\mathbf{S}}(\mathbb{G})$, for $j \in \{ATE, DTE, QTE, CHTE\}$.*

5.1 Bootstrap Confidence Bands

The functional central limit theorems for multiplier bootstrap established in the previous section can be used to conduct point-wise and uniform inference for the TEBF estimators. We provide an algorithm for constructing uniform confidence bands of the overall TEBF estimators in what follows. An analogous procedure that produces uniform confidence bands of the conditional TEBF estimators is given in Appendix A.3.2. Point-wise confidence intervals are by-products of these two algorithms.

Let $\hat{\mathbb{G}}_{lb,\xi,j}$ and $\hat{\mathbb{G}}_{ub,\xi,j}$ denote the first and second component of $\hat{\mathbb{G}}_{\xi,j}$, respectively.

Algorithm 1 (Uniform confidence sets of overall TEBFs)

1. Select a finite grid set $\mathcal{U}_m \equiv \{u_1, u_2, ..., u_m\}$ from $\mathcal{U}$, where the index u depends on the type of treatment effect under consideration. Pick a set $\mathbf{\Theta}_l \equiv \{\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_l\}$ with $\boldsymbol{\theta}_s = (\theta_{1,s}, \theta_{2,s}) \in \Theta^2$, and $\theta_{1,s} \leqslant \theta_{2,s}$, for all $s = 1, ..., l$.

   In Steps 2-5, the calculations will be performed for $d, r \in \{0, 1\}$, $t \in \tilde{\mathcal{T}}$, $\tau \in (0, \tau_o)$, $\boldsymbol{\theta} \in \mathbf{\Theta}_l$, and $u \in \mathcal{U}_m$.

2. Estimate $\hat{\gamma}_d$, $\hat{G}_{d,1}(t, x\hat{\gamma}_d)$, $\hat{G}_d(t, x\hat{\gamma}_d)$, and $\hat{s}_{T_d}(t, \theta)$. If $j = QTE$, compute $\hat{q}_{d,\theta}(\tau)$ and $\hat{f}_{T_d}(t, \theta)$.

3. Calculate $\hat{\nu}_j(u, \boldsymbol{\theta})$, $\hat{\varphi}_{d,1}(W, t, \theta)$, $\hat{\varphi}_{d,2}(W, t, \theta)$, and $\hat{\psi}_j(W, \theta)$.

4. Sample $\{\xi_i^b\}_{i=1}^n$ from a distribution with zero mean and unit variance, independently from data. Calculate $\hat{\boldsymbol{\psi}}_j^*$, and $\mathbb{G}_{\xi^b,j}(u, \theta)$.

   Repeat Step 4 for $b = 1, ..., B$, where B is some large integer.

5. For $\ell = lb, ub$, compute the $(1-\alpha)$-th quantile $\hat{c}^B_{n,\ell,j}(\alpha, \mathcal{U}_m, \mathbf{\Theta}_l)$ of $\left\{\max_{1\leqslant i\leqslant m, 1\leqslant s\leqslant l} \left\|\mathbb{G}_{\ell,\xi^b,j}(u_i, \boldsymbol{\theta}_s)\right\|\right\}_{b=1}^B$, and construct the uniform confidence band

$$C^B_{n,\ell,j}(1-\alpha, \mathcal{U}_m, \mathbf{\Theta}_l) \equiv \left\{\hat{\nu}_{\ell,j}(u, \boldsymbol{\theta}) \pm n^{-1/2} \hat{c}^B_{n,\ell,j}(\alpha, \mathcal{U}_m, \mathbf{\Theta}_l) : u \in \mathcal{U}_m, \boldsymbol{\theta} \in \mathbf{\Theta}_l\right\}.$$

The null hypothesis that a given TEBF is identically 0 over the index set $\mathcal{U}_m$ can be tested directly using the simulated bootstrap critical values. More generally, tests of FOSD relations, such as $\{\nu_{j,\ell}(u, \boldsymbol{\theta}) \leqslant 0 : u \in \mathcal{U}_m, \boldsymbol{\theta} \in \mathbf{\Theta}_l\}$, and tests of homogeneity, such as $\{\nu_{j,\ell}(u, \boldsymbol{\theta}) = \int_{\mathcal{U}_m} \nu_{j,\ell}(\tilde{u}, \boldsymbol{\theta}) d\tilde{u} : u \in \mathcal{U}_m, \boldsymbol{\theta} \in \mathbf{\Theta}_l\}$ can also be easily constructed using the simulated bootstrap process $\left\{\mathbb{G}_{\ell,\xi^b,j}\right\}_{b=1}^B$, which is a byproduct of Algorithm 1.

Note that $C^B_{n,\ell,j}(1-\alpha, \mathcal{U}_m, \mathbf{\Theta}_l)$ are uniform across both u and θ. Pointwise confidence sets are immediately available from the two procedures by setting $\mathcal{U}_m = \{u^*\}$, $\mathbf{\Theta}_l = \{\boldsymbol{\theta}^*\}$, for some u^* and $\boldsymbol{\theta}^*$ specified by the researcher.

We denote the confidence set generated by Algorithm 2 as $C^{x,B}_{n,\ell,j}(1-\alpha, \mathcal{U}_m, \mathbf{\Theta}_l)$. The next theorem confirms that the uniform bootstrap confidence bands for both conditional and overall TEBFs are asymptotically accurate.

Theorem 5.2 *Suppose the assumptions of Corollary 3 hold, we have*

$$\lim_{n\to\infty} \inf_{(u,\boldsymbol{\theta})\in\mathcal{U}_m\times\mathbf{\Theta}_l} \mathbb{P}\left(\nu^x_{\ell,j}(u, \boldsymbol{\theta}) \in C^{x,B}_{n,\ell,j}(1-\alpha, \mathcal{U}_m, \mathbf{\Theta}_l)\right) = 1-\alpha,$$

$$\lim_{n\to\infty} \inf_{(u,\boldsymbol{\theta})\in\mathcal{U}_m\times\boldsymbol{\Theta}_l} \mathbb{P}\left(\nu_{\ell,j}(u,\boldsymbol{\theta}) \in C^B_{n,\ell,j}(1-\alpha,\mathcal{U}_m,\boldsymbol{\Theta}_l)\right) = 1-\alpha,$$

for $x \in \mathcal{X}$, $\ell \in \{lb, ub\}$, *and* $j \in \{ATE, DTE, QTE, CHTE\}$.

6 Monte Carlo Study

Results in the previous section imply that the estimators and uniform confidence bands for conditional and overall TEBFs will exhibit desirable properties when sample size is sufficiently large. What of their small-sample performance? We conduct a small scale Monte Carlo experiment to address this question. The DGP for the simulations consists of the following four aspects:

1. The conditional survival functions: for $d \in \{0,1\}$, both T_d and C_d follow the conditional Exponential distribution. Specifically, $S_{\ell_d|X}(t|x) = \exp\left(-\lambda_{\ell,d}(x\gamma_d)t\right)$, where the hazard rate parameter $\lambda_{\ell,d}(x\gamma_d) > 0$, for $\ell \in \{T, C\}$ and $d \in \{0,1\}$. Since the shape parameters are constrained to be positive, we assume both $\beta_{\varpi,d,\ell} > 0$ and $\inf\{\mathcal{X}\gamma_d\} > 0$.

2. The copula function: the true copula is assumed to belong to the Archimedean family with the Gumbel generator function: $\phi_{\theta_0(d)}(u) = \left(\log u^{-1}\right)^{\theta_0(d)}$ and $\mathcal{C}(u,v;\theta_0(d)) = \exp\left(-\left[(\log u^{-1})^{\theta_0(d)} + (\log v^{-1})^{\theta_0(d)}\right]^{1/\theta_0(d)}\right)$, where $\theta_0(d) \in [\underline{\theta}, \bar{\theta}] \subset [1,\infty)$, for $d \in \{0,1\}$.

3. The covariates: $X = (X_1, X_2, X_3, X_4)$, where X_1 and X_2 are drawn from Uniform$[0,1]$, and the remaining two, X_3 and X_4 are binary variables, following Bernoulli(0.5). The four variables are mutually independent.

4. The treatment assignment mechanism: the treatment status D is determined by $D = \mathbb{1}\{p(X) > U\}$, where U follows Uniform[0, 1], and

$$p(X) = \frac{\exp(-0.25X_1 + 0.25X_2 + 0.1X_3 - 0.1X_4)}{1 + \exp(-0.25X_1 + 0.25X_2 + 0.1X_3 - 0.1X_4)},$$

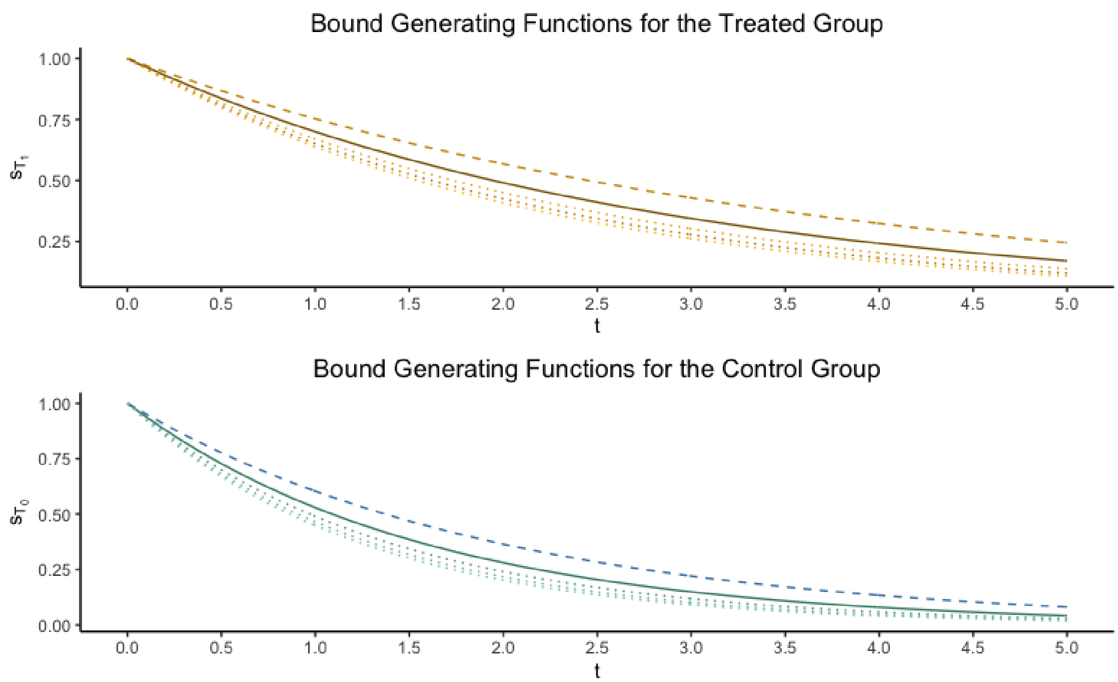
is the true propensity score function.

The BGF admits a analytical form when $\lambda_{T,d}(\cdot) = \lambda_{C,d}(\cdot) \equiv \lambda_d(\cdot)$. Such a simplification is handy when checking the coverage of our bootstrap confidence sets. By symmetry, we have that $S_{Y_d|X}(y|x) = \exp\left(-2^{1/\theta_0(d)}\lambda_d(x\gamma_d)y\right)$, and $S_{Y_d,R_d|X}(y,1|x) = 2^{-1}\exp\left(-2^{1/\theta_0(d)}\lambda_d(x\gamma_d)y\right)$, implying a population censoring rate of 50%. Now, from (4.5) and by direct calculations,

$$s_{T_d}(t, x\gamma_d, \theta) = \phi_\theta^{-1}\left(2^{-1}\phi_\theta\left(\exp\left(-2^{1/\theta_0(d)}\lambda_d(x\gamma_d)t\right)\right)\right).$$

The above formula further simplifies when the true copula is Gumbel, in which case, $s_{T_d}(t, x\gamma_d, \theta) = \exp\left(-2^{1/\theta_0(d)-1/\theta}\lambda_d(x\gamma_d)t\right)$, equivalent to an exponential distribution with the rate parameter equal to $\beta_d(x\gamma_d, \theta) \equiv 2^{1/\theta_0(d)-1/\theta}\lambda_d(x\gamma_d)$. As a direct consequence, $\nu^x_{ATE,lb}(\boldsymbol{\theta}) = \beta_1(x\gamma_1, \theta_2)^{-1} - \beta_0(x\gamma_0, \theta_1)^{-1}$, and the overall average effect is also immediately available via taking expectation with respect to X. In what follows, we set $\gamma_1 = \gamma_0 = (0.5, 0.25, 0.25)'$, $\lambda_0(v) = v^{-0.5}$, and $\lambda_1(v) = (v^{0.5} + v)^{-1}$. To generate variables from the Gumbel copula, we follow the algorithm provided in Section 2.9 in Nelsen (2007), for which purpose, we assume the true copula parameters are $\theta_0(1) = \theta_0(0) = 1.5$.

Figure 1: Bound generating functions with multiple levels of θ



Notes: The top panel: Unconditional BGFs for the treated group. The bottom panel: unconditional BGFs for the control group. In each plot, the dashed curve represents BGF with independence copula. The true potential survival functions are plotted with the solid curve. The dotted curves correspond, from top to bottom, to the cases when $\theta = 2, 2.5,$ and 3.

In Figure 1, we plot unconditional BGFs with various levels of the sensitivity parameter, along with the true potential survival functions of T_1 and T_0. The figure illustrates the stochastic dominance relations induced by the concordance ordering of the copula family. It also demonstrates that erroneously imposing independence may lead to substantial bias.

To assess the performance of TEBF estimators over a index set $\mathcal{U}_m$, we adopt *average* and *median integrated bias*, *integrated root mean square error* (IRMSE), and the coverage rate as the criterion of evaluation.[9] Regarding the index set $\mathcal{U}_m$, we use an equidistant grid between 0 and 8

9 Consider a Monte Carlo experiment with S replications, the average integrated bias is defined by $S^{-1}\sum_{s=1}^{S}\int_{u\in\mathcal{U}_m}\left|\hat{f}_s(u) - f(u)\right|du$, median integrated bias denotes the 50-th percentile of $\left\{\int_{u\in\mathcal{U}_m}\left|\hat{f}_s(u) - f(u)\right|du\right\}_{s=1}^{S}$, and the IRMSE, by $\left(S^{-1}\sum_{s=1}^{S}\int_{u\in\mathcal{U}_m}\left|\hat{f}_s(u) - f(u)\right|^2du\right)^{1/2}$, where $f(\cdot)$ is

with the interval size of 0.2 for the DTE and CHTE. For the QTE, an equidistant grid between 0.2 and 0.8 with a step size of 0.02 is adopted. Additionally, we truncate ATE at $y_o = 9$, which is approximately equal to 99th percentile of $F_{T_0}(\cdot)$. We let both $L(\cdot)$ and $K(\cdot)$ be the Epanechnikov kernel: $L(u) = K(u) = 0.75(1-u^2)\mathbb{1}\{|u| \leqslant 1\}$. The bandwidth b is chosen as the value from the set $\{0.04, 0.06, ..., 0.48, 0.5\}$ that minimizes the estimated criterion $\sum_{d\in\{0,1\}} \hat{\mathcal{J}}_d(\hat{\gamma}_d, \rho)$, where we set $\rho(v) = \exp(-\|v\|^2/2)$. We set $h = b$. Following Remark 1, we evaluate the reliability of the single-index assumption using the specification test of Maistre and Patilea (2019). The null hypothesis that index sufficiency holds is not rejected for both γ_1 and γ_0 at the 10% level.

Table 1 reports simulation results based on 1,000 Monte Carlo replications of samples with size $n = 1,000$. For each type of treatment effect, we show results for two different range of θ: a narrow one with $\boldsymbol{\theta} = (5/4, 5/3)$, and a wider one with $\boldsymbol{\theta} = (1, 2)$. Given the one-to-one mapping between θ and Kendall's τ, the two $\boldsymbol{\theta}$ choices correspond to the latter lying between $[0.2, 0.4]$ and $[0, 0.5]$, respectively. Results from Table 1 indicate that estimators for the ATE and DTE perform relatively better the other two types, in terms of integrated bias, IRMSE, and coverage probability. This is to be expected as nonlinear transformations tend to induce higher bias. Allowing the bandwidth h to vary according to the type of treatment effect under consideration may alleviate this bias issue. This is left for future work. The empirical coverage rates are computed based on Algorithms 1 and 2 with 2,000 multiplier bootstrap replications. According to the table, our bootstrap confidence sets generally achieve close-to-nominal-level coverage, regardless of the type of effect and whether the effect is conditional on a fixed x. Overall, the finite-sample results are consistent with theoretical predictions from Sections 4 and 5.

7 Empirical Illustration

In this section, we revisit Bernasconi *et al.* (2022) on the effect of acute lymphoblastic leukaemia treatment where survival time is subject to dependent censoring caused by the abandonment of treatment. ALL is a major source of cancer diagnosis among people under 18 years old, accounting for nearly 25% of all cancer diagnoses (Howlader, Noone *et al.*, 2016). Wide disparities in the cure rate has been documented between high-income (~80%) and mid-and-low-income (~35%) countries (Gatta, Capocaccia *et al.*, 2005; Howard and Wilimas, 2005). Abandonment of treatment is seen as a major factor for such disparities (Mostert, Arora *et al.*, 2011). Decision to withdraw treatment can be affected by various factors including distance to the treatment facility, family economic status, and personal beliefs, many of which also have an impact on the quality of treatment. As such, the independent censoring assumption is not appropriate in this context.

---

any of the $\nu_{j,\ell}(\cdot)$, for $\ell = lb, ub$, and $j \in \{DTE, QTE, CHTE\}$. When $j = ATE$, these criterion reduce to the usual ones.

(a)	Lower Conditional TEBF				Upper Conditional TEBF			
	Avg. Bias	Med. Bias	RMSE	Cvg. Rate	Avg. Bias	Med. Bias	RMSE	Cvg. Rate
ATE	0.028	0.020	0.149	0.943	-0.019	-0.026	0.141	0.949
DTE	0.050	0.048	0.052	0.940	0.052	0.050	0.054	0.944
QTE	0.493	0.442	0.562	0.932	0.600	0.558	0.673	0.935
CHTE	0.799	0.713	0.910	0.930	0.784	0.678	0.906	0.930
(b)	Lower Conditional TEBF				Upper Conditional TEBF			
	Avg. Bias	Med. Bias	RMSE	Cvg. Rate	Avg. Bias	Med. Bias	RMSE	Cvg. Rate
ATE	0.044	0.047	0.153	0.940	-0.062	-0.074	0.149	0.942
DTE	0.048	0.045	0.050	0.941	0.053	0.051	0.056	0.939
QTE	0.459	0.406	0.523	0.934	0.534	0.462	0.600	0.933
CHTE	0.742	0.552	1.082	0.927	0.797	0.713	1.277	0.932
(c)	Lower Overall TEBF				Upper Overall TEBF			
	Avg. Bias	Med. Bias	RMSE	Cvg. Rate	Avg. Bias	Med. Bias	RMSE	Cvg. Rate
ATE	0.107	0.105	0.147	0.938	0.076	0.073	0.121	0.941
DTE	0.059	0.055	0.063	0.944	0.062	0.059	0.067	0.943
QTE	0.521	0.448	0.593	0.934	0.566	0.513	0.641	0.932
CHTE	0.679	0.616	0.723	0.932	0.606	0.553	0.648	0.931
(d)	Lower Overall TEBF				Upper Overall TEBF			
	Avg. Bias	Med. Bias	RMSE	Cvg. Rate	Avg. Bias	Med. Bias	RMSE	Cvg. Rate
ATE	0.114	0.107	0.130	0.939	0.101	0.105	0.152	0.94
DTE	0.051	0.051	0.053	0.943	0.059	0.057	0.062	0.941
QTE	0.457	0.432	0.504	0.932	0.565	0.517	0.624	0.929
CHTE	0.730	0.685	0.785	0.925	0.544	0.519	0.583	0.93

Note: Simulations based on 1,000 Monte Carlo experiments with samples of size $n = 1,000$. Panels (a) and (c) presents results for conditional and overall TEBF with $(\theta_1, \theta_2) = (5/4, 5/3)$. The numbers in Panels (b) and (d) are associated with $(\theta_1, \theta_2) = (1, 2)$. For all TEBFs except the ATE, the statistics are computed for the MAE with respect to index sets described in the main text. "Avg. Bias", "Med. Bias", "IRMSE", and "Cvg. Rate", stand for the average integrated bias, median integrated bias, integrated root mean squared errors, and 95% coverage probability respectively. Coverage probability is computed based on bootstrap confidence sets produced with 2,000 multiplier bootstrap replications.

Table 1: Monte Carlo results for the conditional and overall TEBFs

The data comes from two subsequent clinical studies conducted in Honduras between 2000 and 2015. During the period of 2000 - 2007, a protocol called GHS-2000 were adopted to treat the ALL patients. In the second period (2008 - 2015), the treatment follows a new protocol denominated AHOPCA ALL-2008.[10] We view GHS-2000 as the control group $(D = 0)$ and AHOPCA ALL-2008 as the treated group $(D = 1)$. Treatment effects in this context translate to comparative effectiveness of the two protocols. The outcome of interest T is formally defined as the time since treatment to the first event among relapse, resistance to treatment, secondary malignant neoplasm, and death. The outcome is subject to both administrative censoring, which is independent of the EFS time, and endogenous censoring in the form of abandonment of treatment. Following Bernasconi *et al.* (2022), we combine the two types of censoring into a composite variable C, and we use potential *event-free-survival* (EFS) to denote S_{T_d} under protocol $d \in \{0, 1\}$.

The baseline covariates X include biological characteristics such as gender, age, white blood cell count, central nervous system involvement (CNS), cancer histology, and socio-economic factors: family unity, living conditions, home phone ownership, and distance to the hospital. Table 2 summarizes these characteristics for patients undergone the each of the two protocols. Instead of performing multiple imputations as in Bernasconi *et al.* (2022), missing cases are removed. Results from the table shows that patients from the AHOPCA ALL-2008 group are less likely to withdraw treatment, to live in a rural neighborhood, and live closer to the clinic. To account for these imbalances across the treatment groups, Bernasconi *et al.* (2022) relies on an *inverse probability of treatment and censoring* weighting strategy, which further depends on the assumption that, conditionally on the baseline covariates, the potential EFS and the abandonment are mutually independent. Such a restriction, however, is not necessary for our proposed methodology.

We discuss the estimation results in what follows. The estimation procedure follows closely that employed in Section 6. To assess the validity of the index sufficiency assumption, we conduct specification tests and find that the nulls are not rejected at 10% level for both treatment groups, indicating that our SICG estimator can be applied to this context. To estimate the BGFs, we assume that the true copula belongs to the Gumbel family and let the sensitivity parameter θ vary between 1 and 2.

Figure 2 plots estimated potential EFS curves along with bounds for the DTE estimates. Panels (a) and (b) depict the unconditional results, whereas Panels (c) and (d) show estimates when each component of X is fixed at its sample mean. The solid and dashed curves in Panels (a) and (c) are calculated, using (4.5) with the independence copula. The shaded areas correspond to the bounds of EFS when the true copula is Gumbel, and the index parameter $\boldsymbol{\theta} = (1, 2)$. The main finding of Bernasconi *et al.* (2022) is that the AHOPCA ALL-2008 protocol yields better potential EFS in the

10 Details of these two protocols can be found in Marjerrison, Antillon *et al.* (2013) and Navarrete, Rossi *et al.* (2014)

	GHS-2000 (No. Obs. = 514)					AHOPCA ALL-2008 (No. Obs. = 536)				
Statistics	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)	Mean	St. Dev.	Pctl(25)	Median	Pctl(75)
EFS Time	4	3.84	0.47	2.73	7.59	2.29	2.05	0.71	1.55	3.47
Abandonments	21%	0.41	0	0	0	15%	0.35	0	0	0
Age	7.67	4.36	3.79	6.79	11.26	7.67	4.94	3.45	6.01	11.82
White Blood Cell Count	4.64	10.01	0.45	1.13	4.24	4.28	8.52	0.53	1.11	4.07
Time to Hospital	4.05	2.85	1.5	4	6	2.87	2.03	1	3	4
Male	61%	0.49	0	1	1	53%	0.5	0	1	1
CNS	6%	0.23	0	0	0	13%	0.34	0	0	0
Linage	92%	0.27	1	1	1	93%	0.26	1	1	1
Living Condition	62%	0.49	0	1	1	46%	0.5	0	0	1
Family Type	45%	0.5	0	0	1	14%	0.35	0	0	0
Phone at Home	36%	0.48	0	0	1	42%	0.49	0	0	1

Note: Summary statistics for two protocols for ALL treatment. The left panel describes the GHS-2000 group (2000 - 2007). The right panel is for the AHOPCA ALL-2008 group (2008-2015). "MALE", "CNS", "Lineage", "Living Condition", "Family Type", "Phone at Home" are dummy variables denoting whether the subject is male, whether the central nervous system is involved, the type of tumor lineage, whether the patient lives in an urban neighborhood, lives in a united family, and owns a home phone.

Table 2: Summary Statistics

(a) Lower Bound	ATE	DTE	QTE	CHTE
Overall	-0.395 [-0.725,-0.065]	-0.048 [-0.169, 0.074]	-0.429 [-1.308,0.451]	-0.139 [-0.657,0.379]
Conditional	-0.532 [-0.916,-0.148]	-0.038 [-0.176,0.101]	-0.629 [-1.559,0.301]	-0.092 [-0.683, 0.499]

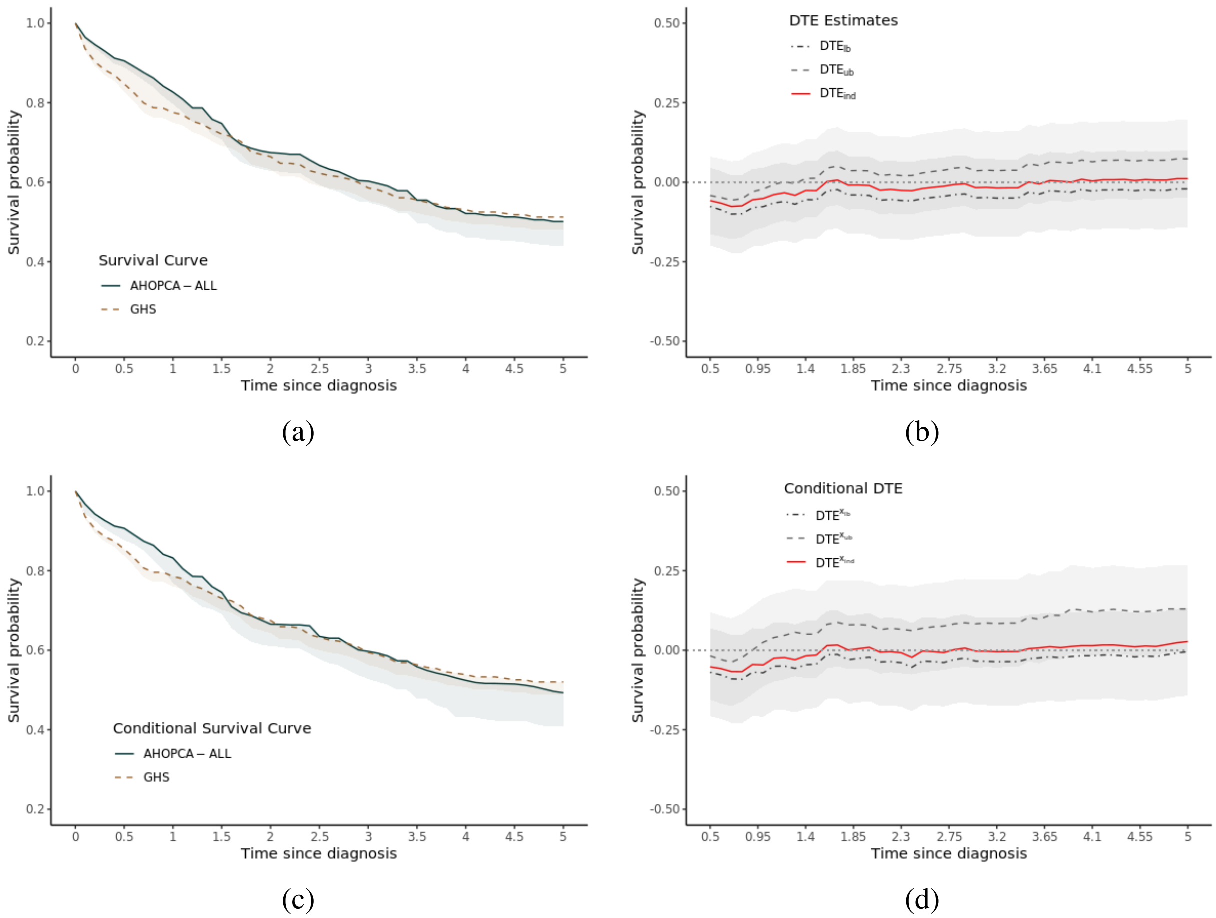
(b) Upper Bound	ATE	DTE	QTE	CHTE
Overall	0.420 [0.089,0.750]	0.040 [-0.082,0.162]	0.543 [-0.337,1.422]	0.156 [-0.362,0.674]
Conditional	0.281 [-0.103,0.665]	0.079 [-0.059,0.217]	0.386 [-0.544,1.316]	0.208 [-0.382,0.799]

Notes: The means of estimated TEBFs over their respective index sets ($\mathcal{U}_m = \{0.5, 0.6, ..., 4.9, 5.0\}$ for the DTE and CHTE, and $\mathcal{U}_m = \{0.2, 0.22, ..., 0.8\}$ for the QTE), are shown in the first row of each panel. Numbers in square brackets are the corresponding 95% bootstrap confidence intervals, based on 2,000 bootstrap replications. They are calculated following Algorithms 1 and 2, for the overall and conditional cases, respectively.

