

Instrumental variable estimation of the proportional hazards model by presmoothing

Lorenzo Tedesco¹, Jad Beyhum², Ingrid Van Keilegom³

¹*Department of Economics, University of Bergamo,
e-mail: lorenzo.tedesco@unibg.it*

²*Department of Economics, KU Leuven,
e-mail: jad.beyhum@kuleuven.be*

³*ORSTAT, KU Leuven,
e-mail: ingrid.vankeilegom@kuleuven.be*

Abstract: We consider instrumental variable estimation of the proportional hazards model. The instrument and the endogenous variable are discrete but there can be (possibly continuous) exogenous covariables. We can reformulate the proportional hazards model into a semiparametric instrumental variable quantile regression model by assuming rank invariance. A naive estimation approach based on conditional moment conditions generated by the model would lead to a highly nonconvex and nonsmooth objective function. To overcome this problem, we propose a new presmoothing methodology. First, we estimate the model nonparametrically - and show that this nonparametric estimator has a closed-form solution in the leading case of interest of randomized experiments with one-sided noncompliance. Second, we use the nonparametric estimator to generate “proxy” observations for which exogeneity holds. Third, we apply the usual partial likelihood estimator to the “proxy” data. While the paper focuses on the proportional hazards model, our presmoothing approach could be applied to estimate other semiparametric formulations of the instrumental variable quantile regression model. Our estimation procedure allows for random right-censoring. We show asymptotic normality of the resulting estimator. The approach is illustrated via simulation studies and an empirical application to the Illinois Unemployment Incentive Experiment.

MSC2020 subject classifications: Primary 62N, 62P20; secondary 62G08, 62G07, 62M.

Keywords and phrases: Proportional hazards model, instrumental variable, duration models, presmoothing, nonparametric estimation.

Received August 2024.

1. Introduction

The proportional hazards (henceforth, PH) model, first introduced in [14]’s seminal paper, has become a popular tool for analyzing time-to-event data, such as the unemployment duration of job seekers or the duration until the bankruptcy of some firms. It is a semiparametric model of the hazard rate, which has the convenient property that the nonparametric component of the model - called

the baseline hazard - can be profiled out in the likelihood. The resulting partial likelihood allows for a simple and stable estimation procedure of the parametric part of the model.

In practice, observational data may be affected by unmeasured confounding that influences both the treatment and the outcome, creating an endogeneity issue. In econometrics, a common solution to the endogeneity problem is the use of an instrumental variable (henceforth, IV) approach. There exists several different IV models, which differ in terms of the unobserved heterogeneity that they allow.

In this paper, we address the problem of endogeneity in the PH model by embedding the problem in the instrumental variable quantile regression (henceforth, IVQR) model introduced in [12]. Indeed, by making a rank invariance assumption, we reformulate the proportional hazards model into a semiparametric version of the instrumental variable quantile regression model of [12]. This allows us to use some of the tools from the literature on the IVQR model. The IVQR model and its associated estimation procedures have become popular in econometrics research, see Chapter 9 in [22] for a recent review. By constraining the treatment effects to only vary by quantile, this approach allows estimating quantile treatment effects over the full population. This is in contrast to the local average treatment effects framework of [3], which does not restrict the heterogeneity of treatment effects but requires a monotonicity assumption to hold and only allows to estimate treatment effects on subpopulations. See [33] for a detailed comparison of the two approaches.

One of the main difficulties generated by the IVQR model is estimation. The causal regression function of interest can be shown to be the solution of a system of nonlinear integral equations. A natural approach to estimation is to solve an empirical version of this system. However, as noted in [13], when the causal regression function follows a (semi)parametric model this leads to a severely nonsmooth and nonconvex objective function. For the linear-in-parameters semiparametric quantile regression model, some ingenious solutions to this problem have been developed in [13], [21] and [20] among others. Based on our knowledge, no general method for estimating the semiparametric IVQR relying on convex optimization exists.

In this paper, we focus on the case where the endogenous variable and the instrument are both discrete and there are (possibly continuous) exogenous covariables. To simplify the exposition, we assume that the endogenous variable and the instrument have the same number of modalities (i.e. the same cardinality of the support). This setting is very relevant in practice since it includes randomized experiments with noncompliance (where both the instrument and the endogenous variable are binary), which is the main focus of [12], or, more recently, [32]. In the remainder of the paper, we often call “treatment” the discrete endogenous variable because of this link. As common with time-to-event data, we assume that the outcome variable is randomly right-censored, where the censoring time is conditionally independent of the outcome.

We propose a new solution to the problem of estimating IVQR models with a causal semiparametric quantile regression function. We apply this new method-

ology to the PH model but stress that it could be useful in other semiparametric models, such as the proportional odds model or distribution regression. This approach is based on a *presmoothing* strategy. Our procedure can be summarized in three steps.

1. First, we estimate nonparametrically the causal quantile regression function.
2. Second, we use this nonparametrically estimated causal quantile regression function to generate “proxy” observations for which the treatment is exogenous.
3. Finally, we apply the celebrated partial likelihood estimator of [14] to the “proxy” observations.