Table 3: Estimation results for conditional and overall TEBFs

first three years and the difference tapers off in the long term. A similar pattern holds here under random censoring. However, these findings do not appear generalizable to other levels of dependent censoring. For instance, the potential EFS is nearly 10% lower under the newer treatment when $\theta = 2$.

Figure 2: Estimates of the potential EFS and the DTE



Notes: In Panels (a) and (c), the solid and the dashed curves represent SICG estimates for the AHOPCA-ALL and GHS protocols, respectively, with the independence copula. The shaded areas are bounded from above and below by SICG estimates with Gumbel copula parameters $\theta_1 = 1$ and $\theta_2 = 2$, respectively. Panels (b) and (d) depict DTE estimates along with their 95% uniform confidence bands, over $\mathcal{T}_m$. The red lines denote the DTE estimates with the independence copula. The confidence bands in Panels (b) and (d) are calculated following the bootstrap procedures in Algorithms 1 and 2, respectively, with 2,000 bootstrap replications. Finally, estimates in Panels (c) and (d) are conditional on X at it sample average.

The red lines of Panels (b) and (d) of Figure 2 represent the DTE estimates over an equidistant grid $\mathcal{T}_m \equiv \{0.5, 0.6, ..., 4.9, 5.0\}$ with $\boldsymbol{\theta} = (1, 1)$, i.e. under the independent censoring. We spot a similar pattern as the original finding by Bernasconi *et al.* (2022). AHOPCA ALL-2008 outperforms GHS-2000 in the initial years post treatment. Would this result carry over to the case of dependent censoring? The question can be addressed using the estimated lower and upper bounds of the DTEs, along with their uniform 95% bootstrap confidence bands, depicted in the two panels. These results are calculated under the maintained range $\boldsymbol{\theta} = (1, 2)$. Judging from the uniform confidence bands for both bounds, we cannot draw the conclusion that the newer protocol tends to fare better in the early years post treatment. That being said, the current conclusion is valid, only under the maintained assumptions on θ. In this case, we are focusing on a mild positive dependence between T and C with Kendall's $\tau \in [0, 0.5]$. Generalizations to other dependence patterns require similar robustness checks with corresponding levels of θ.

In Table 3, a similar finding is recorded for other types of treatment effects. We report the average of the estimated TEBFs over an index set $\mathcal{U}_m$,[11] when $\boldsymbol{\theta} = (1, 2)$. Also shown in the table are 95% uniform confidence sets, evaluated at the mean (over $\mathcal{U}_m$) of these TEBF estimates. Although we cannot conclude from Table 3 that treatment effect nullity holds uniformly over the index set, and over the entire Gumbel copula family, it does seem to suggest that there do not exist non-zero differences between the two protocols on average, regardless of the type of policy effect under consideration.

Overall, when we deviate from the conditional independence censoring mechanism, we do not find enough evidence supporting that AHOPCA ALL-2008 leads to more favorable early-year survival prospects.

8 Conclusion

In this paper, we proposed a framework for conducting sensitivity analysis on various treatment effect parameters when the latent duration is subject to dependent censoring. In order to obtain bounds of policy effects, we first derived bounds on the distribution of the potential outcome. Such bounds follow naturally from the concordance ordering we imposed on the Archimedean copulas. Moreover, we embedded a single-index structure into our identification framework, as an attempt to curb the "curse of dimensionality" and to make our method practically feasible. Given these results, we then proposed estimation procedures and established asymptotic properties of the resulting semiparametric estimators.

To conduct uniformly valid inference, we proposed easy-to-implement multiplier bootstrap procedures, and showed the uniform confidence sets thus constructed are asymptotically accurate. Monte Carlo simulations confirm our theoretical findings. Applying our methodology to real data, we revisited Bernasconi *et al.* (2022). Under a conditional independence assumption, and when the early-year survival prospect is concerned, they conclude in favor of AHOPCA ALL-200. Our sensitivity analysis demonstrates that such a conclusion may not continue to hold when we depart from the random censoring,

We have limited discussion to the classical unconfoundedness design in our main text, but the proposed methodology can be applied to other policy setups, such as the local treatment effect framework, as considered by Imbens and Angrist (1994), Angrist, Imbens, and Rubin (1996), Abadie (2003) and Frölich and Melly (2013); difference-in-differences models, cf. Card and

11 For the DTE and CHTE, $\mathcal{U}_m = \mathcal{T}_m$, and for the QTE, $\mathcal{U}_m = \{0.2, 0.21, ..., 0.79, 0.8\}$.

Krueger (1994), Heckman, Ichimura, and Todd (1997), Abadie (2005), and Athey and Imbens (2006); and marginal treatment effect setup, such as Heckman and Vytlacil (2001), and Heckman and Vytlacil (2005). Upon appropriate modifications of the identification assumptions, BGFs and TEBFs can be easily derived along similar lines of Theorem 3.2 and Proposition 1, based on which, many policy relevant questions can be addressed.

Supplementary Appendix

This supplemental appendix contains proofs of the main theorems, auxiliary lemmas, and results. Appendix A collects the proofs of the main results of the paper. Appendix B introduce additional results on the single-index estimator, and Appendix C presents auxiliary results.

Appendix A Proofs of Main Results

A.1 Proof of Results from Section 3

A.1.1 Identification of Single-index Parameters

Proof of Lemma 3.1. The proof is based on the so-called decomposition and contraction relationships of the Graphoid axioms by Dawid (1979):

$$\text{Decomposition: } A \perp\!\!\!\perp (B, C)|D \text{ implies that } A \perp\!\!\!\perp B|D,$$
$$\text{Contraction: } A \perp\!\!\!\perp B|D \text{ and } A \perp\!\!\!\perp C|(B, D) \text{ implies that } A \perp\!\!\!\perp (B, C)|D,$$

for generic random variables A, B, C, D. For each $d = 0, 1$, $(T_d, C_d) \perp\!\!\!\perp D|X$, under Assumption 1. Since the sigma field generated by $X\gamma_d$ is a subset of that generated by X, we have $(T_d, C_d) \perp\!\!\!\perp D|X, X\gamma_d$. Together with the index sufficiency condition, $(T_d, C_d) \perp\!\!\!\perp X|X\gamma_d$, we deduce from the contraction relationship that $(T_d, C_d) \perp\!\!\!\perp (D, X)|X\gamma_d$, which implies that $(T_d, C_d) \perp\!\!\!\perp D|X\gamma_d$ by the decomposition relationship. Since Y_d and R_d are deterministic functions of (T_d, C_d), the desired result follows. ■

Proof of theorem 3.1. For the first half of part (i), note that for any $(t, x, d, r) \in \mathcal{T} \times \mathcal{X} \times \{0, 1\}^2$,

$$\mathbb{E}[\mathbb{1}\{D = d, R = r, Y \leqslant t\} |X] = F_{Y,R|D,X}(t, r|d, X)\, \mathbb{E}[\mathbb{1}\{D = d\} |X],$$

where $F_{Y,R|D,X}(t|d, X) = F_{Y_d,R_d|D,X}(t, r|d, X) = F_{Y_d,R_d|X}(t, r|X) = F_{Y_d,R_d|X\gamma_d}(t, r|X\gamma_d)$, almost surely. The second equality follows under Assumption 1, and the last holds under Assumption 2. On the other hand,

$$\mathbb{E}[G_{d,r}(t, X\gamma_d)|X] = G_{d,r}(t, X\gamma_d) = F_{Y_d,R_d|D,X\gamma_d}(t, r|d, X\gamma_d) = F_{Y_d,R_d|X\gamma_d}(t, r|X\gamma_d),$$

almost surely. The third equality is due to Lemma 3.1. Thus, (3.3) holds almost surely.

The converse part can be established by contradiction, applying similar arguments as in the proof of Theorem 4.1 in Ichimura (1993). Suppose there exists $d \in \{0, 1\}$ and $\gamma^* \in \Gamma$, such that $\gamma^* \neq \gamma_d$, and $\mathbb{E}[U_{\gamma^*}(t,d,r)|X] = 0$ almost everywhere for $(t,r) \in \mathcal{T}_0 \times \{0,1\}^2$. Note that under Assumption 2, it continues to hold that $\mathbb{E}[\mathbb{1}\{D=d, R=r, Y \leqslant t\}|X] = F_{Y_d,R_d|X\gamma_d}(t,r|X\gamma_d)\,\mathbb{E}[\mathbb{1}\{D=d\}|X]$ almost surely.

For any $x \in \mathcal{X}_0$, let $v = x\gamma^*$ and $\bar{\gamma} = \gamma_d - \gamma^*$, we have

$$F_{Y_d,R_d|X\gamma_d}(t,r|x\gamma_d) = F_{Y_d,R_d|X\gamma_d}\left(t,r|v + \sum_{\ell=1}^{k-1}\bar{\gamma}_\ell x_{[\ell+1]}\right) = G_{d,r}(t,v),$$

where the second equality holds almost surely. Fix v and take partial derivative of the middle term in the above display with respect to $\{x_{[\ell]}\}_{\ell=2}^{k_1}$. It follows that $\partial_{x\gamma}F_{Y_d,R_d|X\gamma_d}(t,r|x\gamma_d)\bar{\gamma}_{[\ell]} = 0$, for $\ell = 1,...,k_1-1$. Recall our assumption on $\mathcal{X}_0$, $\partial_{x\gamma}F_{Y_d,R_d|X\gamma_d}(t,r|X\gamma_d) \neq 0$ with positive probability. Consequently, $\gamma^*_\ell = \gamma_{d,\ell}$, for $\ell = 1,...,k_1-1$.

Under Assumption 3.3 with $\gamma = \gamma_d$, there exists an open interval $\mathcal{V}_0$ such that for all $v \in \mathcal{V}_0$,

$$F_{Y_d,R_d|X\gamma_d}(t,r|v) = F_{Y_d,R_d|X\gamma_d}(t,r|v+\bar{\gamma}_\ell) = G_{d,r}(t,v),$$

where $\ell = k_1,...,k-1$. In view of the first equality, we find from Assumption 3.4 (ii) that $\bar{\gamma}_\ell = 0$, and thus, $\gamma^*_\ell = \gamma_{d,\ell}$, for $\ell = k_1,...,k-1$. This contradicts the supposition that $\gamma^* \neq \gamma_d$. Hence, part (ii) follows.

To show part (iii), we first note that $\mathcal{J}_d(\gamma;\vartheta) \geqslant$ due to its construction. Next, when $\gamma \neq \gamma_d$,

$$\mathcal{J}_d(\gamma;\vartheta) \geqslant \int_{\mathcal{T}_0\times\{0,1\}}\int_{z\in\mathcal{Z}} \|\mathbb{E}[U_{d,\gamma,\ell}(t,r)\vartheta(X;z)]\|^2\, d\Pi_Z(z)d\Pi_{T,R}(t,r) > 0$$

where the second inequality follows because ϑ is chosen such that the equivalence in (3.1) holds, and from part (ii), we have $\mathbb{E}[U_{d,\gamma,\ell}(t,r)|X] > 0$, $\forall(d,r,t) \in \{0,1\}^2 \times \mathcal{T}_0$. ∎

A.1.2 Identification of the Treatment Effects

Proof of Theorem 3.2. Proof of the first half of part (i) is a slight modification of Lemma 1 of Braekers and Veraverbeke (2005). If the true copula is index by θ^*,

$$\begin{aligned}\partial_y S_{Y_d,R|X}(y,1|x) &= \partial_y S_{T_d,C_d|X}(t,c|x)|_{t=c=y} = \partial_y \mathcal{C}(S_{T_d|X}(t|x), S_{C_d|X}(c|x);\theta^*)|_{t=c=y} \\ &= -\frac{\phi'_{\theta*}(S_{T_d|X}(y|x))S'_{T_d|X}(y|x)}{\phi'_{\theta*}(S_{Y_d|X}(y|x))}.\end{aligned}$$

The first equality is due to Tsiatis (1975), the second is by Assumption 4, and the last is due to the construction of the Archimedean copula. Multiplying both sides by $\phi'_{\theta*}(S_{Y_d|X}(y|x))$ and integrating with respect to y over $[0, t]$ gives

$$\int_0^t \phi_{\theta*}^{'}(S_{Y_d|X}(y|x))S_{Y_d,R|X}(dy, 1|x) = \phi_{\theta*}\left(S_{T_d}(t|x)\right). \tag{A.1}$$

Note that

$$\begin{aligned} s_d(y, x\gamma_d) &= \mathbb{E}\left[\mathbb{E}[D\mathbb{1}\{Y > y\}|D = d, X]|D = d, X\gamma_d = x\gamma_d\right] \\ &= \mathbb{E}\left[\mathbb{E}[\mathbb{1}\{Y_d > y\}|X]|D = d, X\gamma_d = x\gamma_d\right] \\ &= \mathbb{E}\left[S_{Y_d|X}(y|X)|D = d, X\gamma_d = x\gamma_d\right] \\ &= \mathbb{E}\left[S_{Y_d|X\gamma_d}(y|X\gamma_d)|D = d, X\gamma_d = x\gamma_d\right] \\ &= S_{Y_d|X\gamma_d}(y|x\gamma_d) = S_{Y_d|X}(y|x), \end{aligned}$$

where the first equality follows by the tower property of conditional expectation. The second is by Assumption 1, and the third is due to Assumption 2. The last one holds under Assumption 2. Same lines of arguments leads to $s_{d,1}(y, x\gamma_d) = S_{Y_d,R|X}(y, 1|x)$. Substituting these equations into (A.1) and letting θ^* range over Θ yields the desired result. When the true copula is invariant to x, θ does not depend on x, and therefore, the second half follows by taking expectation of $s_{T_d}(\cdot, X\gamma_d, \theta)$ with respect to X.

Part (ii) follows directly from the first half of part (i) and Proposition 2 of Rivest and Wells (2001), and therefore, the proof is omitted. ■

Proof of Proposition 1. Under Assumption 1 - 3, γ is identified by Theorem 3.1. For $\theta \in [\theta_1, \theta_2]$, where $(\theta_1, \theta_2) \in \Theta^2$, $s_{T_d}(\cdot, \theta)$ then belongs to the identified set of the potential causal curve $S_{T_d}(\cdot)$ by Theorem 3.2. The identified sets for ν_j are then derived using the fact that treatment responses respect the stochastic dominance relations of S_{T_d} for each $j \in \{ATE, DTE, QTE, CHTE\}$. Sharpness is inherited from that of the potential causal curves. Results for the conditional treatment effects can be shown with similar arguments. ■

A.2 Proof of Results from Section 4

A.2.1 Uniform Linear Representation

Proof of Theorem 4.1. The proof is similar to that of Theorem 4.1 in Fan and Liu (2018), with substantial differences due to the use of single-index estimator. We first provide a uniform linear expansion for $\phi_\theta\left(\hat{s}_{T_d}(t, x\hat{\gamma}_d, \theta)\right) - \phi_\theta\left(s_{T_d}(t, x\gamma_d, \theta)\right)$, the desired result then follows by a second order Taylor expansion of ϕ_θ^{-1}. We let $s_{d,r} \equiv 1/2 - G_{d,r}$ and its estimator $\hat{s}_{d,r} \equiv 1/2 - \hat{G}_{d,r}$ for

$r = 0, 1$. Observe that

$$
\begin{aligned}
&\phi_\theta\left(\hat{s}_{T_d}(t, x\hat{\gamma}_d, \theta)\right) - \phi_\theta\left(s_{T_d}(t, x\gamma_d, \theta)\right) &\text{(A.2)}\\
=&\int_0^t \phi'_\theta\left(\hat{s}_d(y, x\hat{\gamma}_d)\right)\hat{s}_{d,1}(dy, x\hat{\gamma}_d) - \int_0^t \phi'_\theta\left(s_d(y, x\gamma_d)\right)s_{d,1}(dy, x\gamma_d)\\
=&\int_0^t \left\{\phi'_\theta\left(\hat{s}_d(y, x\hat{\gamma}_d)\right) - \phi'_\theta\left(s_d(y, x\gamma_d)\right)\right\}s_{d,1}(dy, x\gamma_d)\\
&+ \int_0^t \phi'_\theta\left(s_d(y, x\gamma_d)\right)\left\{\hat{s}_{d,1}(dy, x\hat{\gamma}_d) - s_{d,1}(dy, x\gamma_d)\right\}\\
&+ \int_0^t \left\{\phi'_\theta\left(\hat{s}_d(y, x\hat{\gamma}_d)\right) - \phi'_\theta\left(s_d(y, x\gamma_d)\right)\right\}\left\{\hat{s}_{d,1}(dy, x\hat{\gamma}_d) - s_{d,1}(dy, x\gamma_d)\right\}\\
=&\int_0^t \phi''_\theta\left(s_d(y, x\gamma_d)\right)\left\{\hat{s}_d(y, x\hat{\gamma}_d) - s_d(y, x\gamma_d)\right\}s_{d,1}(dy, x\gamma_d) &\text{(A.3)}\\
&+ \phi'_\theta\left(s_d(t, x\gamma_d)\right)\left\{\hat{s}_{d,1}(t, x\hat{\gamma}_d) - s_{d,1}(t, x\gamma_d)\right\} &\text{(A.4)}\\
&- \int_0^t \phi''_\theta\left(s_d(y, x\gamma_d)\right)\left\{\hat{s}_{d,1}(y, x\hat{\gamma}_d) - s_{d,1}(y, x\gamma_d)\right\}s_d(dy, x\gamma_d) &\text{(A.5)}\\
&+ r_{n,1}(t, x, \theta) + r_{n,2}(t, x, \theta)
\end{aligned}
$$

where

$$
\begin{aligned}
r_{n,1}(t, x, \theta) &= \frac{1}{2}\int_0^t \phi'''_\theta\left(\zeta(y, x)\right)\left\{\hat{s}_d(y, x\hat{\gamma}_d) - s_d(y, x\gamma_d)\right\}^2 s_{d,1}(dy, x\gamma_d),\\
r_{n,2}(t, x, \theta) &= \int_0^t \left\{\phi'_\theta\left(\hat{s}_d(y, x\hat{\gamma}_d)\right) - \phi'_\theta\left(s_d(y, x\gamma_d)\right)\right\}\left\{\hat{s}_{d,1}(dy, x\hat{\gamma}_d) - s_{d,1}(dy, x\gamma_d)\right\}.
\end{aligned}
$$

The random function ζ lies between $\hat{s}_d$ and s_d. The second equality follows by direct manipulation. The fourth line is due to a second order Taylor expansion of $\phi'_\theta\left(\hat{s}_d\right)$ around s_d, which also produces the remainder $r_{n,1}$, and the fifth one follows by an integration by part on the term in the third line.

The proof proceed in two steps: we first derive the dominating terms of (A.3) - (A.5), and then we show th two remainder terms $r_{n,1}$ and $r_{n,2}$ are asymptotically negligible.

Step 1: expansion of first-order terms.

It suffices to show (A.3). (A.4) and (A.5) can be handled analogously. Let $\ddot{\phi}^\theta_{d,\gamma}(y, x) \equiv \phi''_\theta\left(s_d(y, x\gamma)\right)$. A second order Taylor expansion of $\hat{s}_d$ with respect to γ around γ_d yields

$$
\begin{aligned}
&\int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y, x)\left\{\hat{s}_d(y, x\hat{\gamma}_d) - s_d(y, x\gamma_d)\right\}s_{d,1}(dy, x\gamma_d)\\
=&\int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y, x)(\hat{s}_d(y, x\gamma_d) - s_d(y, x\gamma_d))s_{d,1}(dy, x\gamma_d)\\
&- (\hat{\gamma}_d - \gamma_d)'\int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y, x)\int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y, x)\partial_\gamma \hat{G}_d(y, x\gamma_d)s_{d,1}(dy, x\gamma_d)
\end{aligned}
$$

$$
\begin{aligned}
&+ (\hat{\gamma}_d - \gamma_d)' \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y,x) \left\{ \partial_\gamma \hat{G}_d(y, x\tilde{\gamma}_d) - \partial_\gamma \hat{G}_d(y, x\gamma_d) \right\} s_{d,1}(dy, x\gamma_d) \\
\equiv & L_{n,1} + L_{n,21} + L_{n,22},
\end{aligned}
$$

where $\tilde{\gamma}_d$ lies between $\hat{\gamma}_d$ and γ_d. We rewrite the first term as

$$
\begin{aligned}
L_{n,1} =& - \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y,x) \left(\frac{\hat{\kappa}_{d,y}(x\gamma_d)}{\hat{f}_d(x\gamma_d)} - G_d(y, x\gamma_d) \right) s_{d,1}(dy, x\gamma_d) \\
=& - \frac{1}{n\hat{f}_d(x\gamma_d)} \sum_{i=1}^n K_h\left(x\gamma_d, X_i\gamma_d\right) \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y,x) \mathcal{E}_{d,\gamma_d,i}(y,x) s_{d,1}(dy, x\gamma_d) \\
=& - \frac{1}{nf_d(x\gamma_d)} \sum_{i=1}^n K_h\left(x\gamma_d, X_i\gamma_d\right) \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y,x) \mathcal{E}_{d,\gamma_d,i}(y,x) s_{d,1}(dy, x\gamma_d) \qquad (A.6) \\
&+ \frac{\hat{f}_d(x\gamma_d) - f_d(x\gamma_d)}{nf_d(x\gamma_d)\hat{f}_d(x\gamma_d)} \sum_{i=1}^n K_h\left(x\gamma_d, X_i\gamma_d\right) \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y,x) \mathcal{E}_{d,\gamma_d,i}(y,x) s_{d,1}(dy, x\gamma_d). \qquad (A.7)
\end{aligned}
$$

where $\mathcal{E}_{d,\gamma,\ell}(y,x) \equiv \mathbb{1}\{D_\ell = d\}\left(\mathbb{1}\{Y_\ell \leqslant y\} - G_d(y, x\gamma)\right)$. Note that the difference between $\mathcal{E}_{d,\gamma,\ell}(y,x)$ and $\mathcal{E}_{d,\gamma,\ell}(y)$ lies in whether X is fixed at x.

We divide (A.6) into two parts,

$$
\begin{aligned}
&- \frac{1}{nf_d(x\gamma_d)} \sum_{i=1}^n K_h\left(x\gamma_d, X_i\gamma_d\right) \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y,x) \mathcal{E}_{d,\gamma_d,i}(y) s_{d,1}(dy, x\gamma_d), \qquad (A.8) \\
&- \frac{1}{nf_d(x\gamma_d)} \sum_{i=1}^n K_h\left(x\gamma_d, X_i\gamma_d\right) \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y,x) (\mathcal{E}_{d,\gamma_d,i}(y,x) - \mathcal{E}_{d,\gamma_d,i}(y)) s_{d,1}(dy, x\gamma_d) \qquad (A.9)
\end{aligned}
$$

The first term in the preceding display is centered and belongs to $\eta_{s,d}$. The second term corresponds to the first-order bias and is part of $\eta_{b,d}$.

By the definition of $\hat{G}_d$, (A.7) is equal to

$$
\begin{aligned}
&\frac{\hat{f}_d(x\gamma_d) - f_d(x\gamma_d)}{f_d(x\gamma_d)} \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y,x) \left(\hat{G}_d(y, x\gamma_d) - G_d(y, x\gamma_d) \right) s_{d,1}(dy, x\gamma_d) \\
\lesssim & \sup_{\tilde{\mathcal{T}} \times \mathcal{X}} \left| \hat{f}_d(x\gamma_d) - f_d(x\gamma_d) \right| \cdot \sup_{\tilde{\mathcal{T}} \times \mathcal{X}} \left| \hat{G}_d(y, x\gamma_d) - G_d(y, x\gamma_d) \right| = O_p\left(\frac{\log n}{nh} \right),
\end{aligned}
$$

uniformly over Θ. The inequality follows by (B.5) and (B.6). The equality is due to Assumption 9.2.

Regarding $L_{n,21}$, we have

$$
- (\hat{\gamma}_d - \gamma_d)' \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y,x) G^{(1)}_d(y, x\gamma_d) s_{d,1}(dy, x\gamma_d)
$$

$$- (\hat{\gamma}_d - \gamma_d)' \int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x) \left\{ \partial_\gamma \hat{G}_d(y, x\gamma_d) - G_d^{(1)}(y, x\gamma_d) \right\} s_{d,1}(dy, x\gamma_d)$$

Under Assumption 6.4, $\phi''_\theta(u)$ is bounded on $[\upsilon_o, 1]$ uniformly in θ. Meanwhile, $s_d(y,x)$ is bounded in the same interval whenever $y \in \tilde{\mathcal{T}}$, uniformly in $x \in \mathcal{X}$. Consequently, $\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)$ is bounded on $\tilde{\mathcal{T}} \times \mathcal{X} \times \Theta$. By (B.1), the last term is bounded from above by $\sup_{\tilde{\mathcal{T}}\times\mathcal{X}} \left\| \partial_\gamma \hat{G}_d(t, x\gamma_d) - G_d^{(1)}(t, x\gamma_d) \right\| \|\hat{\gamma}_d - \gamma_d\| = \left(O_p\left((\log n)^{1/2} n^{-1/2} h^{-3/2} \right) + O(h^s) \right) \cdot O_p\left(n^{-1/2}\right)$, which is $\left(O_p\left((\log n)^{1/2} n^{-1} h^{-3/2} \right) \right)$ under our rate condition on the bandwidth. Therefore, $L_{n,21}$ is dominated by

$$-\frac{1}{n} \sum_{i=1}^n \int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x) G_d^{(1)}(y, x\gamma_d)' s_{d,1}(dy, x\gamma_d) V_d^{-1} \psi_2(X_i, \gamma_d).$$

From Lemma C.4, we deduce that $L_{n,22}$ has a uniform rate of $O_p\left(n^{1/2}\right) \cdot O_p\left((\log n)^{1/2} n^{-1} h^{-5/2} \right) = o_p\left((\log n)^{1/2} n^{-1} h^{-3/2} \right)$.

So far, we have derived the leading terms of (A.3). The other two terms, (A.4) and (A.5), can be treated analogously.

Step 2: uniform asymptotic negligibility of $r_{n,1}$ and $r_{n,2}$.

By the mean value theorem, we have

$$\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta} r_{n,1}(t,x,\theta) \lesssim \sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta} \left| \phi'''_\theta\left(\zeta(t,x)\right) \right| \left\{ \hat{s}_d(t,x) - s_d(t,x) \right\}^2,$$

We can deduce from (B.5), and Assumption 6.4 that the right hand side is of order $O_p\left(\log n \cdot n^{-1} h^{-1}\right) + O(h^{2s})$, which is $O_p\left(\log n \cdot n^{-1} h^{-1}\right)$ under Assumption 9.2.

Next, we show $r_{n,2} = O_p\left(\log n \cdot n^{-1} h^{-1}\right)$ as well. Perform a third order Taylor expansion of ϕ'_θ with respect to s_d,

$$\begin{aligned} r_{n,2}(t,x,\theta) &= \int_0^t \left\{ \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x) \left(\hat{s}_d(y, x\hat{\gamma}_d) - s_d(y, x\gamma_d) \right) \right\} \left\{ \hat{s}_{d,1}(dy, x\hat{\gamma}_d) - s_{d,1}(dy, x\gamma_d) \right\} \quad \text{(A.10)} \\ &+ \frac{1}{2} \int_0^t \phi'''_\theta\left(\zeta(y,x)\right) \left\{ \hat{s}_d(y, x\hat{\gamma}_d) - s_d(y, x\gamma_d) \right\}^2 \left\{ \hat{s}_{d,1}(dy, x\hat{\gamma}_d) - s_{d,1}(dy, x\gamma_d) \right\}. \end{aligned}$$

The second term is asymptotically dominated in view of Lemma B.2. Focusing on the first term, we have

$$\begin{aligned} \text{(A.10)} &= \int_0^t \left\{ \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x) \left(\hat{s}_d(y, x\hat{\gamma}_d) - \hat{s}_d(y, x\gamma_d) \right) \right\} \left\{ \hat{s}_{d,1}(dy, x\hat{\gamma}_d) - \hat{s}_{d,1}(dy, x\gamma_d) \right\} \quad \text{(A.11)} \\ &+ \int_0^t \left\{ \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x) \left(\hat{s}_d(y, x\hat{\gamma}_d) - \hat{s}_d(y, x\gamma_d) \right) \right\} \left\{ \hat{s}_{d,1}(dy, x\gamma_d) - s_{d,1}(dy, x\gamma_d) \right\} \quad \text{(A.12)} \end{aligned}$$

$$+\int_0^t \left\{\ddot{\phi}^\theta_{d,\gamma_d}(y,x)\left(\hat{s}_d(y,x\gamma_d)-s_d(y,x\gamma_d)\right)\right\}\left\{\hat{s}_{d,1}(dy,x\hat{\gamma}_d)-\hat{s}_{d,1}(dy,x\gamma_d)\right\} \tag{A.13}$$

$$+\int_0^t \left\{\ddot{\phi}^\theta_{d,\gamma_d}(y,x)\left(\hat{s}_d(y,x\gamma_d)-s_d(y,x\gamma_d)\right)\right\}\left\{\hat{s}_{d,1}(dy,x\gamma_d)-s_{d,1}(dy,x\gamma_d)\right\} \tag{A.14}$$

We analyze (A.11), (A.12), and (A.14) in turn. Via integration by parts, (A.13) can be handled in the same fashion as (A.12), and therefore, the proof is omitted.

We provide results for (A.11). By a first-order Taylor expansion of $\hat{s}_d(y,x\hat{\gamma}_d)$ in γ around γ_d, we get

$$(\hat{\gamma}_d-\gamma_d)'\int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y,x)\partial_\gamma\hat{G}_d(y,x\tilde{\gamma}_d)\partial_{\gamma'}\hat{G}_d(dy,x\breve{\gamma}_d)(\hat{\gamma}_d-\gamma_d),$$

where $\tilde{\gamma}_d$ and $\breve{\gamma}_d$ lie between $\hat{\gamma}_d$ and γ_d. Let $I_{d,t,1,i}=R_i\mathbb{1}\left\{D_i=d,Y_i\leqslant t\right\}$. Expanding the partial derivative in the integrator, we get

$$\begin{aligned}
&\frac{(\hat{\gamma}_d-\gamma_d)'}{nh^2\hat{f}_d(x\breve{\gamma}_d)}\sum_{i=1}^n I_{d,t,1,i}\ddot{\phi}^\theta_{d,\gamma_d}(Y_i,x)\partial_\gamma\hat{G}_d(Y_i,x\tilde{\gamma}_d)K^{(1)}\left((X_i\breve{\gamma}_d-x\breve{\gamma}_d)/h\right)(X_{[-1],i}-x_{[-1]})'(\hat{\gamma}_d-\gamma_d)\\
&-\frac{(\hat{\gamma}_d-\gamma_d)'}{nh\hat{f}_d(x\breve{\gamma}_d)^2}\sum_{i=1}^n I_{d,t,1,i}\ddot{\phi}^\theta_{d,\gamma_d}(Y_i,x)\partial_\gamma\hat{G}_d(Y_i,x\tilde{\gamma}_d)K((X_i\breve{\gamma}_d-x\breve{\gamma}_d)/h)\partial_{\gamma'}\hat{f}_d(x\breve{\gamma}_d)(\hat{\gamma}_d-\gamma_d)\\
&\equiv L_{n,31}+L_{n,32}.
\end{aligned}$$

Rewrite $L_{n,31}$ as

$$\frac{(\hat{\gamma}_d-\gamma_d)'}{nh^2 f_d(x\breve{\gamma}_d)}\sum_{i=1}^n I_{d,t,1,i}\ddot{\phi}^\theta_{d,\gamma_d}(Y_i,x)\partial_\gamma\hat{G}_d(Y_i,x\tilde{\gamma}_d)K^{(1)}\left((X_i\breve{\gamma}_d-x\breve{\gamma}_d)/h\right)(X_{[-1],i}-x_{[-1]})'(\hat{\gamma}_d-\gamma_d) \tag{A.15}$$

$$\begin{aligned}
&-\frac{(\hat{\gamma}_d-\gamma_d)'(\hat{f}_d(x\breve{\gamma}_d)-f_d(x\breve{\gamma}_d))}{nh^2 f_d(x\breve{\gamma}_d)\hat{f}_d(x\breve{\gamma}_d)}\\
&\cdot\sum_{i=1}^n I_{d,t,1,i}\ddot{\phi}^\theta_{d,\gamma_d}(Y_i,x)\partial_\gamma\hat{G}_d(Y_i,x\tilde{\gamma}_d)K^{(1)}\left((X_i\breve{\gamma}_d-x\breve{\gamma}_d)/h\right)(X_{[-1],i}-x_{[-1]})'(\hat{\gamma}_d-\gamma_d).
\end{aligned} \tag{A.16}$$

The term in (A.15) can be further decomposed as

$$\begin{aligned}
&\frac{(\hat{\gamma}_d-\gamma_d)'}{nhf_d(x\breve{\gamma}_d)}\sum_{i=1}^n I_{d,t,1,i}\ddot{\phi}^\theta_{d,\gamma_d}(Y_i,x)G^{(1)}_d(Y_i,x\tilde{\gamma}_d)K^{(1)}_h(x\gamma,X_i\gamma)(X_{[-1],i}-x_{[-1]})'(\hat{\gamma}_d-\gamma_d)\\
&-\frac{(\hat{\gamma}_d-\gamma_d)'}{nhf_d(x\breve{\gamma}_d)}\sum_{i=1}^n I_{d,t,1,i}\ddot{\phi}^\theta_{d,\gamma_d}(Y_i,x)(\partial_\gamma\hat{G}_d(Y_i,x\tilde{\gamma}_d)-G^{(1)}_d(Y_i,x\tilde{\gamma}_d))K^{(1)}_h(x\gamma,X_i\gamma)(X_{[-1],i}-x_{[-1]})'(\hat{\gamma}_d-\gamma_d).
\end{aligned}$$

From (B.2), we find that, Assumptions 3.1 and 6, $G^{(1)}_d(y,x\gamma)$ is bounded uniformly on $\tilde{\mathcal{T}}\times\mathcal{X}_\Gamma$. Additionally, $\ddot{\phi}^\theta_{d,\gamma_d}(y,x)$ is bounded on $\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta$ under Assumption 6.4. Hence, the first term can be bounded from above by

$$\frac{M_1}{h}\left\|\hat{\gamma}_d-\gamma_d\right\|^2 \sup_{(y,x,\gamma)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Gamma_{d,n}}\left\{\left\|G_d^{(1)}(y,x\gamma)\right\|\left\|x_{[-1]}\right\|\right\}\sup_{(x,\gamma)\in\mathcal{X}\times\Gamma_{d,n}}\left\{\frac{1}{n}\sum_{i=1}^n\left|K_h^{(1)}(x\gamma,X_i\gamma)\right|\right\}$$
$$=O_p(n^{-1}h^{-1})=o_p\left(\log n\cdot n^{-1}h^{-1}\right),$$

where the last equality holds uniformly over Θ. In view of (B.6), we deduce that the second is bounded by

$$\frac{M_2}{h}\left\|\hat{\gamma}_d-\gamma_d\right\|^2 \sup_{(y,x,\gamma)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Gamma_{d,n}}\left\{\left\|\partial_\gamma\hat{G}_d(y,x\gamma)-G_d^{(1)}(y,x\gamma)\right\|\left\|x_{[-1]}\right\|\right\}\sup_{x\in\mathcal{X}}\left\{\frac{1}{n}\sum_{i=1}^n\left|K_h^{(1)}(x\gamma,X_i\gamma)\right|\right\}$$
$$=O_p(n^{-1}h^{-1})\cdot\left(O_p\left((\log n)^{1/2}n^{-1/2}h^{-3/2}\right)+O(h^s)\right)=o_p\left(\log n\cdot n^{-1}h^{-1}\right),$$

uniformly over Θ. Applying similar reasoning, we are able to show that (A.16) is the order $O_p\left((\log n)^{1/2}n^{-3/2}h^{-3/2}\right)$, and $L_{n,32}=O_p\left(n^{-1}\right)$, both of which are $o_p\left(\log n\cdot n^{-1}h^{-1}\right)$.

Next, we bound (A.12). A Taylor expansion of $\hat{s}_d(y,x\hat{\gamma}_d)$ around γ_d yields,

$$\begin{aligned}(\text{A.12}) &= -(\hat{\gamma}_d-\gamma_d)'\int_0^t\left\{\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\partial_\gamma\hat{G}_d(y,x\gamma_d)\right\}\{\hat{s}_{d,1}(dy,x\gamma_d)-s_{d,1}(dy,x\gamma_d)\}\\
&+(\hat{\gamma}_d-\gamma_d)'\int_0^t\left\{\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\left(\partial_\gamma\hat{G}_d(y,x\tilde{\gamma}_d)-\partial_\gamma\hat{G}_d(y,x\gamma_d)\right)\right\}\{\hat{s}_{d,1}(dy,x\gamma_d)-s_{d,1}(dy,x\gamma_d)\}\\
&\equiv(\hat{\gamma}_d-\gamma_d)'B_{n,1}+(\hat{\gamma}_d-\gamma_d)'B_{n,2}.\end{aligned}$$

$B_{n,1}$ can be rewritten as

$$\begin{aligned}B_{n,1}=&\int_0^t\left\{\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\partial_\gamma\hat{G}_d(y,x\gamma_d)\right\}d\left\{\frac{\hat{\kappa}_{d,1,y}(x\gamma_d)}{f_d(x\gamma_d)}-G_{d,1}(y,x\gamma_d)\right\}\\
&-\frac{\hat{f}_d(x\gamma_d)-f(x\gamma_d,d)}{f(x\gamma_d,d)\hat{f}_d(x\gamma_d)}\int_0^t\left\{\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\partial_\gamma\hat{G}_d(y,x\gamma_d)\right\}d\hat{\kappa}_{d,1,y}(x\gamma_d).\end{aligned}$$

Similar analysis along the lines of $L_{n,31}$ gives a uniform bound of $O_p\left((\log n)^{1/2}n^{-1/2}h^{-1/2}\right)$ for the second term. Expanding the partial derivative in the first term leads to

$$\frac{1}{f(x\gamma_d,d)^2}\int_0^t\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\partial_\gamma\hat{\kappa}_{d,y}(x\gamma_d)d\left[\hat{\kappa}_{d,1,y}(x\gamma_d)-f(x\gamma_{d,1},d)G_{d,1}(y,x\gamma_d)\right] \tag{A.17}$$
$$-\frac{\partial_\gamma\hat{f}_d(x\gamma_d)}{f_d(x\gamma_d)\hat{f}_d(x\gamma_d)^2}\int_0^t\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\hat{\kappa}_{d,y}(x\gamma_d)d\left[\hat{\kappa}_{d,1,y}(x\gamma_{d,1})-f_d(x\gamma_d)G_{d,1}(y,x\gamma_d)\right] \tag{A.18}$$
$$-\frac{(\hat{f}_d(x\gamma_d)-f(x\gamma_d,d))}{f(x\gamma_d,d)^2\hat{f}_d(x\gamma_d)}\int_0^t\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\partial_\gamma\hat{\kappa}_{d,y}(x\gamma_d)d\left[\hat{\kappa}_{d,1,y}(x\gamma_d)-f_d(x\gamma_d)G_{d,1}(y,x\gamma_d)\right]$$

$$+\frac{\partial_\gamma \hat{f}_d(x\gamma_d)(\hat{f}_d(x\gamma_d)^2 - f(x\gamma_d, d)^2)}{f(x\gamma_d, d)^3 \hat{f}_d(x\gamma_d)^2} \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y,x)\hat{\kappa}_{d,y}(x\gamma_d) d\left[\hat{\kappa}_{d,1,y}(x\gamma_d) - f_d(x\gamma_d)G_{d,1}(y, x\gamma_{d,1})\right].$$

As $\hat{f}_d(x\gamma_d)$ converges uniformly to $f_d(x\gamma_d)$ in probability, the last two terms are clearly dominated by the first two in the limit. We therefore focus on the convergence of (A.17) and (A.18).

Let $\kappa_{d,y}(x\gamma) = \mathbb{E}[\mathbb{1}\{D = d, Y \leqslant y\} K_h(x\gamma, X\gamma)]$ and $\kappa_{d,1,y}(x\gamma) = \mathbb{E}[R\mathbb{1}\{D = d, Y \leqslant y\} \cdot K_h(x\gamma, X\gamma)]$. The integrator of (A.17) can be decomposed into a centered term $\nu_{d,1}(y, x, \gamma_d) = (\hat{\kappa}_{d,1,y}(x\gamma_d) - \kappa_{d,1,y}(x\gamma_d))$ and a bias term $\mu_{d,1}(y, x, \gamma_d) = (\kappa_{d,1,y}(x\gamma_d) - f_d(x\gamma_d)G_{d,1}(y, x\gamma_d))$.

Regarding the centered part, let us define

$$\begin{aligned}
L_{n,41,\ell} \equiv & \frac{1}{nhf(x\gamma_d, d)^2} \sum_{i=1}^n \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y,x) I_{d,y,i}(K_h^{(1)}(x\gamma_d, X_i\gamma_d)(X_{\ell,i} - x_\ell) \\
& - \mathbb{E}[I_{d,y}K_h^{(1)}(x\gamma_d, X\gamma_d)(X_\ell - x_\ell)])\nu_{d,1}(dy, x, \gamma_d), \\
L_{n,42,\ell} \equiv & \frac{1}{f(x\gamma_d, d)^2} \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y,x)\, \mathbb{E}[I_{d,y}h^{-1}K_h^{(1)}(x\gamma_d, X\gamma_d)(X_\ell - x_\ell)]\nu_{d,1}(dy, x, \gamma_d).
\end{aligned}$$

The first term can be represented as a degenerate second order U process indexed by ω. Specifically,

$$L_{n,41,\ell}(\omega) = \frac{1}{n^2h^3}\left\{\sum_{i=1}^n g_{1,\ell}(W_i, W_i, \omega) + \sum_{i\neq j}^n g_{1,\ell}(W_i, W_j, \omega)\right\} \equiv L^a_{n,41,\ell}(\omega) + L^b_{n,41,\ell}(\omega),$$

where

$$\begin{aligned}
g_{1,\ell}(W_1, W_2, \omega) = & \frac{1}{f(x\gamma_d, d)^2}\left\{g_{11}(W_1, \omega)g_{12,\ell}(W_2, Y_1, \omega) - \int g_{11}(w_1, \omega)g_{12,\ell}(W_2, y_1, \omega)dF_W(w_1)\right\}, \\
g_{11}(W_1, \omega) = & I_{d,t,1,1}hK_h(x\gamma_d, X_1\gamma_d), \\
g_{12,\ell}(W_2, y, \omega) = & \ddot{\phi}^\theta_{d,\gamma_d}(y,x)\left\{I_{d,y\wedge t,2}hK_h^{(1)}(x\gamma_d, X_2\gamma_d)(X_{2,\ell} - x_\ell)\right. \\
& \left. - \int \mathbb{1}\{d_2 = d, y_2 \leqslant y \wedge t\} hK_h^{(1)}(x\gamma_d, x_2\gamma_d)(x_{2,\ell} - x_\ell)dF_W(w_2)\right\}.
\end{aligned}$$

Direct calculation shows

$$\sup_{\omega\in\Omega}\left|L^a_{n,41,\ell}(\omega)\right| \lesssim \frac{1}{nh}\sup_{x\in\mathcal{X}}\left\{\frac{1}{n}\sum_{i=1}^n\left|K_h^{(1)}(x\gamma_d, X_i\gamma_d)K_h(x\gamma_d, X_i\gamma_d)\right|\right\} = O_p\left(\frac{1}{nh^2}\right).$$

Now, define the following class of functions

$$\mathcal{G}_1 \equiv \{(w_1, w_2) \mapsto g_{1,\ell}(w_1, w_2, \omega) : \ell \in \{2, ..., k\},\ \omega \in \Omega\}. \tag{A.19}$$

By Lemma C.3, it belongs to the VC type class with a bounded envelop. Standard calculations reveal that the maximum variance of U process kernel $\sup_{g\in\mathcal{G}_1}\mathbb{E}[g^2]$ is of the order $O(h^2)$. By the maximal inequality in Lemma C.1, we conclude that $\mathbb{E}\left[\sup_{\omega\in\Omega}\left|n^{-2}\sum_{i\neq j}^{n} g_{1,\ell}(W_i,W_j,\omega)\right|\right] = O\left(\log n\cdot n^{-1}h\right)$. Applying the Markov inequality and multiplying the U statistic by h^{-3}, we deduce that $\sup_{\omega\in\Omega}\left|L^{b}_{n,41,\ell}(\omega)\right| = O_p\left(\log n\cdot n^{-1}h^{-2}\right)$.

By Lemma B.2, the expectation in side $L_{n,42,\ell}$ is uniformly convergent to $\partial_{x\gamma}G_{d,1}(y,x\gamma_d)(x_\ell - \mathbb{E}[X_\ell|x\gamma_d])$. Thus,

$$\begin{aligned}L_{n,42,\ell} \equiv& \frac{1}{nhf(x\gamma_d,d)^2}\sum_{i=1}^{n}\Big\{I_{d,t,1,i}\ddot{\phi}^{\theta}_{d,\gamma_d}(Y_i,x)\left(\partial_{x\gamma}G_{d,1}(Y_i,x\gamma_d)(x_\ell-\mathbb{E}[X_\ell|x\gamma_d])\right)hK_h(x\gamma_d,X_i\gamma_d)\\&-\mathbb{E}\left[I_{d,t,1}\ddot{\phi}^{\theta}_{d,\gamma_d}(Y,x)\left(\partial_{x\gamma}G_{d,1}(Y,x\gamma_d)(x_\ell-\mathbb{E}[X_\ell|x\gamma_d])\right)hK_h(x\gamma_d,X\gamma_d)\right]\Big\} + h^{s}r_{n,3}(t,x,\theta),\end{aligned}$$

where $\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta}|r_{n,3}(t,x,\theta)| = o_p(1)$. The first term on the right hand side can be bounded via the maximal inequality as long as the following class of function is of the VC type:

$$\mathcal{G}_2 \equiv \{w\mapsto g_{2,\ell}(w,\omega) : \ell\in\{2,...,k\},\ \omega\in\Omega\}, \tag{A.20}$$

where

$$\begin{aligned}g_{2,\ell}(W,\omega) =& f(x\gamma_d,d)^{-2}\Big\{I_{d,t,1}\ddot{\phi}^{\theta}_{d,\gamma_d}(Y,x)\left(\partial_{x\gamma}G_{d,1}(Y,x\gamma_d)(x_\ell-\mathbb{E}[X_\ell|x\gamma_d])\right)hK_h(x\gamma_d,X\gamma_d)\\&-\int r_1\mathbb{1}\{d_1=d,y_1\leqslant t\}\ddot{\phi}^{\theta}_{d,\gamma_d}(y_1,x)\left(\partial_{x\gamma}G_{d,1}(y_1,x\gamma_d)(x_\ell-\mathbb{E}[X_\ell|x\gamma_d])\right)hK_h(x\gamma_d,x_1\gamma_d)dF_W(w_1)\Big\}.\end{aligned}$$

Note that $\sup_{g_2\in\mathcal{G}_2}\mathbb{E}[g_2^2] = O(h)$. Consequently, $L_{n,42,\ell} = O_p\left((\log n)^{1/2}\cdot n^{-1/2}h^{-1/2}\right)$ by the maximal inequality from Lemma C.1

Turning to the bias part of (A.17), we define

$$L_{n,5,\ell} \equiv \frac{1}{n}\sum_{i=1}^{n}\frac{X_{\ell,i}-x_\ell}{hf(x\gamma_d,d)^2}\int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)I_{d,y,i}K_h^{(1)}(x\gamma_d,X_i\gamma_d)\mu_{d,1}(dy,x,\gamma_d),$$

for $\ell = 2,...,k$. Integrating by parts gives

$$\begin{aligned}L_{n,5,\ell} =& \frac{1}{n}\sum_{i=1}^{n}\frac{X_{\ell,i}-x_\ell}{hf(x\gamma_d,d)^2}\ddot{\phi}^{\theta}_{d,\gamma_d}(t,x)I_{d,y,i}K_h^{(1)}(x\gamma_d,X_i\gamma_d)\mu_{d,1}(t,x,\gamma_d)\\&-\frac{X_{\ell,i}-x_\ell}{hf(x\gamma_d,d)^2}\int_0^t I_{d,y,i}K_h^{(1)}(x\gamma_d,X_i\gamma_d)\mu_{d,1}(y,x,\gamma_d)\ddot{\phi}^{\theta}_{d,\gamma_d}(dy,x)\\&-\frac{X_{\ell,i}-x_\ell}{hf(x\gamma_d,d)^2}I_{d,t,i}\mu_{d,1}(Y_i,x,\gamma_d)\ddot{\phi}^{\theta}_{d,\gamma_d}(Y_i,x)K_h^{(1)}(x\gamma_d,X_i\gamma_d),\end{aligned}\tag{A.21}$$

Due to the uniform convergence of the gradient estimator by (B.4), the first term is uniformly bounded from above by

$$\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta}\left\{\left|f_d(x\gamma_d)^{-2}\right|\left|\ddot{\phi}^{\theta}_{d,\gamma_d}(t,x)\right|\cdot\left\{\left\|G^{(1)}_d(t,x\gamma_d)\right\|+\left\|\partial_\gamma\hat{G}_d(t,x\gamma_d)-G^{(1)}_d(t,x\gamma_d)\right\|\right\}\right\}$$
$$\cdot\sup_{(t,x)\in\tilde{\mathcal{T}}\times\mathcal{X}}\left\{\left|\mu_{d,1}(t,x,\gamma_d)\right|\right\}=O_p(1)\cdot O(h^s).$$

Since $\phi''(\cdot)$ and $s_d(\cdot,x\gamma_d)$ are both continuously differentiable with bounded derivative under Assumption 6, we can deduce from the mean value theorem that the second term is also of the order $O_p(h^s)$. Standard bias calculation yields

$$(\text{A.21})=-\frac{\int u^sK(u)du}{nh^{2-s}}\sum_{i=1}^n g_{3,\ell}(W_i,\omega)+h^{s+1}r_{n,4,\ell}(t,x,d,\theta),$$

where $\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta}|r_{n,4,\ell}(t,x,d,\theta)|=O_p(1)$ and, for $\ell\in\{2,...,k\}$,

$$g_{3,\ell}(W,\omega)=\frac{X_\ell-x_\ell}{f(x\gamma_d,d)^2}I_{d,t}\ddot{\phi}^{\theta}_{d,\gamma_d}(Y,x)\partial^s_{x\gamma}\left\{f_d(x\gamma_d)G_{d,1}(Y,x\gamma_d)\right\}hK^{(1)}_h(x\gamma_d,X\gamma_d).$$

Once again, Lemma C.3 establishes that

$$\mathcal{G}_3\equiv\left\{w\mapsto g_{3,\ell}(w,\omega):\ell\in\{2,...,k\},\ \omega\in\Omega\right\},\tag{A.22}$$

is of VC type with bounded envelop. Maximal variance is of the order $O(h)$. We then deduce from Lemma C.1 and the Markov inequality that (A.21) is of the order $O_p\left((\log n)^{1/2}n^{-1/2}h^{s-3/2}\right)$, uniformly over $\omega\in\Omega$. Overall, (A.17) is $O_p\left((\log n)^{1/2}\cdot n^{-1/2}h^{-1/2}\right)$. Following same arguments, we can deduce that (A.18) is $O_p\left((\log n)^{1/2}\cdot n^{-1/2}h^{1/2}\right)$.

The same procedure can be followed to decompose $B_{n,2}$. In what follows, we derive the convergence rate for the $\hat{\kappa}_{d,1,y}$ part only since the denominator $\hat{f}_d$ can be treated analogously. Specifically, define

$$\begin{aligned}B_{n,21}\equiv&f(x\gamma_d)^{-1}\int_0^t\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\left\{\partial_\gamma\hat{G}_d(y,x\tilde{\gamma}_d)-\partial_\gamma\hat{G}_d(y,x\gamma_d)\right\}\nu_{d,1}(dy,x,\gamma_d)\\=&f(x\gamma_d)^{-1}\int_0^t\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\left\{\frac{\partial_\gamma\hat{\kappa}_{d,y}(x\tilde{\gamma}_d)}{\hat{f}_d(x\tilde{\gamma}_d)}-\frac{\partial_\gamma\hat{\kappa}_{d,y}(x\gamma_d)}{\hat{f}_d(x\gamma_d)}\right\}\nu_{d,1}(dy,x,\gamma_d),\\&-f(x\gamma_d)^{-1}\int_0^t\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\left\{\frac{\hat{\kappa}_{d,y}(x\tilde{\gamma}_d)\partial_\gamma\hat{f}_d(x\tilde{\gamma}_d)}{\hat{f}^2_d(x\tilde{\gamma}_d)}-\frac{\hat{\kappa}_{d,y}(x\gamma_d)\partial_\gamma\hat{f}_d(x\gamma_d)}{\hat{f}^2_d(x\gamma_d)}\right\}\nu_{d,1}(dy,x,\gamma_d).\end{aligned}$$

From similar arguments applied in Lemma C.4, one deduces that each of the two terms is dominated a degenerate second order U process that converges at a rate of $O_p\left(\log n\cdot n^{-1}h^{-7/2}\delta_n\right)$, uniformly for $\|\tilde{\gamma}_d-\gamma_d\|\leqslant\delta_n$. Since $\delta_n=O_p\left(n^{-1/2}\right)$, we find from Assumption 9.2 that

$B_{n,21} = o_p\left(\log n \cdot n^{-1/2}h^{-1}\right)$.

$$
\begin{aligned}
B_{n,22} \equiv & f(x\gamma_d)^{-1}\int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\left\{\partial_\gamma \hat{G}_d(y,x\tilde{\gamma}_d) - \partial_\gamma \hat{G}_d(y,x\gamma_d)\right\}\mu_{d,1}(dy,x,\gamma_d) \\
= & f_d(x\gamma_d)^{-2}\int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\left\{\partial_\gamma \hat{\kappa}_{d,y}(x\tilde{\gamma}_d) - \partial_\gamma \hat{\kappa}_{d,y}(x\gamma_d)\right\}\mu_{d,1}(dy,x,\gamma_d) \\
& - \frac{\hat{f}_d(x\tilde{\gamma}_d) - \hat{f}_d(x\gamma_d)}{f_d(x\tilde{\gamma}_d)f_d(x\gamma_d)^2}\int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\partial_\gamma \hat{\kappa}_{d,y}(x\gamma_d)\mu_{d,1}(dy,x,\gamma_d) \\
& - f(x\gamma_d)^{-1}\int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\left\{\frac{\hat{\kappa}_{d,y}(x\tilde{\gamma}_d)\partial_\gamma \hat{f}_d(x\tilde{\gamma}_d)}{\hat{f}_d^2(x\tilde{\gamma}_d)} - \frac{\hat{\kappa}_{d,y}(x\gamma_d)\partial_\gamma \hat{f}_d(x\gamma_d)}{\hat{f}_d^2(x\gamma_d)}\right\}\mu_{d,1}(dy,x,\gamma_d) + (s.o.) \\
\equiv & B^a_{n,22}(t,x,\theta) + B^b_{n,22}(t,x,\theta) + B^c_{n,22}(t,x,\theta) + (s.o.)
\end{aligned}
$$

Integration by parts turns $B^a_{n,22}$ into three terms. Applying arguments of Lemma C.4, and properly accounting for the biases, to each of the terms, one deduces that $\sup_{\|\gamma-\gamma_d\|\leqslant\delta_n}\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta}\left\|B^a_{n,22}(t,x,\theta)\right\| = O_p\left((\log n)^{1/2}n^{-1/2}h^{s-5/2}\delta_n\right)$. The same uniform rate applies to $B^b_{n,22}$, and, after further decomposition, to $B^c_{n,22}$.

Collect results on $B_{n,1}$ and $B_{n,2}$, and multiply them by $O_p\left(n^{-1/2}\right)$. We conclude that (A.12) is $o_p\left(\log n \cdot n^{-1}h^{-1}\right)$.

Lastly, we bound (A.14). Rewriting the term as

$$
\begin{aligned}
& \int_0^t \left\{\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\left(\nu_{n,d}(y,x\gamma_d) + \mu_d(y,x\gamma_d)\right)\right\}\left\{\nu_{n,d,1}(dy,x\gamma_d) + \mu_{d,1}(dy,x\gamma_d)\right\} \\
= & \int_0^t \left\{\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\nu_{n,d}(y,x\gamma_d)\right\}\nu_{n,d,1}(dy,x\gamma_d) + \int_0^t \left\{\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\nu_{n,d}(y,x\gamma_d)\right\}\mu_{d,1}(dy,x\gamma_d) \\
& + \int_0^t \left\{\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\mu_d(y,x\gamma_d)\right\}\nu_{n,d,1}(dy,x\gamma_d) + \int_0^t \left\{\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\mu_d(y,x\gamma_d)\right\}\mu_{d,1}(dy,x\gamma_d).
\end{aligned}
$$

Applying arguments similar to those from Lemma 3.1 of Lopez (2011) and Lemma A.2 of Fan and Liu (2018) yields that the last three terms are asymptotically dominated by the first one due to under-smoothing. Hence, we provide detailed derivation for the first term only. Let

$$
\begin{aligned}
L_{n,6} = & \int_0^t \left\{\ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\nu_{n,d}(y,x\gamma_d)\right\}\nu_{n,d,1}(dy,x\gamma_d) \\
= & \frac{1}{n^2h^2}\left\{\sum_{i=1}^n g_4(W_i,W_i,\omega) + \sum_{i\neq j}^n g_4(W_i,W_j,\omega)\right\} \equiv L^a_{n,6} + L^b_{n,6},
\end{aligned}
$$

where

$$
g_4(W_1,W_2,\omega) = \left\{g_{41}(W_1,\omega)g_{42}(W_2,Y_1,\omega) - \int g_{41}(w_1,\omega)g_{42}(W_2,y_1,\omega)dF(y_1,x_1,d_1,r_1)\right\},
$$

$$
\begin{aligned}
g_{41}(W_1,\omega) &= R_1 \mathbb{1}\{D_1 = d, Y_1 \leqslant t\} hK_h(x\gamma_d, X_1\gamma_d),\\
g_{42}(W_2,y,\omega) &= \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\Big\{\mathbb{1}\{D_2 = d, Y_2 \leqslant y \wedge t\} hK_h(x\gamma_d, X_2\gamma_d)\\
&\qquad - \int \mathbb{1}\{d_2 = d, y_2 \leqslant y \wedge t\} hK_h(x\gamma_d, x_2\gamma_d) dF(y_2, x_2, d_2)\Big\}.
\end{aligned}
$$

Note that the second term is a degenerate second order U process. Lemma C.3 indicates that the class

$$
\mathcal{G}_4 = \{(w_1, w_2) \mapsto g_4(w_1, w_2, \omega) : \omega \in \Omega\} \tag{A.23}
$$

is of VC type with a bounded envelop. Standard calculation implies that $\sup_{\omega\in\Omega}\left|L^a_{n,6}\right| = O_p(n^{-1}h^{-1})$ and the maximal variance $\sup_{g\in\mathcal{G}_4}\mathbb{E}[g^2]$ is $O(h^2)$. Another application of Theorem 8 of Giné and Mason (2007) and the Markov inequality yields that $L^b_{n,6}$ is of order $O_p(\log n \cdot n^{-1}h^{-1})$ uniformly over Ω.

Gathering results on (A.11) - (A.14), we conclude that $\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta}|r_{n,2}(t,x,\theta)| = O_p(\log n \cdot n^{-1}h^{-1}) = o_p\left(n^{-1/2}\right)$, concluding the proof of Step 2.

To finish the proof, we deduce from a second order Taylor expansion of ϕ_θ^{-1} that, for each $(t,x,\theta) \in \tilde{\mathcal{T}} \times \mathcal{X} \times \Theta$,

$$
\begin{aligned}
\hat{s}_{T_d}(t, x\hat{\gamma}_d, \theta) - s_{T_d}(t, x\gamma_d, \theta) =& \frac{1}{\dot{\phi}'_\theta(s_{T_d}(t, x\gamma_d, \theta))}\left(\phi_\theta\left(\hat{s}_{T_d}(t, x\hat{\gamma}_d, \theta)\right) - \phi_\theta\left(s_{T_d}(t, x\gamma_d, \theta)\right)\right)\\
&- \frac{\ddot{\phi}_\theta^{-1}(\tilde{s}_d(t,x,\theta))}{\dot{\phi}_\theta^{-1}(\tilde{s}_d(t,x,\theta))^3}\left(\phi_\theta\left(\hat{s}_{T_d}(t, x\hat{\gamma}_d, \theta)\right) - \phi_\theta\left(s_{T_d}(t, x\gamma_d, \theta)\right)\right)^2,
\end{aligned}
$$

where the random function $\tilde{s}_d(t,x,\theta)$ lies between $\hat{s}_{T_d}(t, x\hat{\gamma}_d, \theta)$ and $s_{T_d}(t, x\gamma_d, \theta)$. From Assumption 6.4, it holds that both $1/\dot{\phi}'_\theta(z)$ and $\ddot{\phi}_\theta^{-1}(z)$ are uniformly bounded when $z \in [0, y_o^*]$. Also, by the definition of y_o^*, the event $\mathbb{1}\{\tilde{s}_d(t,x,\theta) \leqslant y_o^*\}$ has the asymptotic probability equal to one, uniformly in (t,x,θ). As a result, the second term is asymptotically negligible. This concludes the proof. ■

A.2.2 Weak Convergence of the CBGP and UBGP

Proof of Corollary 1. The uniform representation from Theorem 4.1 consists of four parts. Standard analysis using maximal inequality implies that $\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta}\left|n^{-1}\sum_{i=1}^n \eta_{l,d}(W_i,x,t,\theta)\right| = O_p(n^{-1/2})$. Multiplying the quantity by $\sqrt{nh}$ gives a rate of $O_p\left(h^{1/2}\right) = o_p(1)$ for the second part. Moreover, $\sqrt{nh}\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta} r_n(x,t,\theta) = O_p\left((\log n)^{1/2} n^{-1/2}h^{-1}\right) = o_p(1)$ under Assumption 9.2. Next, we show that $\sqrt{nh}\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta}\left|n^{-1}\sum_{i=1}^n \eta_{b,d}(W_i,x,t,\theta)\right| = o_p(1)$. Let $\tilde{\eta}_{b,d} \equiv \eta_{b,d} - \mathbb{E}[\eta_{b,d}]$ denote the centered version of $\eta_{b,d}$. Standard bias calculation shows that $\mathbb{E}[\eta_{b,d}(W,x,t,\theta)] = O(h^s)$. Under Assumption 6, the rate of bias holds uniformly in $(t,x,\theta) \in \tilde{\mathcal{T}} \times \mathcal{X} \times \Theta$. Define

$$\mathcal{G}_b \equiv \{\tilde{w} \mapsto K(x\gamma_d, \tilde{x}\gamma_d)\Psi_d(G_d(\cdot,\tilde{x}\gamma_d) - G_d(\cdot,x\gamma_d), G_{d,1}(\cdot,\tilde{x}\gamma_d) - G_{d,1}(\cdot,x\gamma_d))(t,x,\theta) : (t,\theta) \in \tilde{\mathcal{T}} \times \Theta\}. \tag{A.24}$$

From Lemma C.3, this is a VC type class with bounded entropy. We show below that its maximum variance is $O(h^2)$. It suffices to illustrate on the first part, i.e.

$$\begin{aligned}
&\mathbb{E}\left[K(x\gamma_d, X\gamma_d)^2\left(\int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\left\{G_d(y,X\gamma_d) - G_d(y,x\gamma_d)\right\}s_{d,1}(dy,x\gamma_d)\right)^2\right] \\
&= h^3\int_{\mathbb{R}} u^2k(u)^2du \cdot \left(\int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\partial_{x\gamma}G_d(y,\breve{x}\gamma_d)s_{d,1}(dy,x\gamma_d)\right)^2 \\
&\leqslant h^3\int_{\mathbb{R}} u^2k(u)^2du \cdot \sup_{(u,\theta)\in[v_o,1]\times\Theta}\left|\phi''_{\theta}(u)\right|^2 \sup_{(y,x)\in\tilde{\mathcal{T}}\times\mathcal{X}}\left\{\left|\partial_{x\gamma}G_d(y,x\gamma_d)\right|^2\left|s_{d,1}(y,x\gamma_d)\right|^2\right\} = O(h^3),
\end{aligned}$$

where the first equality is due to the mean value theorem and a change of variable. The inequality follows under Assumption 6. It follows by Lemma C.1 and after multiplying by h^{-1}, that $\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta}|n^{-1}\sum_{i=1}^n \eta_{b,d}(W_i,x,t,\theta)| = O_p\left((\log n)^{1/2}n^{-1/2}h^{1/2}\right)$. Combining with the result on bias, we have $\sqrt{nh}\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta}\left|\frac{1}{n}\sum_{i=1}^n \eta_{b,d}(W_i,x,t,\theta)\right| = O_p\left((\log n)^{1/2}h + n^{1/2}h^{s+1}\right)$, which is $o_p(1)$ under Assumption 9.2.

Hence, it remains to prove the weak convergence of

$$\hat{\mathbb{G}}^{x\dagger}_n \equiv \sum_{i=1}^n f^x_{ni}(t,\theta,d),$$

where $f^x_{ni}(t,\theta,d) \equiv n^{-1/2}h^{1/2}\eta_{s,d}(W_i,x,t,\theta)$, for each $(d,t,\theta) \in \{0,1\} \times \tilde{\mathcal{T}} \times \Theta$ and given a fixed value x. This task can be accomplished by invoking the functional central limit theorem for non-identically distributed stochastic process as presented in Lemma C.2. Note that under Assumption 5.1, the triangular array $\{f^x_{ni}(t,\theta,d)\}$ is row-wise independent. By definition of $\eta_{s,d}(W_i,x,t,\theta)$, and Assumption 6, all of the array components $f^x_{ni}(t,\theta,d)$ are right continuous in both t and θ, which implies that the triangular array is separable. Consequently, $\{f^x_{ni}(t,\theta,d)\}$ is AMS by Lemma 2 of Kosorok (2003).

To verify manageability, we first note that

$$\mathcal{G}_\eta = \{\tilde{w} \mapsto K(x\gamma_d, \tilde{x}\gamma_d)\Psi_d(\mathcal{E}_{d,\gamma_d}, \mathcal{E}_{d,1,\gamma_d})(t,x,\theta) : (t,\theta) \in \tilde{\mathcal{T}} \times \Theta\} \tag{A.25}$$

is a VC class with an envelop $\sum_{d=0,1} H_{\eta,d}(\tilde{x}\gamma_d)$ by Lemma C.3, where $0 \leqslant H_{\eta,d}(\cdot) < M$ for all $\tilde{x} \in \mathcal{X}$ and a positive constant M. Multiplying $n^{-1/2}h^{-1/2}$ preserves the VC property and we conclude by Theorem 11.21 in Kosorok (2008) that $\{f_{ni}\}$ is manageable with the envelop $\{F_{ni}\}$, where $F_{ni}(\tilde{w}) = n^{-1/2}h^{-1/2}\sum_{d=0,1} H_{\eta,d}(\tilde{x}\gamma_d)$, for all $i = 1, ..., n$.

For condition (ii), we define $\chi^x_{n,d}(t,\theta) = \sum_{i=1}^n f^x_{ni}(t,\theta,d) - \mathbb{E}[f^x_{ni}(t,\theta,d)]$. As a result of the independence of W_i and W_j when $i \neq j$ and the fact that $\mathbb{E}[f^x_{ni}(t,\theta,d)] = 0$,

$$\mathbb{E}[\chi^x_{n,d_1}(t_1,\theta_1)\chi^x_{n,d_2}(t_2,\theta_2)] = \sum_{i=1}^n \mathbb{E}\left[f^x_{ni}(t_1,\theta_1,d_1)f^x_{ni}(t_2,\theta_2,d_2)\right].$$

Furthermore, the right hand side is identically zero if $d_1 \neq d_2$ due to the definition of $\mathcal{E}_{d,\gamma}$ and $\mathcal{E}_{d,1,\gamma}$. Condition (ii) is trivially satisfied in this case, and thus we focus on $d_1 = d_2 = d$. From direct calculations in Appendix C.2.1, we find

$$\mathbb{E}\left[f^x_{ni}(t_1,\theta_1,d)f^x_{ni}(t_2,\theta_2,d)\right] = \frac{1}{n}\sigma^2_{d,x}(t_1,\theta_1,t_2,\theta_2) + O(n^{-1}h). \tag{A.26}$$

where $\sigma^2_{d,x}$ is defined in (C.1). Under Assumptions 8.1 and 6, $\|K\|$, $\phi'_{\cdot}$, $\phi''_{\cdot}$, G_d and $G_{d,1}$ are all uniformly bounded. In addition, $f_d(x\gamma_d)$ and $\phi'_{\cdot}(s_{T_d}(\cdot, x\gamma_d, \cdot))$ are uniformly bounded away from zero for each $x \in \mathcal{X}$, under Assumptions 3.1 and 6.4. Since $h \to 0$ as $n \to \infty$, $\lim_{n\to\infty}\mathbb{E}[\chi^x_{n,d_1}(t_1,\theta_1)\chi^x_{n,d_2}(t_2,\theta_2)] = \sigma^2_{d,x}(t_1,\theta_1,t_2,\theta_2)$ and the limit is well-defined. As a result, condition (ii) holds.

Next, condition (iii) follows from the fact that

$$\sum_{i=1}^n \mathbb{E}^*[F^2_{ni}] \leqslant 2\sum_{d=0,1}\int_{\mathcal{X}_\Gamma} h^{-1}H^2_{\eta,d}(\tilde{x}\gamma_d)f(\tilde{x}\gamma_d)d\tilde{x}\gamma_d = 2\sum_{d=0,1} C^2_d \int_{[-1,1]} f(x\gamma_d + uh)du < \infty,$$

where the second equality follows from a change of variable, and the last inequality is due to $H_{\eta,d}$ being uniformly bounded.

Regarding the Lindeberg type condition (iv), note that

$$\begin{aligned}
&\sum_{i=1}^n \mathbb{E}^*[F^2_{ni}\mathbb{1}\{F_{ni} > \epsilon\}] \\
&= \int_{\mathcal{X}_{\gamma_0}}\int_{\mathcal{X}_{\gamma_1}} h^{-1}\left(\sum_{d=0,1} H_{\eta,d}(\tilde{x}\gamma_d)\right)^2 \mathbb{1}\left\{n^{-1/2}h^{-1/2}\sum_{d=0,1} H_{\eta,d}(\tilde{x}\gamma_d) > \epsilon\right\} f(\tilde{x}\gamma_1,\tilde{x}\gamma_0)d\tilde{x}\gamma_1 d\tilde{x}\gamma_0 \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} h\left(\sum_{d=0,1} C_d\mathbb{1}\{|u_d| \leqslant 1\}\right)^2 \\
&\quad \cdot \mathbb{1}\left\{n^{-1/2}h^{-1/2}\sum_{d=0,1} C_d\mathbb{1}\{|u_d| \leqslant 1\} > \epsilon\right\} f(x\gamma_1 + u_1 h, x\gamma_0 + u_0 h)du_1 du_0,
\end{aligned}$$

Since $nh \to \infty$, the limit of the right hand side as $n \to \infty$ equals zero for each ϵ by the dominated convergence theorem. Thus condition (iv) is satisfied.

In view of (A.26), we obtain from expanding the square in $\rho_n(s,t)$ that $\rho_n(t_1,\theta_1,t_2,\theta_2) = \rho(t_1,\theta_1,t_2,\theta_2) + O(h)$, with $\rho(t_1,\theta_1,t_2,\theta_2) = \left\{\sigma^2_{d,x}(t_1,t_1,\theta_1,\theta_1) - 2\sigma^2_{d,x}(t_1,t_2,\theta_1,\theta_2) + \sigma^2_{d,x}(t_2,t_2,\theta_2,\theta_2)\right\}^{1/2}$ for each $(t_1,t_2,\theta_1,\theta_2) \in \tilde{\mathcal{T}}^2 \times \Theta^2$. Since the second term vanishes as $n \to \infty$ and the first one is independent of n, we have $\rho(t_{1,n},\theta_{1,n},t_{2,n},\theta_{2,n}) \to 0$ implies $\rho_0(t_{1,n},\theta_{1,n},t_{2,n},\theta_{2,n}) \to 0$, for all deterministic sequences of $\{t_{1,n},\theta_{1,n}\}$ and $\{t_{2,n},\theta_{2,n}\}$.

We have shown that the triangular array $\{f_{ni}\}$ satisfies conditions (i) - (v) of Lemma C.2, which implies that $\hat{\mathbb{G}}^{x\dagger}_n$ converges weakly to a two-dimensional Gaussian process with covariance function $\Sigma^{x\dagger}_\eta(\cdot,\cdot)$. Lemma C.6 shows that $\Sigma^{x}_\eta(\cdot,\cdot) = \Sigma^{x\dagger}_\eta(\cdot,\cdot) + o(1)$. Combining this result with the fact that $\hat{\mathbb{G}}^{x}_n - \hat{\mathbb{G}}^{x\dagger}_n = o_p(1)$ concludes the proof. ∎

Proof of Corollary 2. Proof of part (i). In view of the uniform representation from Theorem 4.1, we obtain

$$
\begin{aligned}
\hat{s}_{T_d}(t,\theta) - s_{T_d}(t,\theta) =& \mathbb{E}_n\left[s_{T_d}(y,X\gamma_d,\theta) - \mathbb{E}[s_{T_d}(y,X\gamma_d,\theta)]\right] \\
&+ \mathbb{E}_n\left[\hat{s}_{T,d}(y,X\hat{\gamma}_d,\theta) - s_{T_d}(y,X\gamma_d,\theta)\right] \\
\equiv& A_{n,1} + A_{n,2}
\end{aligned}
$$

The first term is an empirical process indexed by $\varphi_{d,2}$. Utilizing the uniform linear representation in Theorem 4.1, we can further decompose $A_{n,2}$ as

$$
\frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n\left\{h^{-1}g_5(W_i,W_j,t,h,\theta) + g_6(W_i,W_j,\theta) + \eta_{b,d}(W_i,X_j,t,\theta)\right\} + \frac{1}{n}\sum_{i=1}^n + r_n(X_i,t,\theta),
$$

where

$$
\begin{aligned}
g_5(W_1,W_2,t,h,\theta) &= \frac{hK_h(X_2\gamma_d,X_1\gamma_d)}{f_d(X_2\gamma_d)}\Psi(\mathcal{E}_{d,\gamma_d,1},\mathcal{E}_{d,1,\gamma_d,1})(t,X_2,\theta),\\
g_6(W_1,W_2,\theta) &= \frac{\psi^b_d(X_1)'V_d^{-1}}{f_d(X_2\gamma_d)}\Psi(\psi^a_d,\psi^a_{d,1})(t,X_2,\theta),
\end{aligned}
$$

From the proof of Corollary 1, we find that $\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta}\left|n^{-1}\sum_{i=1}^n \eta_{b,d}(W_i,x,t,\theta)\right| = O_p\left((\log n)^{1/2}n^{-1/2}h^{1/2}\right) + O(h^s)$, thus the double mean involving $\eta_{b,d}$ is uniformly $o_p\left(n^{-1/2}\right)$ under Assumption 9.2. Additionally, from Theorem 4.1 we have $r_n(x,t,\theta) = o_p\left(n^{-1/2}\right)$ uniformly over $\mathcal{X}\times\tilde{\mathcal{T}}\times\Theta$, thus the last term is also asymptotically negligible.

Consequently, it suffices to work on the first two terms. We will show (a) the U-process indexed by g_5 is asymptotically equivalent to an empirical process indexed by $\mathbb{E}[g_5|W_1]$, and (b)

the U-process indexed by $\eta_{l,d}$ is asymptotically negligible.

We focus on (a) first. It is straightforward to show that $\Psi(\mathcal{E}_{d,\gamma_d,1}, \mathcal{E}_{d,1,\gamma_d,1})/f_d(x\gamma_d)$ is uniformly bounded. We therefore deduce from direct calculations that $\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta}\left|\frac{1}{n^2h}\sum_{i=1}^n g_5(W_i, W_i, t, h, \theta)\right| = O_p\left(n^{-1}h^{-1}\right)$. Observe that, by Lemma C.3,

$$\mathcal{G}_5 = \left\{(w_1, w_2) \mapsto g_5(w_1, w_2, t, h, \theta) : (t, h, \theta) \in \tilde{\mathcal{T}} \times \mathcal{H} \times \Theta\right\}, \tag{A.27}$$

is of VC type with bounded envelop. Also, $\mathbb{E}[g_5|W_2] = 0$ and $\sup_{g\in\mathcal{G}_5}\mathbb{E}[g^2] = O(h)$. As a result, we may deduce from the maximal inequality in Lemma C.1 that

$$\begin{aligned}\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta}&\left|\left\{\frac{1}{n(n-1)h}\sum_{i\neq j}^n g_5(W_i, W_j, t, h, \theta) - \frac{1}{nh}\sum_{i=1}^n \mathbb{E}[g_5(W_i, W_j, t, h, \theta)|W_i]\right\}\right| \\ &= O_p\left(\log n \cdot n^{-1}h^{-1/2}\right) = o_p\left(n^{-1/2}\right),\end{aligned}$$

which implies that the second order U process can be uniformly approximated by an empirical process indexed by the conditional mean. Let $A_{n,21}$ denote this process, for which we have the following

$$\begin{aligned}&\mathbb{E}[g_5(W_1, W_2, t, h, \theta)|W_1] \\ =&h\int \frac{K_h(x\gamma_d, X_1\gamma_d)}{f_d(x\gamma_d)}\Psi(\mathcal{E}_{d,\gamma_d,1}, \mathcal{E}_{d,1,\gamma_d,1})(t, x, \theta)f(x\gamma_d)d(x\gamma_d) \\ =&h\int \frac{K(u)}{f_d(X_1\gamma_d + uh)\phi'_\theta\left(s_{T_d}(t, X_1\gamma_d + uh, \theta)\right)} \\ &\cdot\left\{\int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y, X_1\gamma_d + uh)\mathbb{1}\left\{D_1 = d\right\}\left(\mathbb{1}\left\{Y_1 \leqslant y\right\} - G_d(y, X_1\gamma_d)\right)s_{d,1}(dy, X_1\gamma_d + uh)\right. \\ &- \phi'_\theta\left(s_d(t, X_1\gamma_d + uh)\right)\mathbb{1}\left\{D_1 = d\right\}\left(R_1\mathbb{1}\left\{Y_1 \leqslant t\right\} - G_{d,1}(t, X_1\gamma_d)\right) \\ &- \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y, X_1\gamma_d + uh)\mathbb{1}\left\{D_1 = d\right\}\left(R_1\mathbb{1}\left\{Y_1 \leqslant y\right\} - G_{d,1}(y, X_1\gamma_d)\right) \\ &\left.\cdot s_d(dy, X_1\gamma_d + uh)f(X_1\gamma_d + uh)\right\}du \\ =&\frac{hf(X_1\gamma_d)}{f_d(X_1\gamma_d)}\Psi(\mathcal{E}_{d,\gamma_d,1}, \mathcal{E}_{d,1,\gamma_d,1})(t, X_1, \theta) + O\left(h^{s+1}\right).\end{aligned}$$

The second equality follows by a change of variable and the last one is due to Assumption 8.1. Since $\phi'_\theta(\cdot), \phi''_\theta(\cdot), G_d(y,\cdot), G_{d,1}(y,\cdot), f_d(\cdot)$, and $f(\cdot)$ are all $(s+1)$ times continuously differentiable with uniformly bounded derivatives under Assumption 6, the rate of the bias holds uniformly over $\tilde{\mathcal{T}}\times\Theta$.

Now, we show (b). Observe that g_6 is multiplicatively separable in W_1 and W_2. The part involving W_1 is $O_p\left(n^{-1/2}\right)$ and not indexed by (t,θ), it then suffices to show that the empirical process indexed by $g_{61}(W_2, t, \theta) \equiv \Psi_d(\psi_d^a, \psi_{d,1}^a)(t, X_2, \theta)/f_d\left(X_2\gamma_d\right)$ is $o_p(1)$. Note that

$$\begin{aligned}\mathbb{E}[\psi_d^a(y, X_2)|X_2\gamma_d] &= \mathbb{E}[\partial_{x\gamma}G_d(y, X_2\gamma_d)(\mathbb{E}_{X_1\gamma_d}[X_1|X_2\gamma_d] - X_2)|X_2\gamma_d] \\ &= \partial_{x\gamma}G_d(y, X_2\gamma_d)(\mathbb{E}_{X_1\gamma_d}[X_1|X_2\gamma_d] - \mathbb{E}_{X_2\gamma_d}[X_2|X_2\gamma_d]) = 0,\end{aligned}$$

where the last equality holds because X_1 and X_2 are identically distributed. The same result holds for $\psi_{d,1}^a$. By Fubini's theorem, and the law of iterated expectation, it follows that $\mathbb{E}[g_{61}] = 0$. Next, let

$$\mathcal{G}_6 = \left\{(w_1, w_2) \mapsto g_{61}(w_1, w_2, t, \theta) : (t, \theta) \in \tilde{\mathcal{T}} \times \Theta\right\}. \tag{A.28}$$

Lemma C.3 establishes that it is of the VC type with bounded envelop. Moreover, its maximal variance is $O(1)$. We deduce from Lemma C.1 that $\mathbb{E}_n[g_{61}(W_2, t, \theta)] = O_p\left(n^{-1/2}\right)$ uniformly over $\tilde{\mathcal{T}} \times \Theta$. Overall, the U process indexed by g_6 is $O_p(n^{-1})$, and thus, asymptotically negligible.

Proof of part (ii). In view of the uniform representation established in the previous part, weak convergence follows from Theorem 2.1 of Kosorok (2008) if the class of function

$$\mathcal{G}_\varphi \equiv \{w \mapsto \varphi_d(w, t, \theta) : (d, t, \theta) \in \{0, 1\} \times \tilde{\mathcal{T}} \times \Theta\} \tag{A.29}$$

is Donsker. We see from Lemma C.3 that $\mathcal{G}_\varphi$ of VC type, which implies that it is Donsker by Theorem 19.14 in Van der Vaart (1998). ■

A.2.3 Functional Delta Method

Before proving Theorem 4.2, we state a general result on functional delta method. Let $\nu(\cdot)$ denote a generic functional mapping from $\ell_\infty(\tilde{\mathcal{T}} \times \Theta^2) \times \ell_\infty(\tilde{\mathcal{T}} \times \Theta^2)$ to a normed space $\ell_\infty(\mathcal{U} \times \Theta^2) \times \ell_\infty(\mathcal{U} \times \Theta^2)$.

Lemma A.1 *(i) Suppose this functional of interest is Hadamard differentiable at $\mathbf{S}^x$, for a fix $x \in \mathcal{X}$, tangentially to a space $\mathcal{C}(\mathcal{U} \times \Theta^2)$ with derivative $\nu'_{\mathbf{S}^x}$, and that the assumptions of Corollary 1 hold, then*

$$\sqrt{nh}\left(\nu\left(\hat{\mathbf{S}}^x\right)(\cdot,\cdot) - \nu\left(\mathbf{S}^x\right)(\cdot,\cdot)\right) \Rightarrow \nu'_{\mathbf{S}^x}(\mathbb{G})(\cdot,\cdot) \equiv \mathbb{G}_\nu^x,$$

in $\ell_\infty(\mathcal{U} \times \Theta^2) \times \ell_\infty(\mathcal{U} \times \Theta^2)$, where $\mathbb{G}_\nu^x$ is a two-dimensional Gaussian process with zero mean and covariance function,

$$\Sigma_\nu^x(\mathbf{u}, \boldsymbol{\theta}) = \mathbb{E}\left[\boldsymbol{\varphi}_\nu^x(W, u_1, \boldsymbol{\theta}_1)\boldsymbol{\varphi}_\nu^x(W, u_2, \boldsymbol{\theta}_2)'\right],$$

where $\boldsymbol{\varphi}_\nu^x \equiv \nu'_{\mathbf{S}^x}(\boldsymbol{\varphi}^x)$, for each $\mathbf{u} = (u_1, u_2)' \in \mathcal{U} \times \mathcal{U}$, and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \boldsymbol{\theta}_2')' \in \Theta^2 \times \Theta^2$.

(ii) Suppose $\nu(\cdot)$ is Hadamard differentiable at $\mathbf{S}$ tangentially to a space $\mathcal{C}(\mathcal{U} \times \Theta^2)$ with derivative $\nu'_{\mathbf{S}}$, and that the assumptions of Corollary 2 hold, then

$$\sqrt{n}\left(\nu\left(\hat{\mathbf{S}}\right)(\cdot,\cdot) - \nu\left(\mathbf{S}\right)(\cdot,\cdot)\right) \Rightarrow \nu'_{\mathbf{S}}(\mathbb{G})(\cdot,\cdot) \equiv \mathbb{G}_\nu,$$

in $\ell_\infty(\mathcal{U} \times \Theta^2) \times \ell_\infty(\mathcal{U} \times \Theta^2)$, where $\mathbb{G}_\nu$ is a two-dimensional Gaussian process with zero mean and covariance function,

$$\Sigma_\nu(\mathbf{u}, \boldsymbol{\theta}) = \mathbb{E}\left[\boldsymbol{\varphi}_\nu(W, u_1, \boldsymbol{\theta}_1)\boldsymbol{\varphi}_\nu(W, u_2, \boldsymbol{\theta}_2)'\right],$$

where $\boldsymbol{\varphi}_\nu \equiv \nu'_{\mathbf{S}}(\boldsymbol{\varphi})$, for each $\mathbf{u} = (u_1, u_2)' \in \mathcal{U} \times \mathcal{U}$, and $\boldsymbol{\theta} = (\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2)' \in \Theta^2 \times \Theta^2$.

Proof of Theorem 4.2. We establish Hadamard differentiability for each type of treatment effects and the desired result would then follow by a direct application of Lemma A.1. The cases for DTE and CHTE follow immediately from Lemma 3.9.25 in Van Der Vaart and Wellner (1996). Next, note that integration with respect to the Lebesgue measure is a linear operator, which implies that $\boldsymbol{\nu}_{ATE}(\cdot,\cdot)$ is linear in (s_{T_1}, s_{T_0}), and thus, is Hadamard differentiable by definition.

For QTE, it suffices to show that the mapping $q_{d,\cdot}(\cdot) : \ell_\infty(\tilde{\mathcal{T}} \times \Theta) \mapsto \ell_\infty((0,\bar{\tau}) \times \Theta)$ is Hadamard differentiable. The proof is similar to that of Lemma 3.9.23 in Van Der Vaart and Wellner (1996). Let $h_t \to h$ uniformly in $\ell_\infty(\tilde{\mathcal{T}} \times \Theta)$, with h being continuous. Thus, $s_{T_d} + th_t \in \ell_\infty(\tilde{\mathcal{T}} \times \Theta)$ for all $t > 0$. Let $q_{d,\theta,t}(\tau) \equiv \inf\{y : (s_{T_d} + th_t)(y,\theta) \leqslant 1 - \tau\}$. Due to $s_{T_d}(\cdot,\theta)$ and $(s_{T_d} + th_t)(\cdot,\theta)$ being restricted to $\tilde{\mathcal{T}}$ for each θ, it holds that $q_{d,\theta,t}(\tau), q_{d,\theta,t}(\tau) \in \tilde{\mathcal{T}}$, for all $(\tau,\theta) \in (0,\tau_o) \times \Theta$. By the definition of $q_{d,\theta,t}$, we have

$$1 - (s_{T_d} + th_t)(q_{d,\theta,t}(\tau) - \epsilon_{d,t,\theta}(\tau), \theta) \leqslant \tau \leqslant 1 - (s_{T_d} + th_t)(q_{d,\theta,t}(\tau), \theta),$$

where $\epsilon_{d,t,\theta}(\tau) = t^2 \wedge q_{d,\theta,t}(\tau) > 0$. Under Assumptions 6.2 and 6.4, $s_{T_d}(q_{d,\theta,t}(\tau) - \epsilon_{d,t,\theta}(\tau), \theta) = s_{T_d}(q_{d,\theta,t}(\tau), \theta) + O(\epsilon_{d,t,\theta}(\tau))$, uniformly in $(\tau,\theta) \in (0,\tau_o) \times \Theta$. This further implies that

$$-th(q_{d,\theta,t}(\tau) - \epsilon_{d,t,\theta}(\tau)) + r_t(\tau,\theta) \leqslant s_{T_d}(q_{d,\theta,t}(\tau),\theta) - s_{T_d}(q_{d,\theta}(\tau),\theta) \leqslant -th(q_{d,\theta,t}(\tau)) + r_t(\tau,\theta),$$

where $r_t(\tau,\theta) = o(t)$, uniformly in τ and θ. From the continuous differentiability of $s_{T_d}(y,\theta)$ in y and the derivative $-f_{T_d}(y,\theta)$ being uniformly bounded, we deduce that $|q_{d,\theta,t}(\tau) - q_{d,\theta}(\tau)| = O(t)$, uniformly in τ and θ. Applying Taylor expansion of the middle term in the preceding display allows us to conclude that $q_{d,\cdot}(\cdot)$ is Hadamard differentiable at $s_{T_d}(\cdot,\cdot)$, with derivative given by $h \mapsto h(q_{d,\cdot}(\cdot),\cdot)/f_{T_d}(q_{d,\cdot}(\cdot),\cdot)$, for $d \in \{0,1\}$. Another application of Lemma 3.9.25 in Van Der Vaart and Wellner (1996) yields the Hadamard differentiability of $\boldsymbol{\nu}_{QTE}(\cdot,\cdot)$, concluding our proof. ∎

A.3 Proofs for Results from Section 5

A.3.1 Weak Convergence of Multiplier Bootstrap Processes

Proof of Theorem 5.1. Proof of part (i) We first show that $\mathbb{G}_{n,\xi} \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}$. In view of Theorem 11.19 in Kosorok (2008), the conditional weak convergence follows under Assumption 11 if the triangular array $\{f_{ni}^x\}_{i=1}^n$, with $f_{ni}^x(t,\theta,d) = n^{-1/2}h^{1/2}\eta_{s,d}(W_i,x,t,\theta)$, satisfies the conditions of Lemma C.2. Note that we have verified these conditions in Corollary 1. Hence, the desired result holds. Next, we prove

$$\sup_{(t,\boldsymbol{\theta})\in\tilde{\mathcal{T}}\times\Theta^2} \left\| \hat{\mathbb{G}}_{\xi}^x(t,\boldsymbol{\theta}) - \mathbb{G}_{n,\xi}^x(t,\boldsymbol{\theta}) \right\| = o_p(1). \tag{A.30}$$

Decompose $\hat{\eta}_{s,d}(W,\xi,x,t,\theta)$ as

$$\begin{aligned}
&\frac{K_h(x\hat{\gamma}_d, X\hat{\gamma}_d) - K_h(x\gamma_d, X\gamma_d)}{f_d(x\gamma_d)}\xi\Psi_d\left(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}\right)(t,x,\theta)\\
&-\frac{\left(\hat{f}_d(x\hat{\gamma}_d) - f_d(x\gamma_d)\right)K_h(x\gamma_d, X\gamma_d)}{f_d(x\gamma_d)\hat{f}_d(x\hat{\gamma}_d)}\xi\Psi_d\left(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}\right)(t,x,\theta)\\
&+\frac{K_h(x\gamma_d, X\gamma_d)}{f_d(x\gamma_d)}\xi\left\{\hat{\Psi}_d\left(\hat{\mathcal{E}}_{d,\hat{\gamma}_d},\hat{\mathcal{E}}_{d,1,\hat{\gamma}_d}\right)(t,x,\theta) - \Psi_d\left(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}\right)(t,x,\theta)\right\}\\
&-\frac{\left(\hat{f}_d(x\hat{\gamma}_d) - f_d(x\gamma_d)\right)\left(K_h(x\hat{\gamma}_d, X\hat{\gamma}_d) - K_h(x\gamma_d, X\gamma_d)\right)}{f_d(x\gamma_d)\hat{f}_d(x\hat{\gamma}_d)}\xi\Psi_d\left(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}\right)(t,x,\theta)\\
&-\frac{\left(\hat{f}_d(x\hat{\gamma}_d) - f_d(x\gamma_d)\right)K_h(x\gamma_d, X\gamma_d)}{f_d(x\gamma_d)\hat{f}_d(x\hat{\gamma}_d)}\xi\left\{\hat{\Psi}_d\left(\hat{\mathcal{E}}_{d,\hat{\gamma}_d},\hat{\mathcal{E}}_{d,1,\hat{\gamma}_d}\right)(t,x,\theta) - \Psi_d\left(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}\right)(t,x,\theta)\right\}\\
&+\frac{K_h(x\hat{\gamma}_d, X\hat{\gamma}_d) - K_h(x\gamma_d, X\gamma_d)}{f_d(x\gamma_d)}\xi\left\{\hat{\Psi}_d\left(\hat{\mathcal{E}}_{d,\hat{\gamma}_d},\hat{\mathcal{E}}_{d,1,\hat{\gamma}_d}\right)(t,x,\theta) - \Psi_d\left(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}\right)(t,x,\theta)\right\}\\
&-\frac{\left(\hat{f}_d(x\hat{\gamma}_d) - f_d(x\gamma_d)\right)\left(K_h(x\hat{\gamma}_d, X\hat{\gamma}_d) - K_h(x\gamma_d, X\gamma_d)\right)}{f_d(x\gamma_d)\hat{f}_d(x\hat{\gamma}_d)}\\
&\cdot\xi\left\{\hat{\Psi}_d\left(\hat{\mathcal{E}}_{d,\hat{\gamma}_d},\hat{\mathcal{E}}_{d,1,\hat{\gamma}_d}\right)(t,x,\theta) - \Psi_d\left(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}\right)(t,x,\theta)\right\}\\
=&A_{n,1}(W,\xi) + A_{n,2}(W,\xi) + A_{n,3}(W,\xi) + A_{n,4}(W,\xi) + A_{n,5}(W,\xi) + A_{n,6}(W,\xi) + A_{n,7}(W,\xi).
\end{aligned}$$

The first three terms are obtained from the "first-order" expansion, the rate of which is dominating. Crude bounds based on uniform rates from Lemma B.2 can be utilized to control the remaining three terms.

By a Taylor expansion, we have

$$
\begin{aligned}
A_{n,1}(W,\xi)=&\frac{h^{-1}K_h^{(1)}(x'\gamma_d,X_i'\gamma_d)(x_{[-1]}-X_{[-1],i})'(\hat{\gamma}_d-\gamma_d)}{f_d(x\gamma_d)}\xi\Psi_d\left(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}\right)(t,x,\theta)\\
&+\frac{h^{-2}K_h^{(2)}(x'\tilde{\gamma}_d,X_i'\tilde{\gamma}_d)((x_{[-1]}-X_{[-1],i})'(\hat{\gamma}_d-\gamma_d))^2}{f_d(x\gamma_d)}\xi\Psi_d\left(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}\right)(t,x,\theta)\\
=&A_{n,11}(W,\xi)+A_{n,12}(W,\xi).
\end{aligned}
$$

Due to the independence of the bootstrap weights, $\mathbb{E}[A_{n,1}(X,\xi)|X]=0$, the empirical process $n^{-1/2}h^{1/2}\sum_{j=1}^n A_{n,11}(W_j,\xi_j)$ is centered, and by Lemma C.1, is of the order $O_p\left(\log n\cdot n^{-1/2}h^{-1}\right)=o_p(1)$, uniformly in t and θ. For the second term, using the uniform boundedness of $K^{(2)}$, $\mathcal{E}_{d,\gamma}$, and the compactness of $\mathcal{X}$, we deduce that $\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta}\left|n^{-1/2}h^{1/2}\sum_{j=1}^n A_{n,11}(W_j,\xi_j)\right|=o_p(n^{-1/2}h^{1/2})$.

Note that $n^{-1/2}h^{1/2}\sum_{j=1}^n A_{n,2}(W_j,\xi_j)$ can be bounded by

$$
\begin{aligned}
&n^{1/2}h^{1/2}\left|\frac{\hat{f}_d(x\hat{\gamma}_d)-f_d(x\gamma_d)}{f_d(x\gamma_d)\hat{f}_d(x\hat{\gamma}_d)}\right|\cdot\left|\frac{1}{n}\sum_{i=1}^n K_h(x\gamma_d,X_i\gamma_d)\xi\Psi_d\left(\mathcal{E}_{d,\gamma_d,i},\mathcal{E}_{d,1,\gamma_d,i}\right)(t,x,\theta)\right|\\
&=n^{1/2}h^{1/2}\,O_p\left((\log n)^{1/2}n^{-1/2}h^{-1/2}\right)\cdot O_p\left((\log n)^{1/2}n^{-1/2}h^{-1/2}\right)=O_p\left(\log n\cdot n^{-1/2}h^{-1/2}\right),
\end{aligned}
$$

which is $o_p(1)$ uniformly in t and θ.

From Lemma C.5 with $\ell=0$, we deduce that $\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta}\left|n^{-1/2}h^{1/2}\sum_{j=1}^n A_{n,3}(W_j,\xi_j)\right|=O_p\left(\log n\cdot n^{-1/2}h^{-1/2}\right)$. Next, for $s_1=4,5$, $n^{-1/2}h^{1/2}\sum_{j=1}^n A_{n,s_1}(W_j,\xi_j)$ are bounded by

$$
\left|\frac{\hat{f}_d(x\hat{\gamma}_d)-f_d(x\gamma_d)}{f_d(x\gamma_d)\hat{f}_d(x\hat{\gamma}_d)}\right|\cdot\left|n^{-1/2}h^{1/2}\sum_{j=1}^n A_{n,s_2}(W_j,\xi_j)\right|,
$$

for $s_2=1,3$, respectively, both of which converge uniformly at a rate of $O_p\left(\log n\cdot n^{-1}h^{-1}\right)$.

By a Taylor expansion of $K_h(x\hat{\gamma}_d,X\hat{\gamma}_d)$ around γ_d, and applying Lemma C.5 with $\ell=1$, we get $\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta}\left|n^{-1/2}h^{1/2}\sum_{j=1}^n A_{n,6}(W_j,\xi_j)\right|=\|\hat{\gamma}_d-\gamma_d\|\cdot O_p\left((\log n)^{1/2}n^{-1/2}h^{-3/2}\right)=O_p\left((\log n)^{1/2}n^{-1}h^{-3/2}\right)$. Lastly, $n^{-1/2}h^{1/2}\sum_{j=1}^n A_{n,7}(W_j,\xi_j)$ is bounded by

$$
\left|\frac{\hat{f}_d(x\hat{\gamma}_d)-f_d(x\gamma_d)}{f_d(x\gamma_d)\hat{f}_d(x\hat{\gamma}_d)}\right|\cdot\left|n^{-1/2}h^{1/2}\sum_{j=1}^n A_{n,6}(W_j,\xi_j)\right|,
$$

which converges at a rate of $O_p\left(\log n\cdot n^{-3/2}h^{-2}\right)$ uniformly over $\tilde{\mathcal{T}}\times\Theta$. Collecting the results on $A_{n,1}$ to $A_{n,7}$ implies that (A.30) holds.

To finish the proof, note that by the triangular inequality,

$$\left|\mathbb{E}_{\xi|w}[h(\hat{\mathbb{G}}^x_{n,\xi})] - \mathbb{E}[h(\mathbb{G}^x)]\right| \leqslant \left|\mathbb{E}_{\xi|w}[h(\hat{\mathbb{G}}^x_{n,\xi})] - \mathbb{E}_{\xi|w}[h(\mathbb{G}^x_{n,\xi})]\right| + \left|\mathbb{E}_{\xi|w}[h(\mathbb{G}^x_{n,\xi})] - \mathbb{E}[h(\mathbb{G}^x)]\right|,$$

for each $h \in BL_1$. The second term on the right hand side converges to zero since $\mathbb{G}^x_{n,\xi} \overset{p}{\underset{\xi}{\to}} \mathbb{G}^x$. By Jensen's inequality and the definition of BL_1, $\mathbb{E}_{\xi|w}[|h(\hat{\mathbb{G}}^x_{n,\xi}) - h(\mathbb{G}^x_{n,\xi})|] \leqslant 2\mathbb{P}_{\xi|w}\left(|\hat{\mathbb{G}}^x_{n,\xi} - \mathbb{G}^x_{n,\xi}| > \epsilon\right) + \epsilon$, for any $\epsilon \in (0,1)$. Due to the dominated convergence theorem, the first term on the right hand side goes to zero since $\hat{\mathbb{G}}^x_{n,\xi} - \mathbb{G}^x_{n,\xi} = o_p(1)$. Taking "sup" over BL_1 shows $\hat{\mathbb{G}}^x_{n,\xi} \overset{p}{\underset{\xi}{\to}} \mathbb{G}^x$.

Proof of part (ii) The proof is similar in structure to that of the first part. Thus, we provide a sketch of proof only. In view of Theorem 10.4 in Kosorok (2008), $\mathbb{G}_{n,\xi} \overset{p}{\underset{\xi}{\to}} \mathbb{G}$ provided $\mathcal{G}_\varphi$ as defined in (A.29) is a Donsker class. The latter condition is proved in Lemma C.3. In view of the discussion at the end of the last part, it remains to show that $\sup_{(t,\boldsymbol{\theta})\in\tilde{\mathcal{T}}\times\Theta^2}\left\|\hat{\mathbb{G}}_\xi(t,\boldsymbol{\theta}) - \mathbb{G}_{n,\xi}(t,\boldsymbol{\theta})\right\| = o_p(1)$.

We establish uniform convergence of $\xi\cdot\hat{\varphi}_{d,1}(W,t,\theta)$ first. Decompose the term as

$$\begin{aligned}
&\frac{\hat{f}(X\hat{\gamma}_d) - f(X\gamma_d)}{f_d(X\gamma_d)}\xi\Psi_d\left(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}\right)(t,X,\theta)\\
&-\frac{f(X\gamma_d)\left(\hat{f}_d(X\hat{\gamma}_d) - f(X\gamma_d,d)\right)}{f_d(X\gamma_d)\hat{f}_d(X\hat{\gamma}_d)}\xi\Psi_d\left(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}\right)(t,X,\theta)\\
&+\frac{f(X\gamma_d)}{f_d(X\gamma_d)}\xi\left\{\hat{\Psi}_d\left(\hat{\mathcal{E}}_{d,\hat{\gamma}_d},\hat{\mathcal{E}}_{d,1,\hat{\gamma}_d}\right)(t,X,\theta) - \Psi_d\left(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}\right)(t,X,\theta)\right\}\\
&-\frac{\left(\hat{f}(X\hat{\gamma}_d) - f(X\gamma_d)\right)\left(\hat{f}_d(X\hat{\gamma}_d) - f(X\gamma_d,d)\right)}{f_d(X\gamma_d)^2\hat{f}_d(X\hat{\gamma}_d)}\xi\Psi_d\left(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}\right)(t,X,\theta)\\
&+\frac{\hat{f}(X\hat{\gamma}_d) - f(X\gamma_d)}{f_d(X\gamma_d)}\xi\left\{\hat{\Psi}_d\left(\hat{\mathcal{E}}_{d,\hat{\gamma}_d},\hat{\mathcal{E}}_{d,1,\hat{\gamma}_d}\right)(t,X,\theta) - \Psi_d\left(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}\right)(t,X,\theta)\right\}\\
&-\frac{f(X\gamma_d,d)\left(\hat{f}_d(X\hat{\gamma}_d) - f(X\gamma_d,d)\right)}{f_d(X\gamma_d)\hat{f}_d(X\hat{\gamma}_d)}\xi\left\{\hat{\Psi}_d\left(\hat{\mathcal{E}}_{d,\hat{\gamma}_d},\hat{\mathcal{E}}_{d,1,\hat{\gamma}_d}\right)(t,X,\theta) - \Psi_d\left(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}\right)(t,X,\theta)\right\}\\
&-\frac{\left(\hat{f}_d(X\hat{\gamma}_d) - f(X\gamma_d,d)\right)\left(\hat{f}_d(X\hat{\gamma}_d) - f(X\gamma_d,d)\right)}{f_d(X\gamma_d)\hat{f}_d(X\hat{\gamma}_d)}\xi\\
&\quad\cdot\left\{\hat{\Psi}_d\left(\hat{\mathcal{E}}_{d,\hat{\gamma}_d},\hat{\mathcal{E}}_{d,1,\hat{\gamma}_d}\right)(t,X,\theta) - \Psi_d\left(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}\right)(t,X,\theta)\right\}\\
=&A_{n,1}(W,\xi) + A_{n,2}(W,\xi) + A_{n,3}(W,\xi) + A_{n,4}(W,\xi) + A_{n,5}(W,\xi) + A_{n,6}(W,\xi) + A_{n,7}(W,\xi).
\end{aligned}$$

It can be shown, via direct analysis in the case of $A_{n,1}$ and $A_{n,2}$ or by further decomposition à la Lemma C.5 for $A_{n,3}$, that the first three terms are dominated by second order degenerate U processes. Therefore, we can deduce from Lemma C.1 that

$$\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta}\left|n^{-1/2}\sum_{j=1}^{n}A_{n,\ell}(W_j,\xi_j)\right| = O_p\left(\log n\cdot n^{-1/2}h^{-1/2}\right),$$

for $\ell = 1,2,3$. Similar analysis implies that the $A_{n,4}$-$A_{n,6}$ and $A_{n,7}$ are governed by third order degenerate U processes and a fourth order degenerate U processes, respectively, with

$$\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta}\left|n^{-1/2}\sum_{j=1}^{n}A_{n,\ell}(W_j,\xi_j)\right| = O_p\left((\log n)^{3/2}n^{-1}h^{-1}\right),$$
$$\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta}\left|n^{-1/2}\sum_{j=1}^{n}A_{n,7}(W_j,\xi_j)\right| = O_p\left((\log n)^{2}n^{-3/2}h^{-3/2}\right),$$

for $\ell = 4,5,6$. We therefore conclude that

$$\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta}\left|n^{-1/2}\sum_{j=1}^{n}\left\{\hat{\varphi}_{d,\xi,1}(W_j,\xi_j,t,\theta)-\varphi_{d,\xi,1}(W_j,\xi_j,t,\theta)\right\}\right| = O_p\left(\log n\cdot n^{-1/2}h^{-1/2}\right) = o_p(1),$$

for $d = 0,1$. Regarding $\xi\hat{\varphi}_{d,2}$, we note that

$$\begin{aligned}
&n^{-1/2}\sum_{j=1}^{n}\left\{\hat{\varphi}_{d,\xi,2}(W_j,\xi_j,t,\theta)-\varphi_{d,\xi,2}(W_j,\xi_j,t,\theta)\right\}\\
=&n^{-1/2}\sum_{j=1}^{n}\xi_j\left\{\hat{s}_{T,d}(t,X_j\hat{\gamma}_d,\theta)-s_{T_d}(t,X_j\gamma_d,\theta)\right\}\\
&+n^{1/2}\,\mathbb{E}_n[\xi]\cdot\left\{\mathbb{E}_n[\hat{s}_{T,d}(t,X\hat{\gamma}_d,\theta)]-\mathbb{E}[s_{T_d}(t,X\gamma_d,\theta)]\right\}.
\end{aligned}$$

The term in the second line can be analyzed as in the previous part. Due to Assumption 11, we have that $n^{1/2}\,\mathbb{E}_n[\xi] = O_p(1)$. The uniform convergence of $\hat{s}_{T,d}(t,X\hat{\gamma}_d,\theta)$ to $s_{T_d}(t,X\gamma_d,\theta)$ as proved in Theorem 4.1, implies that $\mathbb{E}_n[\hat{s}_{T,d}(t,X\hat{\gamma}_d,\theta)]-\mathbb{E}[s_{T_d}(t,X\gamma_d,\theta)] = o_p(1)$, uniformly over $\tilde{\mathcal{T}}\times\Theta$. Combining the two results, we deduce that the last term in the previous display is also $o_p(1)$, concluding the proof. ∎

Proof of Corollary 3. Since integration with respect to the Lebesgue measure is a linear, and thus, Lipschitz continuous, mapping, the results for the ATE then follow from Theorem 5.1 and the continuous mapping theorem for multiplier bootstrap, cf. Proposition 10.7 in Kosorok (2008). The case for the DTE also follows trivially from the continuous mapping theorem. Now, let us consider the QTE. Let $\Psi^x_{n,d}(t,\theta)\equiv n^{-1/2}h^{1/2}\sum_{i=1}^{n}\xi_i\cdot\hat{\eta}_{s,d}(W_i,x,t,\theta)$. Using this quantity, we write

$$\sup_{(\tau,\theta)\in(0,\tau_o)\times\Theta}\left|n^{-1/2}h^{1/2}\sum_{i=1}^{n}\xi\left(\hat{\psi}^x_{d,QTE}(W,\tau,\theta)-\frac{\hat{\eta}_{s,d}(W,x,q^x_{d,\theta}(\tau),\theta)}{f_{T_d,x}(q^x_{d,\theta}(\tau),\theta)}\right)\right| \tag{A.31}$$

$$
\begin{aligned}
&= \sup_{(\tau,\theta)\in(0,\tau_o)\times\Theta} \left| \frac{\Psi^x_{n,d}(\hat{q}^x_{d,\theta}(\tau),\theta)}{\hat{f}_{T_d,x}(\hat{q}^x_{d,\theta}(\tau),\theta)} - \frac{\Psi^x_{n,d}(q^x_{d,\theta}(\tau),\theta)}{f_{T_d,x}(q^x_{d,\theta}(\tau),\theta)} \right| \\
&\leqslant \sup_{(\tau,\theta)\in(0,\tau_o)\times\Theta} \left| \frac{\Psi^x_{n,d}(\hat{q}^x_{d,\theta}(\tau),\theta)}{\hat{f}_{T_d,x}(\hat{q}^x_{d,\theta}(\tau),\theta)} - \frac{\Psi^x_{n,d}(\hat{q}^x_{d,\theta}(\tau),\theta)}{f_{T_d,x}(q^x_{d,\theta}(\tau),\theta)} \right| \\
&\quad + \sup_{(\tau,\theta)\in(0,\tau_o)\times\Theta} \left| \frac{\Psi^x_{n,d}(\hat{q}^x_{d,\theta}(\tau),\theta)}{f_{T_d,x}(q^x_{d,\theta}(\tau),\theta)} - \frac{\Psi^x_{n,d}(q^x_{d,\theta}(\tau),\theta)}{f_{T_d,x}(q^x_{d,\theta}(\tau),\theta)} \right| \\
&\lesssim \sup_{(\tau,\theta)\in(0,\tau_o)\times\Theta} \left| \Psi^x_{n,d}(\hat{q}^x_{d,\theta}(\tau),\theta) \right| \sup_{(\tau,\theta)\in(0,\tau_o)\times\Theta} \left| \hat{f}_{T_d,x}(\hat{q}^x_{d,\theta}(\tau),\theta) - f_{T_d,x}(q^x_{d,\theta}(\tau),\theta) \right| \\
&\quad + \sup_{(\tau,\theta)\in(0,\tau_o)\times\Theta} \left| \Psi^x_{n,d}(\hat{q}^x_{d,\theta}(\tau),\theta) - \Psi^x_{n,d}(q^x_{d,\theta}(\tau),\theta) \right|.
\end{aligned}
$$

From the fact that $\hat{q}^x_{d,\theta}$ converges uniformly to $q^x_{d,\theta}$, and the definition of τ_o, we deduce that $\sup_{\tau\in(0,\tau_o)} \hat{q}^x_{d,\theta}(\tau) \leqslant y_o$ with probability approaching one. Note that,

$$
\begin{aligned}
&\sup_{(\tau,\theta)\in(0,\tau_o)\times\Theta} \left| \hat{f}_{T_d,x}(\hat{q}^x_{d,\theta}(\tau),\theta) - f_{T_d,x}(q^x_{d,\theta}(\tau),\theta) \right| \\
&\leqslant \sup_{(\tau,\theta)\in(0,\tau_o)\times\Theta} \left| \hat{f}_{T_d,x}(\hat{q}^x_{d,\theta}(\tau),\theta) - f_{T_d,x}(\hat{q}^x_{d,\theta}(\tau),\theta) \right| + \sup_{(\tau,\theta)\in(0,\tau_o)\times\Theta} \left| f_{T_d,x}(\hat{q}^x_{d,\theta}(\tau),\theta) - f_{T_d,x}(q^x_{d,\theta}(\tau),\theta) \right| \\
&\leqslant \sup_{y\in\tilde{\mathcal{T}}} \left| \hat{f}_{T_d,x}(y,\theta) - f_{T_d,x}(y,\theta) \right| + M_1 \sup_{(\tau,\theta)\in(0,\tau_o)\times\Theta} \left| \hat{q}^x_{d,\theta}(\tau) - q^x_{d,\theta}(\tau) \right| = o_p(1).
\end{aligned}
$$

Regarding the second inequality, the first term follows by the observation above the display and Assumption 12, while the second term follows by the continuous differentiability of $f_{T_d,x}$, which implied by Assumption 6.2 and Theorem 3.2. The last equality is due to Assumption 12 and the fact that $\hat{q}^x_{d,\theta}(\tau)$ converges to $q^x_{d,\theta}(\tau)$ uniformly over $(0,\tau_o)\times\Theta$.

We have shown, in Theorem 5.1, that $\Psi^x_{n,d}(\cdot,\cdot)$ converges weakly to a centered Gaussian process. It follows immediately, by Prokhorov's theorem and the fact that $\sup_{\tau\in(0,\tau_o)} \hat{q}^x_{d,\theta}(\tau) \leqslant y_o$ with probability approaching one, that $\sup_{(\tau,\theta)\in(0,\tau_o)\times\Theta} \left| \Psi^x_{n,d}(\hat{q}^x_{d,\theta}(\tau),\theta) \right| = O_p(1)$. Moreover, since $\Psi^x_{n,d}(\cdot,\cdot)$ is equicontinuous and that $\hat{q}^x_{d,\theta}$ converges uniformly to $q^x_{d,\theta}$, we get that, conditional on the sample path $\{W_i\}_{i=1}^n$, $\sup_{(\tau,\theta)\in(0,\tau_o)\times\Theta} \left| \Psi^x_{n,d}(\hat{q}^x_{d,\theta}(\tau),\theta) - \Psi^x_{n,d}(q^x_{d,\theta}(\tau),\theta) \right| = o_p(1)$.

Collecting the results, we deduce that (A.31) $\overset{p}{\underset{\xi}{\to}} 0$. In addition, by same lines of reasoning as in the proof for Theorem 5.1(i), it is straightforward to show that $\Psi^x_{n,d}(q^x_{d,\cdot}(\cdot),\cdot)/f_{T_d,x}(q^x_{d,\cdot}(\cdot),\cdot) \overset{p}{\underset{\xi}{\to}} \nu'_{QTE,\boldsymbol{S}^x}(\mathbb{G}^x)(\cdot,\cdot)$. This completes the proof for $\hat{\mathbb{G}}^x_{\xi,QTE}$.

To conclude, we note that, the proof for the unconditional QTE and for the DTE's will follow by largely parallel analyses, and thus, we omit it. ■

A.3.2 Bootstrap Confidence Sets

In this section, we first describe an algorithm for constructing uniform confidence sets for conditional TEBFs. Then, we validate the resulting confidence sets by proving Theorem 5.2. Let $\hat{\mathbb{G}}^x_{lb,\xi,j}$ and $\hat{\mathbb{G}}^x_{ub,\xi,j}$ denote the first and second component of $\hat{\mathbb{G}}^x_{\xi,j}$, respectively.

Algorithm 2 1. Same as Step 1 of Algorithm 2. In Steps 2-5, the calculations will be performed for $d, r \in \{0,1\}$, $t \in \tilde{\mathcal{T}}$, $\tau \in (0, \tau_o)$, $\boldsymbol{\theta} \in \boldsymbol{\Theta}_l$, and $u \in \mathcal{U}_m$.

2. Estimate $\hat{\gamma}_d$, $\hat{G}_{d,1}(t, x\hat{\gamma}_d)$, $\hat{G}_d(t, x\hat{\gamma}_d)$, $\hat{s}_{T_d}(t, x\hat{\gamma}_d, \theta)$, following (4.2), (4.3), (4.4), and (4.5). If $j = QTE$, compute $\hat{q}^x_{d,\theta}(\tau)$ and $\hat{f}_{T_d,x}(t,\theta)$, following (C.2).

3. Calculate $\hat{\nu}^x_j(u, \boldsymbol{\theta})$, $\hat{\eta}_{s,d}(W, x, t, \theta)$, and $\hat{\psi}^x_j(W, \theta)$, based on (5.1) and (5.6) - (5.9), respectively.

4. Sample $\{\xi^b_i\}^n_{i=1}$ from a distribution with zero mean and unit variance, independently from data. Calculate $\hat{\boldsymbol{\psi}}^{x*}_j$, and $\mathbb{G}^x_{\xi^b,j}(u, \theta)$.

   Repeat Step 4 for $b = 1, ..., B$, where B is some large integer.

5. For $\ell = lb, ub$, compute the $(1-\alpha)$-th quantile $\hat{c}^{x,B}_{n,\ell,j}(\alpha, \mathcal{U}_m, \boldsymbol{\Theta}_l)$ of $\left\{\max_{1 \leqslant i \leqslant m, 1 \leqslant s \leqslant l} \left\|\mathbb{G}^x_{\ell,\xi^b,j}(u_i, \boldsymbol{\theta})\right\|\right\}^B_{b=1}$, and construct the uniform confidence band

$$C^{x,B}_{n,\ell,j}(1-\alpha, \mathcal{U}_m, \boldsymbol{\Theta}_l) \equiv \left\{\hat{\nu}^x_{\ell,j}(u, \boldsymbol{\theta}) \pm n^{-1/2} h^{-1/2} \hat{c}^{x,B}_{n,\ell,j}(\alpha, \mathcal{U}_m, \boldsymbol{\Theta}_l) : u \in \mathcal{U}_m, \boldsymbol{\theta} \in \boldsymbol{\Theta}_l\right\}.$$

Proof of Theorem 5.2. The proof is a direct consequence of Theorem 4.2, Corollary 3, and the continuous mapping theorem for the multiplier bootstrap, cf. Theorem 2.6 in Kosorok (2008). ■

Appendix B Single-Index Estimator

In this section, we establish large sample properties of the index coefficients estimator $\hat{\boldsymbol{\gamma}}$. The results, as presented in the following lemma, are largely based on Proposition 1 in Li and Patilea (2018). We show that $\hat{\boldsymbol{\gamma}}$ is consistent for $\boldsymbol{\gamma}$, converges to $\boldsymbol{\gamma}$ at the parametric rate. Moreover, $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ admits an asymptotic linear representation, and converges in distribution to a normal distribution.

Lemma B.1 *Under Assumptions 1, 3, 4.1, 5, 6.1 - 6.3, 7, 8.1 and 9.1, it holds that*

$$\hat{\gamma}_d - \gamma_d = \frac{1}{n}\sum_{i=1}^n V_d^{-1} \psi^b_d(W_i) + o_p\left(n^{-1/2}\right), \tag{B.1}$$

for $d \in \{0,1\}$, *where* ψ_d^a *and* V_d *are defined in Section 4.3. Furthermore,*

$$\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) \xrightarrow{d} N(0, \Sigma_{\boldsymbol{\gamma}}),$$

where

$$\Sigma_{\boldsymbol{\gamma}} \equiv \begin{pmatrix} \Sigma_{\gamma_1} & \mathbf{0} \\ \mathbf{0} & \Sigma_{\gamma_0} \end{pmatrix},$$

and $\Sigma_{\gamma_d} \equiv V_d^{-1}\mathbb{E}[\psi_d^b(W)\psi_d^b(W)']V_d^{-1}$.

Proof of Lemma B.1. The linear expansion follows from the same lines of argument as in that of Proposition 1 in Li and Patilea (2018). We show how their Assumption 8.1 can be fulfilled by parallel conditions in our context. Condition (1) is satisfied under Assumptions 5.1 and 3.1. Condition (2) holds under Assumptions 4.1 and 5.2. Conditions (3) and (4) follow from Theorem 3.1, which holds under Assumptions 1 - 3. Condition (5) is satisfied under 3.1. Condition (6) is due to Assumptions 6.1 - 6.3. Lastly, Assumptions 7.2, 7.1, 8.1 imply Conditions (7) - (9), respectively. We remark that Assumption 9.1 is slightly weaker than Condition (10). However, their proof carries through, under this weaker condition, if the maximal inequality from Lemma C.1, instead of the main corollary of Sherman (1994), is employed in the proof.

To conclude, we note that $\psi_1^b \cdot \psi_0^b = 0$, and thus the covariance between $\hat{\gamma}_1 - \gamma_1$ and $\hat{\gamma}_0 - \gamma_0$ is 0. ■

B.1 Single-Index Kernel Estimator

In this section, we present a lemma documenting some well-known facts on single-index kernel estimators. These results will be used repeated throughout the appendix. First, we introduce some quantities,

$$K^{(1)}(u) \equiv dK(u)/du, \quad K^{(2)}(u) \equiv d^2K(u)/du^2, \quad K_h^{(j)}(u,v) \equiv K^{(j)}((u-v)/h)/h, \text{ for } j = 1,2,$$
$$G_d^{(1)}(y, x\gamma) \equiv \left\{\varrho_{1,1}^{\gamma}/\varrho_{0,0}^{\gamma} - \varrho_{1,0}^{\gamma}\varrho_{0,1}^{\gamma}/{\varrho_{0,0}^{\gamma}}^2\right\}(y, x\gamma), \tag{B.2}$$

and analogous definition for the sub-distributions should be apparent. We remark that $G_d^{(1)}(y, x\gamma)$ is, in general, not equal to $\partial_\gamma G_d(y, x\gamma)$. When $\gamma = \gamma_d$, the expression simplifies, and we find, by direct calculations, that $G_d^{(1)}(y, x\gamma_d) = \partial_{x\gamma}G_d(y, x\gamma_d) \cdot \mathbb{E}\left[(x - X)|X\gamma_d = x\gamma_d\right]$.

In the following lemma, we provide convergence rates for single-index kernel density and conditional distribution estimators. Let $\mathcal{H} \equiv [h_l n^{-\zeta}, h_u]$, for some positive number h_l, h_u, and $0 < \zeta < 1/3$.

Lemma B.2 *Suppose Assumptions 3.1, 5, 4.1, 6.3, 8.2, and 9.2 hold, then*

$$\sup_{(y,x,h,\gamma)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\mathcal{H}\times\Gamma}\left|\hat{f}_d(x\gamma)-\varrho_{0,0}(y,x\gamma)\right| = O_p\left((\log n)^{1/2}n^{-1/2}h^{-1/2}\right)+O(h^s), \tag{B.3}$$

$$\sup_{(y,x,h,\gamma)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\mathcal{H}\times\Gamma}\left|\hat{\kappa}_{d,y}(x\gamma)-\varrho_{1,0}(y,x\gamma)\right| = O_p\left((\log n)^{1/2}n^{-1/2}h^{-1/2}\right)+O(h^s), \tag{B.4}$$

$$\sup_{(y,x,h,\gamma)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\mathcal{H}\times\Gamma}\left|\hat{G}_d(y,x\gamma)-G_d(y,x\gamma)\right| = O_p\left((\log n)^{1/2}n^{-1/2}h^{-1/2}\right)+O(h^s) \tag{B.5}$$

$$\sup_{(y,x,h,\gamma)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\mathcal{H}\times\Gamma}\left\|\partial_\gamma\hat{G}_d(y,x\gamma)-G_d^{(1)}(y,x\gamma)\right\| = O_p\left((\log n)^{1/2}n^{-1/2}h^{-3/2}\right)+O(h^s), \tag{B.6}$$

for $\ell=1,2$. Analogous results hold for $\hat{G}_{d,r}$, and $\partial_\gamma\hat{G}_{d,r}$, with $r=0,1$.

Proof of Lemma B.2. The proof follows from standard kernel techniques, cf. Einmahl and Mason (2005). It is also implicit in that of Theorem 1 in Chiang and Huang (2012), and hence, is omitted. ■

Appendix C Auxiliary Results

C.1 Definitions and Additional Results

In this section, we provide several results related to the results in the main text. First, we introduce the notion of *covering numbers* and the *VC type* (or *Euclidean*) class. Let $\|\cdot\|_{L^r(P)}$ denote $\{\mathbb{E}[|f(W)|^r]\}^{1/r}$ Given a class of functions $\mathcal{F}$ defined on a space $\mathcal{X}$, a probability measure Q, the covering number $\mathcal{N}(\epsilon,\mathcal{F},L_r(Q))$, is the minimum number of $L_r(Q)$ balls of radius ϵ needed to cover $\mathcal{F}$. The centers of these balls is not required to be in $\mathcal{F}$. A function $F:\mathcal{X}\mapsto\mathbb{R}$ is called an envelop for $\mathcal{F}$ if $|f(x)|\leqslant F(x)$ for all $x\in\mathcal{X}$ and all $f\in\mathcal{F}$. We say $\mathcal{F}$ is of the VC type with respect to an envelop function F if there exists positive constants A and $V\geqslant 1$ satisfying $\sup_Q\mathcal{N}(\epsilon\|F\|_{L_2(Q)},\mathcal{F},L_2(Q))\leqslant(A\|F\|_{L_2(Q)}/\epsilon)^V$, for all $\epsilon\in(0,2\|F\|_{L_2(Q)}]$ and the supremum is taken over all probability measures Q with $0<\|F\|_{L_2(Q)}<\infty$. We say that the VC class $\mathcal{F}$ admits the characteristics A and V.

Given a function g of m variables, let $U_n^{(m)}(g)=\frac{(n-k)!}{n!}\sum_{i\in I_n^m}g(X_{i_1},...,X_{i_m})$, with $I_n^m=\{(i_1,...,i_k):1\leqslant i_j\leqslant n, i_j\neq i_l, if j\neq l\}$. Let the *Hoeffding projections* of g with respect to a measure P be defined as $\pi_k g=(\delta_{x_1}-P)\times\cdots\times(\delta_{x_k}-P)\times P^{m-k}g$ and $\pi_0=\mathbb{E}[g(X_1,...,X_m)]$. If g is symmetric in its entries, we define the *Hoeffding decomposition* as $U_n^{(m)}(g)-\mathbb{E}[g]=\sum_{j=1}^m\binom{m}{j}U_n^{(j)}(\pi_j g)$. The following Lemma, due to Giné and Mason (2007), establishes a maximal inequality for moment of the U processes which plays a crucial role in our derivation of the uniform linear representations.

Lemma C.1 (Giné and Mason, 2007, Theorem 8) *Let $\mathcal{F}$ be a measurable collection of symmetric functions $S^m \mapsto \mathbb{R}$ with an envelop function F and let P be any probability measure on the space $(S, \mathcal{S})$. Assume F is bounded by $M > 0$ and $\mathcal{F}$ is a VC class with respect to F with characteristics A and V. Then for every $m \in \mathbb{N}$, $A \geqslant e^m$, $V \geqslant 1$, there exist constants C_1 and C_2 such that*

$$n^{k/2}\,\mathbb{E}\left[\left\|U_n^{(k)}(\pi_k f)\right\|_{\mathcal{F}}\right] \leqslant C_1 \sigma \left(\log \frac{A\,\|F\|_{L_2(P^m)}}{\sigma}\right)^{k/2},$$

for $k = 0, 1, \ldots, m$, assuming $n\sigma^2 \leqslant C_2 \log\left(2\,\|F\|_{L_2(P^m)}/\sigma\right)$, where σ^2 satisfies $\|P^m f^2\| \leqslant \sigma^2 \leqslant P^m F^2$.

Let inner and outer expectations be denoted by $\mathbb{E}_*$ and $\mathbb{E}^*$ as in Section 1.2 of Van Der Vaart and Wellner (1996). We say a sequence of stochastic process $X_n : \mathbb{E} \mapsto \mathbb{D}$, where $\mathbb{E}$ and $\mathbb{D}$ are metric spaces, *converges weakly* to X, denoted by $X_n \Rightarrow X$ if $\mathbb{E}^*[h(X_n)] \to \mathbb{E}[h(X)]$, for all $h \in \mathcal{C}_b(\mathbb{D})$ where $\mathcal{C}_b(\mathbb{D})$ denotes the space of the real-valued bounded continuous functions defined on $\mathbb{D}$.

We follow the definition of *conditional weak convergence in probability* as appeared in Section 2.2.3 in Kosorok (2008). The notation $X_n \underset{\xi}{\overset{p}{\rightsquigarrow}} X$ means that $\sup_{h \in BL_1} \left|\mathbb{E}_{\xi|w}[h(X_n)] - \mathbb{E}[h(X)]\right| \xrightarrow{p} 0$ and $\mathbb{E}^*_{\xi|w}[h(X_n)] - \mathbb{E}_{\xi|w_*}[h(X_n)] = 0$, where BL_1 is the space of functions $f : \mathbb{D} \mapsto \mathbb{R}$ with Lipschitz norm bounded by 1. Namely, $\|f\|_\infty \leqslant 1$ and $|f(x) - f(y)| \leqslant d(x, y)$, for $x, y \in \mathbb{D}$. The operator $\mathbb{E}_{\xi|w}$ denotes the conditional expectation over the weights ξ given the remaining data.

The following lemma, originated from Theorem 10.6 of Pollard (1990) and restated in Theorem 11.16 of Kosorok (2008) is key to establishing weak convergence of conditional processes.

Lemma C.2 (Kosorok, 2008, Theorem 11.16) *Suppose a triangular array stochastic processes $\{f_{ni}(t) : i = 1, \ldots, n, t \in T\}$ consisting of row-wise independent processes is almost measurable Suslin (AMS). Define $\chi_n(t) = \sum_{i=1}^n f_{ni}(t)$ and $\rho_n(s,t) = \left(\sum_{i=1}^n \mathbb{E}\left[(f_{ni}(s) - f_{ni}(t))^2\right]\right)^{1/2}$, for $s, t \in T$. (i) the $\{f_{ni}\}$ are manageable, with envelops $\{F_{ni}\}$ which are also independent within rows; (ii) $H(s,t) = lim_{n\to\infty} \sum_{i=1}^n \mathbb{E}[\chi_n(s)\chi_n(t)]$ exists for every $s, t \in T$; (iii) $\limsup_{n\to\infty} \mathbb{E}^*[F_{ni}^2] < \infty$; (iv) $\lim_{n\to\infty} \sum_{i=1}^n \mathbb{E}^*\left[F_{ni}^2 \mathbb{1}\{F_{ni}\} > \epsilon\right] = 0$; (v) $\rho(s,t) = \lim_{n\to\infty} \rho_n(s,t)$ exists for every $s, t \in T$. For all deterministic sequences $\{s_n\}$ and $\{t_n\}$, if $\rho(s_n, t_n) \to 0$ then $\rho_n(s_n, t_n) \to 0$. Then T is totally bounded under the ρ pseudometric and χ_n converges weakly on $\ell_\infty(T)$ to a tight, centered Gaussian process χ concentrated on $\{g \in \ell_\infty(T) : g$ is uniformly ρ-continuous$\}$, with covariance function $H(s,t)$.*

Precise definitions of AMS and manageable triangular arrays can be found in Section 11.4.1 in Kosorok (2008). Direct check of these two conditions is usually not easy. To address this issue, Kosorok (2008) presents sufficient conditions: by Lemma 11.15 in Kosorok (2008), the triangular array is AMS whenever it is separable[12], and for manageability to hold, Lemma 11.21 in Kosorok (2008) implies that the VC type condition on the triangular array suffices.

Lastly, we recall the definition of *Hadamard differentiability*, see pp. 272-273 in Van Der Vaart and Wellner (1996). We say a mapping $\nu : \mathbb{D}_\nu \subset \mathbb{D} \to \mathbb{E}$ is called *Hadamard differentiable* at $F \in \mathbb{D}_\nu$, tangentially to a set $\mathbb{D}_0 \subset \mathbb{D}$, if there is a continuous linear map $\nu_F' : \mathbb{D} \to \mathbb{E}$ such that

$$\frac{\nu(F + t_n h_n) - \nu(F)}{t_n} \to \nu_F'(h),$$

for all converging sequences $\{t_n\} \subset \mathbb{R}$ with $t_n \to 0$ and $\{h_n\} \subset \mathbb{D}$ with $h_n \to h \in \mathbb{D}_0$, such that $F + t_n h_n \in \mathbb{D}_\nu$ as $n \to \infty$, for all n.

C.2 Auxiliary Lemmas

Lemma C.3 *Suppose that the assumptions of Theorem 4.1 hold. The function classes, $\mathcal{G}_1$ - $\mathcal{G}_6$, $\mathcal{G}_b$, $\mathcal{G}_\eta$, and $\mathcal{G}_\varphi$ as defined in* (A.19), (A.20),(A.22), (A.23), (A.27), (A.28), (A.24), (A.25), *and* (A.29) *are of VC type with bounded envelop.*

Proof of Lemma C.3. We first identify the sub-classes that constitute the above functional classes and show that the uniform entropy condition is satisfied for each of these sub-classes. Then, we illustrate on how we use results on the sub-classes to show that the functional classes in the theorem is of VC type. Define

$$\begin{aligned}
\mathcal{M}_1 &\equiv \{y \mapsto \mathbb{1}\{y \leqslant t\} : t \in \tilde{\mathcal{T}}\},\\
\mathcal{M}_2 &\equiv \{x_1 \mapsto K((x_1\gamma_d - x\gamma_d)/h)\mathbb{1}\{|x_1\gamma_d - x\gamma_d| \leqslant h\} : (x,h) \in \mathcal{X} \times \mathcal{H}\},\\
\mathcal{M}_3 &\equiv \{x_1 \mapsto K^{(1)}((x_1\gamma_d - x\gamma_d)/h)\mathbb{1}\{|x_1\gamma_d - x\gamma_d| \leqslant h\} : (x,h) \in \mathcal{X} \times \mathcal{H}\},\\
\mathcal{M}_{4,1} &\equiv \{y \mapsto \partial^\ell_{x\gamma} G_d(y, x\gamma_d) : \ell \in \{0,1,2,...,s\}, (d,x) \in \{0,1\} \times \mathcal{X}\},\\
\mathcal{M}_{4,2} &\equiv \{y \mapsto \partial^\ell_{x\gamma} G_{d,1}(y, x\gamma_d) : \ell \in \{0,1,2,...,s\}, (d,x) \in \{0,1\} \times \mathcal{X}\}.\\
\mathcal{M}_{4,3} &\equiv \{\partial_y G_d(t, x\gamma_d) : (d,t,x) \in \{0,1\} \times \tilde{\mathcal{T}} \times \mathcal{X}\}.\\
\mathcal{M}_{4,4} &\equiv \{\partial_y G_{d,1}(t, x\gamma_d) : (d,t,x) \in \{0,1\} \times \tilde{\mathcal{T}} \times \mathcal{X}\}.\\
\mathcal{M}_5 &\equiv \{\partial^\ell_{x\gamma} f_d(x\gamma_d) : \ell \in \{0,1,2,...,s\}, (d,x) \in \{0,1\} \times \mathcal{X}\}.
\end{aligned}$$

12 A triangular array of stochastic process $\{f_{ni}(t) : i = 1,...,n, t \in T\}$ is separable if, for all $n \geqslant 1$, there exists a countable set $T_n \in T$ such that

$$\mathbb{E}^*\left[\mathbb{1}\left\{\sup_{t\in T}\sup_{s\in T_n}\sum_{i=1}^n (f_{ni}(s) - f_{ni}(t))^2 > 0\right\}\right] = 0.$$

By Lemma 19.15 in Van der Vaart (1998), $\mathcal{M}_1$ is of VC type with the constant envelop. Under Assumption 8.2, both $K(\cdot)$ and $K^{(1)}(\cdot)$ are of bounded variation, Lemma 22(i) of Nolan and Pollard (1987) implies that $\mathcal{M}_{2,1}$ and $\mathcal{M}_{2,2}$ belong to the VC class with a constant envelop. Next, since $\partial_v^{\ell} F_{Y_d,R_d|D,X\gamma_d}(y,r|d,v)$, $\ell = 0,...,s$, is Lipschitz continuous with respect to $x\gamma_d$ under Assumption 6.2(i), Lemma 2.13 of Pakes and Pollard (1989) implies $\mathcal{M}_{4,1}$ and $\mathcal{M}_{4,2}$ are of VC type with bounded envelop functions. The proof for $\mathcal{M}_{4,3}, \mathcal{M}_{4,4}$, and $\mathcal{M}_5$ follows the same arguments based on the Lipschitz continuity of $\partial_y F_{Y_d,R_d|D,X\gamma_d}(y,r|d,v)$ with respect to y and v, and of $\partial_v f_{d,\gamma_d}(v)$ with respect to v, as implied by Assumption 6.2(iv), and 6.1(i), respectively.

Now we are ready to show why the functional classes in the lemma are of VC type. We illustrate on $\mathcal{G}_1$ and $\mathcal{G}_\eta$. All others follow by same lines of reasoning.

We focus on $\mathcal{G}_1$ first. Note that the class that g_{11} belongs to is a product of a finite set $\{(r_1,d_1) \mapsto r_1 \mathbb{1}\{d_1 = d\}, d \in \{0,1\}\}$, $\mathcal{M}_1$, and $\mathcal{M}_2$, and thus, it is of VC type by Corollary A.1 in Chernozhukov, Chetverikov, and Kato (2014). Since all three sub-classes have finite envelops, their product also does. Regarding g_{12}, we first show that $\mathcal{M}_\phi \equiv \{y \mapsto \phi''_\theta(s_d(y, x\gamma_d)) : (x,\theta) \in \mathcal{X} \times \Theta\}$ is also a VC class with bounded envelop. For any $x_1, x_2 \in \mathcal{X}$ and $\theta_1, \theta_2 \in \Theta$, we have

$$
\begin{aligned}
&\left|\phi''_{\theta_1}(s_d(y,x_1\gamma_d)) - \phi''_{\theta_2}(s_d(y,x_2\gamma_d))\right| \\
\leqslant& \left|\phi''_{\theta_1}(s_d(y,x_1\gamma_d)) - \phi''_{\theta_2}(s_d(y,x_1\gamma_d))\right| + \left|\phi''_{\theta_2}(s_d(y,x_1\gamma_d)) - \phi''_{\theta_2}(s_d(y,x_2\gamma_d))\right| \\
\leqslant& M_1 |\theta_1 - \theta_2| + \sup_{(\theta,u)\in\Theta\times[v_o,1]} \left|\phi'''_\theta(u)\right| \sup_{(y,x)\in\tilde{\mathcal{T}}\times\mathcal{X}} |\partial_{x\gamma} G_d(y,x\gamma_d)| \, \|x_1 - x_2\| \, \|\gamma_d\| \\
\leqslant& M_1 |\theta_1 - \theta_2| + M_2 \|x_1 - x_2\| \leqslant \sqrt{2}\max\{M_1, M_2\} \left\|(\theta_1, x_1')' - (\theta_2, x_2')'\right\|,
\end{aligned}
$$

where M_1 and M_2 are positive constants. The second inequality is due to the Lipschitz continuity condition on ϕ''_θ, and the third follows because ϕ'''_θ and $\partial_{x\gamma} G_d$ are uniformly bounded under Assumption 6.4, and 6.2. The last one is by Hölder's inequality. Another application of Lemma 2.13 of Pakes and Pollard (1989) yields the desired result.

Let $\mathcal{M}_x = \{\tilde{x} \mapsto \tilde{x}_\ell - x_\ell : \ell = 2,...,k, x \in \mathcal{X}\}$. Since $\mathcal{X}$ is compact, $\mathcal{M}_x$ is a VC class because $\mathcal{N}\left(\epsilon \sup_{x\in\mathcal{X}} \left\|x_{[-1]}\right\|, \mathcal{M}_x, L_2(Q)\right) \leqslant C\left(diam(\mathcal{X})/\epsilon\right)$, for a positive constant C independent of ϵ. Applying Corollary A.1 in Chernozhukov *et al.* (2014) again on the product of $\mathcal{M}_1$, $\mathcal{M}_3$, $\mathcal{M}_\phi$, $\mathcal{M}_x$, and the finite set $\{(y_1,y_2,d_2) \mapsto \mathbb{1}\{d_2 = d, y_2 \leqslant y_1\}, d \in \{0,1\}\}$ yields that the first half of g_{12} belongs to a VC class. Next, by Lemma 5 of Sherman (1994), we deduce that $\{y_1 \mapsto \int \mathbb{1}\{d_2 = d, y_2 \leqslant y_1 \wedge t\} hK^{(1)}(x\gamma_d, x_2'\gamma_d)(x_{2,l} - x_l)dF(w_2) : \omega \in \Omega\}$ and $\{w_2 \mapsto \int g_{11}(w_1,\omega)g_{12}(w_2,y_1,\omega)dF(w_1) : \omega \in \Omega\}$ are both of the VC type. Since $f_d(x\gamma_d)$ is uniformly bounded away from 0, $\{1/f_d(x\gamma_d)^2 : (d,x) \in \{0,1\} \times \mathcal{X}\}$ admits a finite envelop. Applying Corollary A.1 in Chernozhukov *et al.* (2014) yet again concludes the proof.

Turning to $\mathcal{G}_\eta$, we first show that for a fixed x, the set $\mathcal{M}_\theta \equiv \{1/\phi'_\theta(s_{T_d}(t, x\gamma_d, \theta)) : (t,\theta) \times$

$\tilde{\mathcal{T}} \times \Theta\}$ belongs to the VC class. Recall that $s_{T_d}(t, x\gamma_d, \theta) = \phi_\theta^{-1}\left(\int_0^t \phi_\theta'(s_d(y, x\gamma_d))s_{d,1}(dy, x\gamma_d)\right)$. Hence,

$$\begin{aligned}
&1/\phi_{\theta_1}'(s_{T_d}(t_1, x\gamma_d, \theta_1)) - 1/\phi_{\theta_2}'(s_{T_d}(t_2, x\gamma_d, \theta_2)) \\
\leqslant &\left\{1/\phi_{\theta_1}'(s_{T_d}(t_1, x\gamma_d, \theta_1)) - 1/\phi_{\theta_2}'(s_{T_d}(t_1, x\gamma_d, \theta_2))\right\} \\
&+ \left\{1/\phi_{\theta_2}'(s_{T_d}(t_1, x\gamma_d, \theta_2)) - 1/\phi_{\theta_2}'(s_{T_d}(t_2, x\gamma_d, \theta_2))\right\} \\
\equiv &\Delta_1 + \Delta_2.
\end{aligned}$$

Decomposing the first term further into,

$$\begin{aligned}
|\Delta_1| \leqslant &\left|1/\dot{\phi}_{\theta_1}^{-1}\left(\int_0^{t_1} \phi_{\theta_1}'(s_d(y, x\gamma_d))s_{d,1}(dy, x\gamma_d)\right) - 1/\dot{\phi}_{\theta_1}^{-1}\left(\int_0^{t_1} \phi_{\theta_2}'(s_d(y, x\gamma_d))s_{d,1}(dy, x\gamma_d)\right)\right| \\
&+ \left|1/\dot{\phi}_{\theta_1}^{-1}\left(\int_0^{t_1} \phi_{\theta_2}'(s_d(y, x\gamma_d))s_{d,1}(dy, x\gamma_d)\right) - 1/\dot{\phi}_{\theta_2}^{-1}\left(\int_0^{t_1} \phi_{\theta_2}'(s_d(y, x\gamma_d))s_{d,1}(dy, x\gamma_d)\right)\right| \\
\equiv &\Delta_{11} + \Delta_{12}.
\end{aligned}$$

For the first term, we have

$$|\Delta_{11}| \leqslant (1 - v_o) \sup_{(z,\theta)\in[0,y_o^*]\times\Theta} \left|\frac{\phi_\theta''(\phi_\theta^{-1}(z))}{\left(\dot{\phi}_\theta^{-1}(z)\right)^3}\right| \sup_{(u,\theta)\in[v_o,1]\times\Theta} \left|\phi_\theta'(u)\right| |\theta_1 - \theta_2| = M_3 |\theta_1 - \theta_2|.$$

Under Assumption 10.(ii), $\Delta_{12} \leqslant M_4 |\theta_1 - \theta_2|$.

$$\begin{aligned}
|\Delta_2| \leqslant &\sup_{(z,\theta)\in[0,y_o^*]\times\Theta} \left|\frac{\ddot{\phi}_\theta^{-1}(z)}{\left(\dot{\phi}_\theta^{-1}(z)\right)^3}\right| \\
&\cdot \left|\int_0^{t_1} \phi_{\theta_2}'(s_d(y, x\gamma_d))s_{d,1}(dy, x\gamma_d) - \int_0^{t_2} \phi_{\theta_2}'(s_d(y, x\gamma_d))s_{d,1}(dy, x\gamma_d)\right| \\
\leqslant &\sup_{(z,\theta)\in[0,y_o^*]\times\Theta} \left|\frac{\ddot{\phi}_\theta^{-1}(z)}{\left(\dot{\phi}_\theta^{-1}(z)\right)^3}\right| \cdot \sup_{(u,\theta)\in[v_o,1]\times\Theta} \left|\phi_\theta'(u)\right| \sup_{(y,x)\in\tilde{\mathcal{T}}\times\mathcal{X}} |\partial_{x\gamma} G_{d,1}(y, x\gamma_d)| \, |t_1 - t_2| \\
= &M_5 |t_1 - t_2|.
\end{aligned}$$

where inequalities hold by the mean value theorem and under Assumptions 6.2, and 6.4. Combining the bounds and applying Hölder's inequality, we conclude by Lemma 2.13 of Pakes and Pollard (1989) that $\mathcal{M}_\theta$ is a VC class.

Next, following similar analysis as in the previous part, we deduce from Corollary A.1 of Chernozhukov *et al.* (2014) that $\{w_1 \mapsto \mathbb{1}\{d_1 = d\}(\mathbb{1}\{y_1 \leqslant y\} - G_d(y, x_1\gamma_d))\,\partial_y G_{d,1}(y, x\gamma_d) :$

$(d, y, \theta) \in \{0,1\} \times \tilde{\mathcal{T}} \times \Theta\}$ for a given $x \in \mathcal{X}$ is a VC class with a finite envelop. Applying Lemma 5 of Sherman (1994), we get $\{w_1 \mapsto \int \mathbb{1}\{y_1 \leqslant t\} \mathbb{1}\{d_1 = d\} (\mathbb{1}\{y_1 \leqslant y\} - G_d(y, x_1\gamma_d)) \partial_y G_{d,1}(y, x\gamma_d) dy : (d, t, \theta) \in \{0,1\} \times \tilde{\mathcal{T}} \times \Theta\}$ also belongs to the VC class with an envelop $F_{\eta,1} = G_{d,1}(y_o \wedge y_c, x\gamma_d)$. This is due to

$$
\begin{aligned}
&\int \mathbb{1}\{y_1 \leqslant t\} \mathbb{1}\{d_1 = d\} (\mathbb{1}\{y_1 \leqslant y\} - G_d(y, x_1\gamma_d)) \partial_y G_{d,1}(y, x\gamma_d) dy \\
&\quad \leqslant 2 \int_0^t \partial_y G_{d,1}(y, x\gamma_d) dy \leqslant G_{d,1}(y_o \wedge y_c, x\gamma_d).
\end{aligned}
$$

Analogous results can be established for the other two parts of Ψ_d.

Combining these results with the fact that $\{x_1 \mapsto K((x_1\gamma_d - x\gamma_d)/h) : (t, h) \in \tilde{\mathcal{T}} \times \mathcal{H}\}$ is VC with an envelop $C\mathbb{1}\{|x_1\gamma_d - x\gamma_d| \leqslant h\}$, we deduce that $\mathcal{G}_\eta$ is of the VC type, with the envelop given by $\sum_{d=0,1} C_d \mathbb{1}\{|x_1\gamma_d - x\gamma_d| \leqslant h\}$ where C_0 and C_1 are positive constants. Setting $H_{\eta,d}(x_1\gamma_d) = C_d \mathbb{1}\{|x_1\gamma_d - x\gamma_d| \leqslant h\}$ concludes the proof. ■

Lemma C.4 *Under the assumptions of Theorem 4.1, for any $\delta_n = O_p\left(n^{-1/2}\right)$,*

$$
\begin{aligned}
\sup_{\|\tilde{\gamma}_d - \gamma_d\| \leqslant \delta_n} \sup_{(t,x) \in \tilde{\mathcal{T}} \times \mathcal{X}} &\left\| \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y, x) \left\{ \partial_\gamma \hat{G}_d(y, x\tilde{\gamma}_d) - \partial_\gamma \hat{G}_d(y, x\gamma_d) \right\} s_{d,1}(dy, x\gamma_d) \right\| \\
&= O_p\left((\log n)^{1/2} n^{-1/2} h^{-5/2} \delta_n \right) + O(\delta_n).
\end{aligned}
$$

Proof of Lemma C.4. Split the term inside the norm operator into

$$
\begin{aligned}
\Delta_1(t, x, \theta) &\equiv \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y, x) \left\{ \frac{\partial_\gamma \hat{\kappa}_{d,y}(x\tilde{\gamma}_d)}{\hat{f}_d(x\tilde{\gamma}_d)} - \frac{\partial_\gamma \hat{\kappa}_{d,y}(x\gamma_d)}{\hat{f}_d(x\gamma_d)} \right\} s_{d,1}(dy, x\gamma_d), \\
\Delta_2(t, x, \theta) &\equiv - \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y, x) \left\{ \frac{\hat{\kappa}_{d,y}(x\tilde{\gamma}_d) \partial_\gamma \hat{f}_d(x\tilde{\gamma}_d)}{\hat{f}^2_d(x\tilde{\gamma}_d)} - \frac{\hat{\kappa}_{d,y}(x\gamma_d) \partial_\gamma \hat{f}_d(x\gamma_d)}{\hat{f}^2_d(x\gamma_d)} \right\} s_{d,1}(dy, x\gamma_d).
\end{aligned}
$$

Decomposing Δ_1 and ignoring smaller order terms gives

$$
\begin{aligned}
\Delta_1(t, x, \theta) = &f_d(x\gamma_d)^{-1} \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y, x) \{ \partial_\gamma \hat{\kappa}_{d,y}(x\tilde{\gamma}_d) - \partial_\gamma \hat{\kappa}_{d,y}(x\gamma_d) \} s_{d,1}(dy, x\gamma_d) \\
&+ \frac{\hat{f}_d(x\tilde{\gamma}_d) - \hat{f}_d(x\gamma_d)}{f_d(x\tilde{\gamma}_d) f_d(x\gamma_d)} \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y, x) \partial_\gamma \hat{\kappa}_{d,y}(x\gamma_d) s_{d,1}(dy, x\gamma_d) + (s.o.)
\end{aligned}
$$

We investigate the uniform rate of the first term only. The second term exhibits the same rate and is simpler. Define

$$
\Delta_{11}(t, x, \theta, \tilde{\gamma}_d) \equiv \int_0^t \ddot{\phi}^\theta_{d,\gamma_d}(y, x) \mathbb{E}[\partial_\gamma \hat{\kappa}_{d,y}(x\tilde{\gamma}_d) - \partial_\gamma \hat{\kappa}_{d,y}(x\gamma_d)] s_{d,1}(dy, x\gamma_d),
$$

$$\Delta_{12}(t,x,\theta,\tilde{\gamma}_d) \equiv \int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\{\partial_\gamma \hat{\kappa}_{d,y}(x\tilde{\gamma}_d) - \partial_\gamma \hat{\kappa}_{d,y}(x\gamma_d) \\ - \mathbb{E}[\partial_\gamma \hat{\kappa}_{d,y}(x\tilde{\gamma}_d) - \partial_\gamma \hat{\kappa}_{d,y}(x\gamma_d)]\} s_{d,1}(dy, x\gamma_d).$$

By Fubini's theorem and standard change of variables,

$$\begin{aligned}
\Delta_{11}(t,x,\theta,\tilde{\gamma}_d) = {} & h^{-2}\int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\cdot \\
& \mathbb{E}\left[\varrho^{\tilde{\gamma}_d}_{1,1}(y,X\tilde{\gamma}_d)K^{(1)}((X\tilde{\gamma}_d - x\tilde{\gamma}_d)/h) - \varrho^{\gamma_d}_{1,1}(y,X\gamma_d)K^{(1)}((X\gamma_d - x\gamma_d)/h)\right] s_{d,1}(dy,x\gamma_d) \\
= {} & h^{-1}\int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\cdot\left\{\int_{\mathbb{R}} K^{(1)}(u)\varrho^{\tilde{\gamma}_d}_{1,1}(y,x\tilde{\gamma}_d+uh)f_d(x\tilde{\gamma}_d+uh)du\right. \\
& \left. - \int_{\mathbb{R}} K^{(1)}(u)\varrho^{\gamma_d}_{1,1}(y,x\gamma_d+uh)f_d(x\gamma_d+uh)du\right\} s_{d,1}(dy,x\gamma_d) \\
= {} & \int_{\mathbb{R}} uK^{(1)}(u)du\cdot\int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\cdot\left\{\left(\partial z\varrho^{\tilde{\gamma}_d}_{1,1}(y,z)|_{z=x\tilde{\gamma}_d}f_d(x\tilde{\gamma}_d) + \varrho^{\tilde{\gamma}_d}_{1,1}(y,x\tilde{\gamma}_d)\partial_z f_d(z)|_{z=x\tilde{\gamma}_d}\right)\right. \\
& \left. - \left(\partial z\varrho^{\gamma_d}_{1,1}(y,z)|_{z=x\gamma_d}f_d(x\gamma_d) + \varrho^{\gamma_d}_{1,1}(y,x\gamma_d)\partial_z f_d(z)|_{z=x\gamma_d}\right)\right\} s_{d,1}(dy,x\gamma_d),
\end{aligned}$$

where $\varrho^{\gamma}_{1,1}$ is defined in (4.8). The second equality follows by Taylor expansion and the fact that $\int_{[-1,1]} K^{(1)}(u)du = 0$. By the Lipschitz continuity of $\varrho^{\gamma}_{1,1}(y,x\gamma)$, $\partial_{x\gamma}\varrho^{\gamma}_{1,1}(y,x\gamma)$, $f_d(x\gamma)$, and $\partial_{x\gamma}f_d(x\gamma)$, with respect to γ as implied by Assumption 7.1, and by the fact that $\|\tilde{\gamma}_d - \gamma_d\| \leqslant \delta_n$, we conclude that $\sup_{\|\tilde{\gamma}_d-\gamma_d\|\leqslant\delta_n}\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta}\|\Delta_{11}(t,x,\theta,\tilde{\gamma}_d)\| = O(\delta_n)$.

The centered term Δ_{12} can be bounded using following empirical process

$$\mathcal{G}_{\delta,n} \equiv \{w \mapsto g_\delta(w,\omega,\tilde{\gamma}_d) : \omega\in\Omega, \|\tilde{\gamma}_d - \gamma_d\| \leqslant \delta_n\},$$

where $g_\delta(W,\omega,\tilde{\gamma}_d) \equiv g_{\delta,1}(W,\omega,\tilde{\gamma}_d) - \int g_{\delta,1}(W,\omega,\tilde{\gamma}_d)dF_W(W)$, and

$$g_{\delta,1}(W,\omega,\tilde{\gamma}_d) \equiv \int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\mathbb{1}\{D=d, Y\leqslant y\} s_{d,1}(dy,x\gamma_d) \\ \cdot\left\{K^{(1)}((X\tilde{\gamma}_d - x\tilde{\gamma}_d)/h) - K^{(1)}((X\gamma_d - x\gamma_d)/h)\right\}.$$

Applying similar lines of arguments as in Lemma C.3, it is straightforward to show that $\mathcal{G}_{\delta,n}$ is a VC type class with bounded envelop, for each δ_n. From the continuous differentiability of $K^{(1)}(\cdot)$, we deduce by similar arguments to Lemma 8.4 in Maistre and Patilea (2019) that $\left|K^{(1)}((X\tilde{\gamma}_d - x\tilde{\gamma}_d)/h) - K^{(1)}((X\gamma_d - x\gamma_d)/h)\right| \leqslant \delta_n h^{-1}\left\|K^{(2)}((X\gamma_d - x\gamma_d)/h)\right\| + C\delta_n^2 h^{-2}$, for some positive constant C. Combine this fact with the uniform boundedness of $\ddot{\phi}^{\theta}_{d,\gamma_d}$, and $s_{d,1}$, and we find that $\sup_{g_\delta\in\mathcal{G}_\delta}\mathbb{E}[g_\delta^2]$ is bounded from above at the rate of $O(\delta_n^2 h^{-1})$. We then conclude from applying the maximal inequality in Lemma C.1 that $\sup_{\|\tilde{\gamma}_d-\gamma_d\|}\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta}\|\Delta_{12}(t,x,\theta,\tilde{\gamma}_d)\| = O_p\left((\log n)^{1/2} n^{-1/2}h^{-5/2}\delta_n\right)$.

For Δ_2, we have

$$\begin{aligned}\Delta_2(t,x,\theta,\tilde{\gamma}_d) = &- f_d^{-2}(x\tilde{\gamma}_d)\int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\left\{\hat{\kappa}_{d,y}(x\tilde{\gamma}_d)-\hat{\kappa}_{d,y}(x\gamma_d)\right\}\partial_\gamma \hat{f}_d(x\tilde{\gamma}_d)s_{d,1}(dy,x\gamma_d)\\ &- f_d^{-2}(x\gamma_d)\int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\kappa_{d,y}(x\gamma_d)\left\{\partial_\gamma \hat{f}_d(x\tilde{\gamma}_d)-\partial_\gamma \hat{f}_d(x\gamma_d)\right\}s_{d,1}(dy,x\gamma_d)\\ &+ \frac{(f_d(x\tilde{\gamma}_d)+f_d(x\gamma_d))(\hat{f}_d(x\tilde{\gamma}_d)-\hat{f}_d(x\gamma_d))}{f_d^{-2}(x\tilde{\gamma}_d)f_d^{-2}(x\gamma_d)}\\ &\cdot\int_0^t \ddot{\phi}^{\theta}_{d,\gamma_d}(y,x)\kappa_{d,y}(x\gamma_d)\partial_\gamma \hat{f}_d(x\gamma_d)s_{d,1}(dy,x\gamma_d).\end{aligned}$$

Arguing as in the case of Δ_1, one finds that the second term in the above display dominates the other two with a uniform rate of $O_p\left((\log n)^{1/2} n^{-1/2}h^{-5/2}\delta_n\right)+O(\delta_n)$. Gathering results on Δ_1 and Δ_2 completes our proof. ■

Lemma C.5 *Suppose the conditions of Theorem 5.1 hold. Then*

$$\begin{aligned}\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta}&\left|n^{-1/2}h^{1/2}\sum_{i=1}^n g_{d,\gamma_d,\ell}(X_i,x)\left\{\hat{\Psi}_d\left(\hat{\mathcal{E}}_{d,\hat{\gamma}_d,i},\hat{\mathcal{E}}_{d,1,\hat{\gamma}_d,i}\right)(t,x,\theta)-\Psi_d\left(\mathcal{E}_{d,\gamma_d,i},\mathcal{E}_{d,1,\gamma_d,i}\right)(t,x,\theta)\right\}\right|\\ &= O_p\left((\log n)^{1/2} n^{-1/2}h^{-(2\ell+1)/2}\right),\end{aligned}$$

where $g_{d,\gamma_d,\ell}(X,x)=h^{-(\ell+1)}K^{(\ell)}(x\gamma_d,X\gamma_d)/f(x\gamma_d,d)$*, for* $\ell=0,1$.

Proof of Lemma C.5. Define $\eta_{3,1}(t,x\gamma)=\phi'_\theta(s_{T_d}(t,x\gamma,\theta))$, $\eta_{3,2}(t,x\gamma)=\phi''_\theta(s_d(t,x\gamma))$, $\eta_{3,3}(W,t,\gamma)=\mathbb{1}\{D=d\}(\mathbb{1}\{Y\leqslant t\}-G_d(t,X\gamma))$, $\eta_{3,4}(t,x\gamma)=s_{d,1}(t,x\gamma)$, $\eta_{3,5}(W,t,\gamma)=\mathbb{1}\{D=d\}(R\mathbb{1}\{Y\leqslant t\}-G_{d,1}(t,X\gamma))$, and for $\ell=1,...,5$, let the estimator of $\eta_{3,\ell}$ be denoted by $\hat{\eta}_{3,\ell}$. Their definitions should be apparent. Index on θ is suppressed.

From Theorem 4.1 and Lemma B.2, we have

$$\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta}\left|\hat{\eta}_{3,\ell}(t,x\gamma_d)-\eta_{3,\ell}(t,x\gamma_d)\right| = O_p\left((\log n)^{1/2} n^{-1/2}h^{-1/2}\right),$$

for $\ell=1,2,4$.

Given these notations, we divide Ψ_d into

$$\begin{aligned}\Psi_{d,1}(W,t,x\gamma) &= \frac{1}{\eta_{3,1}(t,x\gamma)}\int_0^t \eta_{3,2}(y,x\gamma)\eta_{3,3}(W,y,x\gamma)\eta_{3,4}(dy,x\gamma),\\ \Psi_{d,2}(W,t,x\gamma) &= \frac{-1}{\eta_{3,1}(t,x\gamma)}\eta_{3,4}(t,x\gamma)\eta_{3,5}(W,t,x\gamma),\\ \Psi_{d,3}(W,t,x\gamma) &= \frac{1}{\eta_{3,1}(t,x\gamma)}\int_0^t \eta_{3,2}(y,x\gamma)\eta_{3,5}(W,y,x\gamma)\eta_{3,4}(dy,x\gamma),\end{aligned}$$

and thus $\Psi_d(\mathcal{E}_{d,\gamma}, \mathcal{E}_{d,1,\gamma}) = \sum_{\ell=1}^{3} \Psi_{d,\ell}$. We illustrate on $\Psi_{d,1}$, since the other two terms share a similar structure. From tedious manipulation, it can be shown that $\hat{\Psi}_{d,1}(W, t, x\hat{\gamma}_d) - \Psi_{d,1}(W, t, x\gamma_d) = \sum_{\ell=1}^{10} A_{3,\ell}(W, t, x)$, where

$$
\begin{aligned}
A_{3,1}(W,t,x) &= -\frac{\hat{\eta}_{3,1}(t, x\hat{\gamma}_d) - \eta_{3,1}(t, x\gamma_d)}{\eta_{3,1}(t, x\gamma_d)\hat{\eta}_{3,1}(t, x\hat{\gamma}_d)} \int_0^t \eta_{3,2}(y, x\gamma_d)\eta_{3,3}(W, y, \gamma_d)\eta_{3,4}(dy, x\gamma_d),\\
A_{3,2}(W,t,x) &= \frac{1}{\eta_{3,1}(t, x\gamma_d)} \int_0^t (\hat{\eta}_{3,2}(y, x\hat{\gamma}_d) - \eta_{3,2}(y, x\gamma_d))\eta_{3,3}(W, y, \gamma_d)\eta_{3,4}(dy, x\gamma_d),\\
A_{3,3}(W,t,x) &= \frac{1}{\eta_{3,1}(t, x\gamma_d)} \int_0^t \eta_{3,2}(y, x\gamma_d)(\hat{\eta}_{3,3}(W, y, \hat{\gamma}_d) - \eta_{3,3}(W, y, \gamma_d))\eta_{3,4}(dy, x\gamma_d),\\
A_{3,4}(W,t,x) &= \frac{1}{\eta_{3,1}(t, x\gamma_d)} \int_0^t \eta_{3,2}(y, x\gamma_d)\eta_{3,3}(W, y, \gamma_d)(\hat{\eta}_{3,4}(dy, x\hat{\gamma}_d) - \eta_{3,4}(dy, x\gamma_d)),\\
A_{3,5}(W,t,x) &= -\frac{\hat{\eta}_{3,1}(t, x\hat{\gamma}_d) - \eta_{3,1}(t, x\gamma_d)}{\eta_{3,1}(t, x\gamma_d)\hat{\eta}_{3,1}(t, x\hat{\gamma}_d)}\\
&\quad \cdot \int_0^t (\hat{\eta}_{3,2}(y, x\hat{\gamma}_d) - \eta_{3,2}(y, x\gamma_d))\eta_{3,3}(W, y, \gamma_d)\eta_{3,4}(dy, x\gamma_d),\\
A_{3,6}(W,t,x) &= -\frac{\hat{\eta}_{3,1}(t, x\hat{\gamma}_d) - \eta_{3,1}(t, x\gamma_d)}{\eta_{3,1}(t, x\gamma_d)\hat{\eta}_{3,1}(t, x\hat{\gamma}_d)}\\
&\quad \cdot \int_0^t \eta_{3,2}(y, x\gamma_d)(\hat{\eta}_{3,3}(W, y, \hat{\gamma}_d) - \eta_{3,3}(W, y, \gamma_d))\eta_{3,4}(dy, x\gamma_d),\\
A_{3,7}(W,t,x) &= -\frac{\hat{\eta}_{3,1}(t, x\hat{\gamma}_d) - \eta_{3,1}(t, x\gamma_d)}{\eta_{3,1}(t, x\gamma_d)\hat{\eta}_{3,1}(t, x\hat{\gamma}_d)}\\
&\quad \cdot \int_0^t \eta_{3,2}(y, x\gamma_d)\eta_{3,3}(W, y, \gamma_d)(\hat{\eta}_{3,4}(dy, x\hat{\gamma}_d) - \eta_{3,4}(dy, x\gamma_d)),\\
A_{3,8}(W,t,x) &= \frac{1}{\eta_{3,1}(t, x\gamma_d)} \int_0^t (\hat{\eta}_{3,2}(y, x\hat{\gamma}_d) - \eta_{3,2}(y, x\gamma_d))\\
&\quad \cdot (\hat{\eta}_{3,3}(W, y, \hat{\gamma}_d) - \eta_{3,3}(W, y, \gamma_d))\eta_{3,4}(dy, x\gamma_d),\\
A_{3,9}(W,t,x) &= \frac{1}{\eta_{3,1}(t, x\gamma_d)} \int_0^t \eta_{3,2}(y, x\gamma_d)(\hat{\eta}_{3,3}(W, y, \hat{\gamma}_d) - \eta_{3,3}(W, y, \gamma_d))\\
&\quad \cdot (\hat{\eta}_{3,4}(dy, x\hat{\gamma}_d) - \eta_{3,4}(dy, x\gamma_d)),\\
A_{3,10}(W,t,x) &= \frac{1}{\eta_{3,1}(t, x\gamma_d)} \int_0^t (\hat{\eta}_{3,2}(y, x\hat{\gamma}_d) - \eta_{3,2}(y, x\gamma_d))\\
&\quad \cdot \eta_{3,3}(y, x\gamma_d)(\hat{\eta}_{3,4}(dy, x\hat{\gamma}_d) - \eta_{3,4}(dy, x\gamma_d)),\\
A_{3,11}(W,t,x) &= -\frac{\hat{\eta}_{3,1}(t, x\hat{\gamma}_d) - \eta_{3,1}(t, x\gamma_d)}{\eta_{3,1}(t, x\gamma_d)\hat{\eta}_{3,1}(t, x\hat{\gamma}_d)}\\
&\quad \cdot \int_0^t (\hat{\eta}_{3,2}(y, x\hat{\gamma}_d) - \eta_{3,2}(y, x\gamma_d))(\hat{\eta}_{3,3}(W, y, \hat{\gamma}_d) - \eta_{3,3}(W, y, \gamma_d))\eta_{3,4}(dy, x\gamma_d),
\end{aligned}
$$

$$A_{3,12}(W,t,x) = \frac{1}{\eta_{3,1}(t,x\gamma_d)} \int_0^t (\hat{\eta}_{3,2}(y,x\hat{\gamma}_d) - \eta_{3,2}(y,x\gamma_d))$$
$$\cdot (\hat{\eta}_{3,3}(W,y,\hat{\gamma}_d) - \eta_{3,3}(W,y,\gamma_d))(\hat{\eta}_{3,4}(dy,x\hat{\gamma}_d) - \eta_{3,4}(dy,x\gamma_d)),$$
$$A_{3,13}(W,t,x) = -\frac{\hat{\eta}_{3,1}(t,x\hat{\gamma}_d) - \eta_{3,1}(t,x\gamma_d)}{\eta_{3,1}(t,x\gamma_d)\hat{\eta}_{3,1}(t,x\hat{\gamma}_d)}$$
$$\cdot \int_0^t (\hat{\eta}_{3,2}(y,x\hat{\gamma}_d) - \eta_{3,2}(y,x\gamma_d))\eta_{3,3}(W,y,\gamma_d)(\hat{\eta}_{3,4}(dy,x\hat{\gamma}_d) - \eta_{3,4}(dy,x\gamma_d)),$$
$$A_{3,14}(W,t,x) = -\frac{\hat{\eta}_{3,1}(t,x\hat{\gamma}_d) - \eta_{3,1}(t,x\gamma_d)}{\eta_{3,1}(t,x\gamma_d)\hat{\eta}_{3,1}(t,x\hat{\gamma}_d)}$$
$$\cdot \int_0^t \eta_{3,2}(y,x\gamma_d)(\hat{\eta}_{3,3}(W,y,\hat{\gamma}_d) - \eta_{3,3}(W,y,\gamma_d))(\hat{\eta}_{3,4}(dy,x\hat{\gamma}_d) - \eta_{3,4}(dy,x\gamma_d)),$$
$$A_{3,15}(W,t,x) = -\frac{\hat{\eta}_{3,1}(t,x\hat{\gamma}_d) - \eta_{3,1}(t,x\gamma_d)}{\eta_{3,1}(t,x\gamma_d)\hat{\eta}_{3,1}(t,x\hat{\gamma}_d)} \cdot \int_0^t (\hat{\eta}_{3,2}(y,x\hat{\gamma}_d) - \eta_{3,2}(y,x\gamma_d))$$
$$\cdot (\hat{\eta}_{3,3}(W,y,\hat{\gamma}_d) - \eta_{3,3}(W,y,\gamma_d))(\hat{\eta}_{3,4}(dy,x\hat{\gamma}_d) - \eta_{3,4}(dy,x\gamma_d)).$$

Following the same type of analysis we have used so far, namely performing Taylor expansion, integration by parts, and applying the maximal inequality from Lemma C.1 whenever appropriate, we get

$$\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta} \left| n^{-1/2}h^{1/2} \sum_{i=1}^n g_{d,\gamma_d,\ell}(X_i,x) A_{3,\ell_1}(W_i,t,x) \right| = O_p\left((\log n)^{1/2} n^{-1/2} h^{-(2\ell+1)/2} \right)$$
$$\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta} \left| n^{-1/2}h^{1/2} \sum_{i=1}^n g_{d,\gamma_d,\ell}(X_i,x) A_{3,\ell_2}(W_i,t,x) \right| = O_p\left(\log n \cdot n^{-1} h^{-(\ell+1)} \right)$$
$$\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta} \left| n^{-1/2}h^{1/2} \sum_{i=1}^n g_{d,\gamma_d,\ell}(X_i,x) A_{3,\ell_3}(W_i,t,x) \right| = O_p\left((\log n)^{3/2} n^{-3/2} h^{-(2\ell+3)/2} \right)$$
$$\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta} \left| n^{-1/2}h^{1/2} \sum_{i=1}^n g_{d,\gamma_d,\ell}(X_i,x) A_{3,15}(W_i,t,x) \right| = O_p\left((\log n)^2 n^{-2} h^{-(\ell+2)} \right).$$

for $\ell = 0, 1$, $\ell_1 = 1, 2, 3, 4$, $\ell_2 = 5, 6, 7, 8, 9, 10$, and $\ell_3 = 11, 12, 13, 14$. As a result, $\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta} \left| n^{-1/2}h^{1/2} \sum_{i=1}^n g_{d,\gamma_d,\ell}(X_i,x) \left(\hat{\Psi}_{d,1}(W_i,t,x\hat{\gamma}_d) - \Psi_{d,1}(W_i,t,x\gamma_d) \right) \right| = O_p\left((\log n)^{1/2} n^{-1/2} h^{-(2\ell+1)/2} \right)$. Analogous results hold for $\Psi_{d,2}$ and $\Psi_{d,3}$, concluding the proof. ■

C.2.1 Covariance Functions

Lemma C.6 *Suppose the assumptions of Corollary 1 hold. Then, it holds that,* $\Sigma_\eta^x(\cdot,\cdot) = \Sigma_\eta^{x\dagger}(\cdot,\cdot) + o(1)$, *where*

$$\Sigma_\eta^{x\dagger}(\mathbf{t},\boldsymbol{\theta}) = \begin{pmatrix} \sigma^2_{1,x}(t_1,t_2,\theta_1,\theta_2) & 0 \\ 0 & \sigma^2_{0,x}(t_1,t_2,\theta_3,\theta_4) \end{pmatrix},$$

and

$$\begin{aligned}
\sigma^2_{d,x}(t_1,t_2,\theta_1,\theta_2) &= \frac{\|K\|_2^2}{f_d(x\gamma_d)\phi'_{\theta_1}(s_{T_d}(t_1,x\gamma_d,\theta_1))\phi'_{\theta_2}(s_{T_d}(t_2,x\gamma_d,\theta_2))} \\
&\cdot \Bigg\{ \int_0^{t_1}\int_0^{t_2} \ddot{\phi}^{\theta_1}_{d,\gamma_d}(y_1,x)\ddot{\phi}^{\theta_2}_{d,\gamma_d}(y_2,x) \\
&\qquad \cdot \{G_d(y_1\wedge y_2,x\gamma_d) - G_d(y_1,x\gamma_d)G_d(y_2,x\gamma_d)\}\, s_{d,1}(dy_2,x\gamma_d)s_{d,1}(dy_1,x\gamma_d) \\
&+ \int_0^{t_1} \ddot{\phi}^{\theta_1}_{d,\gamma_d}(y_1,x)\dot{\phi}^{\theta_2}_{d,\gamma_d}(t_2,x)\{G_{d,1}(y_1\wedge t_2,x\gamma_d) - G_d(y_1,x\gamma_d)G_{d,1}(t_2,x\gamma_d)\}\, s_{d,1}(dy_1,x\gamma_d) \\
&- \int_0^{t_1}\int_0^{t_2} \ddot{\phi}^{\theta_1}_{d,\gamma_d}(y_1,x)\ddot{\phi}^{\theta_2}_{d,\gamma_d}(y_2,x) \\
&\qquad \cdot \{G_{d,1}(y_1\wedge y_2,x\gamma_d) - G_d(y_1,x\gamma_d)G_{d,1}(y_2,x\gamma_d)\}\, s_d(dy_2,x\gamma_d)s_{d,1}(dy_1,x\gamma_d) \\
&+ \int_0^{t_2} \dot{\phi}^{\theta_1}_{d,\gamma_d}(t_1,x)\ddot{\phi}^{\theta_2}_{d,\gamma_d}(y_2,x)\{G_{d,1}(y_2\wedge t_1,x\gamma_d) - G_d(y_2,x\gamma_d)G_{d,1}(t_1,x\gamma_d)\}\, s_{d,1}(dy_2,x\gamma_d) \\
&+ \dot{\phi}^{\theta_1}_{d,\gamma_d}(t_1,x)\dot{\phi}^{\theta_2}_{d,\gamma_d}(t_2,x)\{G_{d,1}(t_1\wedge t_2,x\gamma_d) - G_{d,1}(t_1,x\gamma_d)G_{d,1}(t_2,x\gamma_d)\} \\
&- \int_0^{t_2} \dot{\phi}^{\theta_1}_{d,\gamma_d}(t_1,x)\ddot{\phi}^{\theta_2}_{d,\gamma_d}(y_2,x)\{G_{d,1}(t_1\wedge y_2,x\gamma_d) - G_{d,1}(y_2,x\gamma_d)G_{d,1}(t_1,x\gamma_d)\}\, s_d(dy_2,x\gamma_d) \\
&- \int_0^{t_1}\int_0^{t_2} \ddot{\phi}^{\theta_1}_{d,\gamma_d}(y_1,x)\ddot{\phi}^{\theta_2}_{d,\gamma_d}(y_2,x) \\
&\qquad \cdot \{G_{d,1}(y_1\wedge y_2,x\gamma_d) - G_d(y_2,x\gamma_d)G_{d,1}(y_1,x\gamma_d)\}\, s_{d,1}(dy_2,x\gamma_d)s_d(dy_1,x\gamma_d) \\
&- \int_0^{t_1} \dot{\phi}^{\theta_2}_{d,\gamma_d}(t_2,x)\ddot{\phi}^{\theta_1}_{d,\gamma_d}(y_1,x)\{G_{d,1}(y_1\wedge t_2,x\gamma_d) - G_{d,1}(t_2,x\gamma_d)G_{d,1}(y_1,x\gamma_d)\}\, s_d(dy_1,x\gamma_d) \\
&+ \int_0^{t_1}\int_0^{t_2} \ddot{\phi}^{\theta_1}_{d,\gamma_d}(y_1,x)\ddot{\phi}^{\theta_2}_{d,\gamma_d}(y_2,x) \\
&\qquad \cdot \{G_{d,1}(y_1\wedge y_2,x\gamma_d) - G_{d,1}(y_1,x\gamma_d)G_{d,1}(y_2,x\gamma_d)\}\, s_d(dy_2,x\gamma_d)s_d(dy_1,x\gamma_d)\Bigg\}
\end{aligned} \tag{C.1}$$

Proof of Lemma C.6. When $d_1 \neq d_2$, $\Psi_{d_1}(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}) \cdot \Psi_{d_2}(\mathcal{E}_{d,\gamma_d},\mathcal{E}_{d,1,\gamma_d}) = 0$, implying the off-diagonal element of the covariance matrix is 0, regardless of t and θ. For terms on the main diagonal, the proof is analogous to that of Lemma C.1 in Fan and Liu (2018), and thus, we omit the details. ∎

C.3 First-Stage Estimator for the QTE

To estimate the bound curves for the QTE, we replace $f_{T_d,x}(t,\theta)$ and $f_{T_d}(t,\theta)$, with preliminary estimates $\hat{f}_{T_d,x}(t,\theta)$ and $\hat{f}_{T_d}(t,\theta)$, respectively. Validity of multiplier bootstrap procedure in Section 5.1 hinges on the estimates being uniformly consistent in (t,θ). In what follows, we provide estimators, based on the analytical expression of f_{T_d}, that satisfy this property. Using the closed form expression for s_{T_d} from Theorem 3.2, we deduce that

$$f_{T_d,x}(t,\theta) = \frac{\phi_\theta'(s_d(t,x\gamma_d))}{\phi_\theta'(s_{T_d}(t,x\gamma_d,\theta))} f_{d,1}(t,x\gamma_d),$$

where $f_{d,1}(t,x\gamma) = -\partial_t s_{d,1}(t,x\gamma)$. The fraction in the above display can be estimated by reusing $\hat{\gamma}_d$, $\hat{s}_d$, and $\hat{s}_{T_d}$ from (4.2), (4.4), and (4.5), respectively. For $f_{d,1}(t,x\gamma)$, we will use the SIM conditional density estimator as follows.

Let $H(\cdot)$ be a kernel function and λ be a sequence of bandwidths that fulfills the conditions in Assumption C.1. Define

$$\hat{f}_{d,1}(t,x\gamma) = \frac{\sum_{i=1}^n I_{d,y,1,i} H_\lambda(x\gamma, X_i\gamma)}{\sum_{i=1}^n H_\lambda(x\gamma, X_i\gamma)}, \tag{C.2}$$

where $H_\lambda(u,v) = \lambda^{-1}H(\lambda^{-1}(v-u))$. Now, we will have $f_{T_d,x}$ estimated by $\hat{f}_{T_d,x}(t,\theta) = \phi_\theta'(\hat{s}_d(t,x\hat{\gamma}_d))\hat{f}_{d,1}(t,x\hat{\gamma}_d)/\phi_\theta'(\hat{s}_{T_d}(t,x\hat{\gamma}_d,\theta))$. For the unconditional density f_{T_d}, we can estimate it by taking the sample average of $\hat{f}_{T_d,X}(t,\theta)$ with respect to X, i.e. $\hat{f}_{T_d}(t,\theta) \equiv n^{-1}\sum_{i=1}^n \hat{f}_{T_d,X_i}(t,\theta)$.

Assumption C.1 *(i) The kernel function, $H(\cdot)$ is symmetric, supported on $[-1,1]$, and of bounded variation; (ii) it is twice continuously differentiable and the second order derivative is continuous and of bounded variation; (ii) $\lambda \to 0$, $\log n \cdot n^{-1}\lambda^{-2} \to 0$, as $n \to \infty$; (iii) for $\ell_1 = 0,1$, and $\ell_2 = 0,1$; Define $\varrho_f^\gamma(y,x\gamma) \equiv f_d(x\gamma)\,\mathbb{E}\left[\partial_y G_{d,1}(y,X\gamma_d)|X\gamma = x\gamma\right]$; (iii) $v \mapsto \varrho_f^\gamma(y,v)$, is continuously differentiable and the derivative are bounded uniformly on $\tilde{\mathcal{T}} \times \mathcal{X} \times \Gamma_{d,n}$; (iv) $\partial_v \varrho_f^\gamma(y,v)$ is Lipschitz continuous in v with the Lipschitz constant being independent of y, x, and $\gamma \in \Gamma_{d,n}$.*

Lemma C.7 *Under the assumptions of Corollary 2 and Assumption C.1, it holds that $\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta}\left|\hat{f}_{T_d,x}(t,\theta) - f_{T_d,x}(t,\theta)\right| = o_p(1)$, and $\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta}\left|\hat{f}_{T_d}(t,\theta) - f_{T_d}(t,\theta)\right| = o_p(1)$, for $d \in \{0,1\}$.*

Proof of Lemma C.7. We first show that $\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta}\left|\hat{f}_{d,1}(t,x\hat{\gamma}_d) - f_{d,1}(t,x\gamma_d)\right| = o_p(1)$. By the triangular inequality, we have that

$$\left|\hat{f}_{d,1}(t,x\hat{\gamma}_d) - f_{d,1}(t,x\gamma_d)\right| \leqslant \left|\hat{f}_{d,1}(t,x\hat{\gamma}_d) - \hat{f}_{d,1}(t,x\gamma_d)\right| + \left|\hat{f}_{d,1}(t,x\gamma_d) - f_{d,1}(t,x\gamma_d)\right| \tag{C.3}$$

$$\equiv \Delta_{f,1}(t, x, \hat{\gamma}_d) + \Delta_{f,2}(t, x, \hat{\gamma}_d).$$

Along lines of arguments similar to that of Lemma C.4, we get that $\sup_{\|\gamma-\gamma_d\|\leqslant\delta_n} \sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta} |\Delta_{f,1}(t, x, \gamma, \theta)| = O_p\left((\log n)^{1/2} n^{-1/2}\lambda^{-3/2}\delta_n\right) + O(\delta_n)$. The second part arises from bias calculations, which depend crucially on Assumptions C.1.(iii) and C.1.(iv). Next, it follows, by direct application of Theorems 1 and 4 in Einmahl and Mason (2005), that $\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta} |\Delta_{f,2}(t, x)| = O_p\left((\log n)^{1/2} n^{-1/2}\lambda^{-1} + \lambda^2\right)$, which is $o_p(1)$ under Assumption C.1. Combining these results, we conclude that the left hand side of (C.3) is $o_p(1)$.

Now, observe that

$$\hat{f}_{T_d,x}(t,\theta) - f_{T_d,x}(t,\theta) = \left\{ \frac{\phi'_\theta(\hat{s}_d(t, x\hat{\gamma}_d))}{\phi'_\theta(\hat{s}_{T_d}(t, x\hat{\gamma}_d, \theta))} - \frac{\phi'_\theta(s_d(t, x\gamma_d))}{\phi'_\theta(s_{T_d}(t, x\gamma_d, \theta))} \right\} \hat{f}_{d,1}(t, x\hat{\gamma}_d)$$
$$+ \frac{\phi'_\theta(s_d(t, x\gamma_d))}{\phi'_\theta(s_{T_d}(t, x\gamma_d, \theta))} \left\{ \hat{f}_{d,1}(t, x\hat{\gamma}_d) - f_{d,1}(t, x\gamma_d) \right\}.$$

From the fact that $\hat{s}_d$ and $\hat{s}_{T_d}$ are uniformly convergent, that $\dot{\phi}_\theta^{-1}(z)$ is uniformly bounded away from 0 on $[0, y_o^*]$, and that, for each $(t, x) \in \tilde{\mathcal{T}} \times \mathcal{X}$, $\phi(\hat{s}_{T_d}(t, x\hat{\gamma}_d))$ belongs to $[0, y_o^*]$ with probability approaching 1, we deduce that the difference inside the curly braces in the first line is $o_p(1)$. Under Assumption 6.3, we have $\phi'_\theta(s_d(t, x\gamma_d))/\phi'_\theta(s_{T_d}(t, x\gamma_d, \theta))$ is uniformly $O(1)$.

The function $f_{d,1}(\cdot, \cdot\gamma_d)$ is uniformly bounded from Assumption 6.2. It then follows from the uniform convergence results we derived earlier, that $\hat{f}_{d,1}(t, x\hat{\gamma}_d)$ is also uniformly bounded across $(t, x) \in \tilde{\mathcal{T}} \times \mathcal{X}$. Consequently, $\sup_{(t,x,\theta)\in\tilde{\mathcal{T}}\times\mathcal{X}\times\Theta} \left|\hat{f}_{T_d,x}(t,\theta) - f_{T_d,x}(t,\theta)\right| = o_p(1)$. We notice that this also implies that $\sup_{(t,\theta)\in\tilde{\mathcal{T}}\times\Theta} \left|\hat{f}_{T_d}(t,\theta) - f_{T_d}(t,\theta)\right| = o_p(1)$, for $d \in \{0, 1\}$, which concludes our proof. ■

References

Abadie, A. (**2003**). "Semiparametric instrumental variable estimation of treatment response models," Journal of econometrics **113**(2), 231–263.

Abadie, A. (**2005**). "Semiparametric difference-in-differences estimators," The Review of Economic Studies **72**(1), 1–19.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (**1996**). "Identification of causal effects using instrumental variables," Journal of the American statistical Association **91**(434), 444–455.

Anstrom, K. J., and Tsiatis, A. A. (**2001**). "Utilizing propensity scores to estimate causal treatment effects with censored time-lagged data," Biometrics **57**(4), 1207–1218.

Athey, S., and Imbens, G. W. (**2006**). "Identification and inference in nonlinear difference-in-differences models," Econometrica **74**(2), 431–497.

Beran, R. (**1981**). "Nonparametric regression with randomly censored survival data," .

Bernasconi, D. P., Antolini, L. *et al.* (**2022**). "A causal inference approach to compare leukaemia treatment outcome in the absence of randomization and with dependent censoring," International Journal of Epidemiology **51**(1), 314–323.

Beyhum, J., Florens, J.-P., and Van Keilegom, I. (**2021**). "A nonparametric instrumental approach to endogeneity in competing risks models," arXiv preprint arXiv:2105.00946 .

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

Bouaziz, O., and Lopez, O. (**2010**). "Conditional density estimation in a censored single-index regression model," Bernoulli **16**(2), 514–542.

Braekers, R., and Veraverbeke, N. (**2005**). "A copula-graphic estimator for the conditional survival function under dependent censoring," Canadian Journal of Statistics **33**(3), 429–447.

Bücher, A., El Ghouch, A., and Van Keilegom, I. (**2021**). "Single-index quantile regression models for censored data," in *Advances in Contemporary Statistics and Econometrics* (Springer), pp. 177–196.

Card, D., and Krueger, A. B. (**1994**). "Minimum wages and employment: A case study of the fast-food industry in new jersey and pennsylvania," The American Economic Review **84**(4), 772.

Chen, Y.-H. (**2010**). "Semiparametric marginal regression analysis for dependent competing risks under an assumed copula," Journal of the Royal Statistical Society: Series B (Statistical Methodology) **72**(2), 235–251.

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

Chernozhukov, V., Fernández-Val, I., and Galichon, A. (**2010**). "Quantile and probability curves without crossing," Econometrica **78**(3), 1093–1125.

Chiang, C.-T., and Huang, M.-Y. (**2012**). "New estimation and inference procedures for a single-index conditional distribution model," Journal of Multivariate Analysis **111**, 271–285.

Crommen, G., Beyhum, J., and Van Keilegom, I. (**2022**). "A gaussian model for survival data subject to dependent censoring and confounding," arXiv preprint arXiv:2208.04184 .

Czado, C., and Van Keilegom, I. (**2021**). "Dependent censoring based on copulas," arXiv preprint arXiv:2104.06872 .

Dabrowska, D. M. (**1989**). "Uniform consistency of the kernel conditional kaplan-meier estimate," The Annals of Statistics 1157–1167.

Dawid, A. P. (**1979**). "Conditional independence in statistical theory," Journal of the Royal Statistical Society: Series B (Methodological) **41**(1), 1–15.

Delecroix, M., Hristache, M., and Patilea, V. (**2006**). "On semiparametric m-estimation in single-index regression," Journal of Statistical Planning and Inference **136**(3), 730–769.

Deresa, N. W., and Van Keilegom, I. (**2020**). "Flexible parametric model for survival data subject to dependent censoring," Biometrical Journal **62**(1), 136–156.

Deresa, N. W., Van Keilegom, I., and Antonio, K. (**2022**). "Copula-based inference for bivariate survival data with left truncation and dependent censoring," Insurance: Mathematics and Economics **107**, 1–21.

Domínguez, M. A., and Lobato, I. N. (**2004**). "Consistent estimation of models defined by conditional moment restrictions," Econometrica **72**(5), 1601–1615.

Einmahl, U., and Mason, D. M. (**2005**). "Uniform in bandwidth consistency of kernel-type function estimators," The Annals of Statistics **33**(3), 1380–1403.

Escanciano, J. C. (**2006**a). "A consistent diagnostic test for regression models using projections," Econometric Theory **22**(6), 1030–1051.

Escanciano, J. C. (**2006**b). "Goodness-of-fit tests for linear and nonlinear time series models," Journal of the American Statistical Association **101**(474), 531–541.

Fan, Y., and Liu, R. (**2018**). "Partial identification and inference in censored quantile regression," Journal of Econometrics **206**(1), 1–38.

Firpo, S. (**2007**). "Efficient semiparametric estimation of quantile treatment effects," Econometrica **75**(1), 259–276.

Frandsen, B. R. (**2015**). "Treatment effects with censoring and endogeneity," Journal of the American Statistical Association **110**(512), 1745–1752.

Frölich, M., and Melly, B. (**2013**). "Unconditional quantile treatment effects under endogeneity," Journal of Business & Economic Statistics **31**(3), 346–357.

Gatta, G., Capocaccia, R. *et al.* (**2005**). "Childhood cancer survival trends in europe: a eurocare working group study," Journal of Clinical Oncology **23**(16), 3742–3751.

Genest, C., and MacKay, J. (**1986**a). "The joy of copulas: Bivariate distributions with uniform marginals," The American Statistician **40**(4), 280–283.

Genest, C., and MacKay, R. J. (**1986**b). "Copules archimédiennes et families de lois bidimensionnelles dont les marges sont données," Canadian journal of statistics **14**(2), 145–159.

Giné, E., and Mason, D. M. (**2007**). "On local u-statistic processes and the estimation of densities of functions of several sample variables," The Annals of Statistics **35**(3), 1105–1145.

Giné, E., and Zinn, J. (**1984**). "Some limit theorems for empirical processes," The Annals of Probability 929–989.

González-Manteiga, W., and Crujeiras, R. M. (**2013**). "An updated review of goodness-of-fit tests for regression models," Test **22**(3), 361–411.

Heckman, J. J., Ichimura, H., and Todd, P. E. (**1997**). "Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme," The review of economic studies **64**(4), 605–654.

Heckman, J. J., and Vytlacil, E. (**2001**). "Policy-relevant treatment effects," American Economic Review **91**(2), 107–111.

Heckman, J. J., and Vytlacil, E. (**2005**). "Structural equations, treatment effects, and econometric policy evaluation 1," Econometrica **73**(3), 669–738.

Hirano, K., Imbens, G. W., and Ridder, G. (**2003**). "Efficient estimation of average treatment effects using the estimated propensity score," Econometrica **71**(4), 1161–1189.

Howard, S. C., and Wilimas, J. A. (**2005**). "Delays in diagnosis and treatment of childhood cancer: where in the world are they important?," Pediatric blood & cancer **44**(4), 303–304.

Howlader, N., Noone, A. *et al.* (**2016**). "Seer cancer statistics review, 1975–2013. bethesda, md: National cancer institute; 2016" .

Huang, X., and Zhang, N. (**2008**). "Regression survival analysis with an assumed copula for dependent censoring: a sensitivity analysis approach," Biometrics **64**(4), 1090–1099.

Hubbard, A. E., Laan, M. J., and Robins, J. M. (**2000**). "Nonparametric locally efficient estimation of the treatment specific survival distribution with right censored data and covariates in observational studies," in *Statistical Models in Epidemiology, the Environment, and Clinical Trials* (Springer), pp. 135–177.

Ichimura, H. (**1993**). "Semiparametric least squares (sls) and weighted sls estimation of single-index models," Journal of econometrics **58**(1-2), 71–120.

Imbens, G. W., and Angrist, J. D. (**1994**). "Identification and estimation of local average treatment effects," Econometrica: Journal of the Econometric Society 467–475.

Joe, H. (**1997**). *Multivariate models and multivariate dependence concepts* (CRC press).

Kaplan, E. L., and Meier, P. (**1958**). "Nonparametric estimation from incomplete observations," Journal of the American statistical association **53**(282), 457–481.

Khan, S., and Tamer, E. (**2009**). "Inference on endogenously censored regression models using conditional moment inequalities," Journal of Econometrics **152**(2), 104–119.

Kim, D. (**2021**). "Partially identifying competing risks models: An application to the war on cancer," Journal of Econometrics .

Kosorok, M. R. (**2003**). "Bootstraps of sums of independent but not identically distributed stochastic processes," Journal of Multivariate Analysis **84**(2), 299–318.

Kosorok, M. R. (**2008**). *Introduction to empirical processes and semiparametric inference.* (Springer, New York).

Lavergne, P., and Patilea, V. (**2013**). "Smooth minimum distance estimation and testing with conditional estimating equations: Uniform in bandwidth theory," Journal of Econometrics **177**(1), 47–59.

Lee, M.-j., and Lee, S.-j. (**2005**). "Analysis of job-training effects on korean women," Journal of Applied Econometrics **20**(4), 549–562.

Li, W., and Patilea, V. (**2018**). “A dimension reduction approach for conditional kaplan–meier estimators,” Test **27**(2), 295–315.

Lo, S. M., and Wilke, R. A. (**2010**). “A copula model for dependent competing risks,” Journal of the Royal Statistical Society: Series C (Applied Statistics) **59**(2), 359–376.

Lopez, O. (**2011**). “Nonparametric estimation of the multivariate distribution function in a censored regression model with applications,” Communications in Statistics-Theory and Methods **40**(15), 2639–2660.

Lopez, O., Patilea, V., and Van Keilegom, I. (**2013**). “Single index regression models in the presence of censoring depending on the covariates,” Bernoulli **19**(3), 721–747.

Maistre, S., and Patilea, V. (**2019**). “Nonparametric model checks of single-index assumptions,” Statistica Sinica **29**(1), 113–138.

Mammen, E. (**1993**). “Bootstrap and wild bootstrap for high dimensional linear models,” The annals of statistics **21**(1), 255–285.

Marjerrison, S., Antillon, F. *et al.* (**2013**). “Outcome of children treated for relapsed acute lymphoblastic leukemia in central america,” Cancer **119**(6), 1277–1283.

Mostert, S., Arora, R. S. *et al.* (**2011**). “Abandonment of treatment for childhood cancer: position statement of a siop podc working group,” The lancet oncology **12**(8), 719–720.

Navarrete, M., Rossi, E. *et al.* (**2014**). “Treatment of childhood acute lymphoblastic leukemia in central america: A lower-middle income countries experience,” Pediatric blood & cancer **61**(5), 803–809.

Nelsen, R. B. (**2007**). *An introduction to copulas* (Springer Science & Business Media).

Nolan, D., and Pollard, D. (**1987**). “U-processes: rates of convergence,” The Annals of Statistics 780–799.

Pakes, A., and Pollard, D. (**1989**). “Simulation and the asymptotics of optimization estimators,” Econometrica: Journal of the Econometric Society 1027–1057.

Peterson, A. V. (**1976**). “Bounds for a joint distribution function with fixed sub-distribution functions: Application to competing risks,” Proceedings of the National Academy of Sciences **73**(1), 11–13.

Pollard, D. (**1990**). “Empirical processes: theory and applications,” Ims.

Rivest, L.-P., and Wells, M. T. (**2001**). “A martingale approach to the copula-graphic estimator for the survival function under dependent censoring,” Journal of Multivariate Analysis **79**(1), 138–155.

Rosenbaum, P. R., and Rubin, D. B. (**1983**). “The central role of the propensity score in observational studies for causal effects,” Biometrika **70**(1), 41–55.

Rubin, D. B. (**1974**). “Estimating causal effects of treatments in randomized and nonrandomized studies.,” Journal of educational Psychology **66**(5), 688.

Sant’Anna, P. H. (**2016**). “Program evaluation with right-censored data,” arXiv preprint arXiv:1604.02642 .

Sant’Anna, P. H., Song, X., and Xu, Q. (**2022**). “Covariate distribution balance via propensity scores,” Journal of Applied Econometrics .

Sant’Anna, P. H. (**2021**). “Nonparametric tests for treatment effect heterogeneity with duration outcomes,” Journal of Business & Economic Statistics **39**(3), 816–832.

Sherman, R. P. (**1994**). “Maximal inequalities for degenerate u-processes with applications to optimization estimators,” The Annals of Statistics **22**(1), 439–459.

Sklar, M. (**1959**). “Fonctions de repartition an dimensions et leurs marges,” Publ. inst. statist. univ. Paris **8**, 229–231.

Strzalkowska-Kominiak, E., and Cao, R. (**2014**). “Beran-based approach for single-index models under censoring,” Computational Statistics **29**(5), 1243–1261.

Stute, W. (**1997**). “Nonparametric model checks for regression,” The Annals of Statistics 613–641.

Tsiatis, A. (**1975**). “A nonidentifiability aspect of the problem of competing risks.,” Proceedings of the National Academy of Sciences **72**(1), 20–22.

Van der Vaart, A. (**1998**). *Asymptotic statistics* (Cambridge Univ. Press, Cambridge).

Van Der Vaart, A. W., and Wellner, J. A. (**1996**). *Weak Convergence and Empirical Processes* (Springer, New York), pp. 16–28.

Xia, Y., Zhang, D., and Xu, J. (**2010**). “Dimension reduction and semiparametric estimation of survival models,” Journal of the American Statistical Association **105**(489), 278–290.

Zheng, M., and Klein, J. P. (**1995**). “Estimates of marginal survival for dependent competing risks based on an assumed copula,” Biometrika **82**(1), 127–138.