For Step 1, we use [4]’s estimator of the conditional survival function to construct an empirical version of the system of equations generated by the model and then solve the estimated system to obtain the causal regression function of interest. This allows us to deal with random right-censoring, a common feature of duration data. We note that Step 1 is the most critical phase of the estimation due to its nonparametric nature. However, the primary rationale behind the presmoothing methodology is to project an infinite-dimensional space onto a finite-dimensional one, thereby enhancing the robustness of the final parametric estimator. Specifically, the nonparametric estimation phase proposed in Step 1 is used in Steps 2 and 3 to estimate the parametric component of the model. Consequently, the challenge of obtaining a precise estimator for causal quantile regression becomes less critical, as the final parametric estimator is based on the average values derived from the nonparametric estimator. Moreover, the complexity of Step 1 is mitigated by the fact that the treatment and instrument are discrete, rendering the inverse problem well-posed. Additionally, in the practically relevant case of randomized experiments with *one-sided noncompliance* – where both the treatment and the instrument are binary, and there is full compliance in the control group – our nonparametric estimator has a closed-form expression. Finally, we observe that the extensive literature on [4]’s estimator can be leveraged to simplify, improve, and innovate the first step procedure.

The fact that we can use the partial likelihood estimator in Step 3 is very convenient since the latter has a convex objective function and is implemented in traditional statistical software. We stress again that the proposed procedure could be applied to other models (such as the proportional odds model or distribution regression) by replacing Step 3 with standard estimators of these models (under exogeneity). The main advantage of our procedure is to bring us back (in Step 2) to the case of exogenous data, which is well-studied and for which most semiparametric models were originally motivated. One may wonder why we do not use directly the nonparametric estimator of Step 1 to analyze the causal effect of the treatment variable. This is because the estimate of the regression coefficient in the Cox model is likely to be much more precise. Moreover, it has a direct interpretation as a causal hazard ratio.

We show that the resulting estimator is asymptotically normal. We also develop a theory for the nonparametric estimator of Step 1 which is new to the

literature. Simulations demonstrate that our estimator possesses good finite sample properties. Finally, the procedure is applied to evaluate the causal effect of reemployment bonuses using data from the Illinois Unemployment Incentive Experiment.

We call our methodology *presmoothing* because it is related to literature in (semi)parametric statistics advocating for estimation procedures that first use nonparametric estimators before projecting the nonparametric estimators on (semi)parametric spaces in a second step, see [15, 1] or, more recently, [23]. The main takeaway from this statistical literature is that, asymptotically, the two-step estimator recovers the property of the “oracle” second-step estimator (which assumes that the nonparametric object estimated in the first step is known). *Presmoothing* is particularly useful in cases where estimating *directly* the semiparametric model is complicated while nonparametric estimation is simpler. One of the contributions of this article is to bring this *presmoothing* tool to econometrics.

Finally, before outlining the different sections of the paper, we note that there already exist quite a few papers studying instrumental variable methods for duration models. To the best of our knowledge, no other paper considers the PH model embedded in the IVQR model. Please note, that for reasons of space, we do not cite all relevant works. [8] consider a mixed proportional hazards model in which the unobserved heterogeneity enters multiplicatively, that is differently from the IVQR model. [30] considers a PH model with endogeneity but they make an additive (rather than quantile) homogeneity of treatment effects assumption. As argued in [12], the restrictions on unobserved heterogeneity in the IVQR model can be seen as weaker. [5] studies the nonparametric IVQR model with random right censoring with discrete treatment and instrument but no covariates. The present paper generalizes [5] by allowing for continuous covariables in nonparametric estimation and studying the three-step estimation of the PH model. [19, 11, 29, 7] study the linear-in-parameters IVQR model under random right-censoring but do not cover the PH model. The way we reformulate the PH duration model into an IVQR model is related to the reformulation of the dynamic duration model into a dynamic IVQR model in [6].

The rest of the paper is organized as follows. In Section 2, we specify the model. Identification results are derived in Section 3. Section 4 is devoted to the estimation theory. In Section 5 we discuss the asymptotic theory of the proposed estimator. In Section 6, the finite sample performance of the proposed method is investigated through simulations. In Section 7, we illustrate the method by means of an empirical application of the method. Technical details and proofs can be found in the Appendix.

2. Model

2.1. The duration model

We are interested in the effect of an endogenous treatment Z on a non-negative outcome T . We consider a treatment $Z = (Z_1, \dots, Z_{d_Z})^\top$ expressed as a vector

of d_Z binary variables and taking L different values $\{z_1, \dots, z_L\} = \mathcal{Z}$. In our data example e.g., we have a categorical treatment with three levels that can be defined by means of two binary (dummy) variables, so $d_Z = 2$. Since there are three levels, $L = 3$ in this case and $z_1 = (0, 0)^\top$, $z_2 = (1, 0)^\top$ and $z_3 = (0, 1)^\top$. There are some covariables X with compact support $\mathcal{X} \subset \mathbb{R}^{d_X}$. Let $T(z)$ be the potential outcome duration under treatment equal to $z \in \mathcal{Z}$. Let us assume that the distribution of $T(z)$ given $X = x$ is continuous for all $x \in \mathcal{X}$ and also define by

$$\lambda(z, t|x) := \lim_{dt \rightarrow 0} \frac{\mathbb{P}(T(z) \in [t, t + dt] | T(z) \geq t, X = x)}{dt}$$

the structural hazard of $T(z)$ given $X = x$ at time $t \in \mathbb{R}_+$. We impose the consistency condition $T = T(Z)$. We also assume that $T(z)$ given $X = x$ follows a proportional hazards model, that is there exists a baseline hazard $\lambda_0 : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ and a vector of regression coefficients $\beta_0 = (\beta_{0,z}^\top, \beta_{0,x}^\top)^\top \in \mathbb{R}^{d_Z + d_X}$ such that

$$\lambda(z, t|x) = \lambda_0(t) \exp((z^\top, x^\top)\beta_0), \text{ for all } (z, x, t) \in \mathcal{Z} \times \mathcal{X} \times \mathbb{R}_+. \quad (1)$$

Remark that this proportional hazards assumption implies that the support of $T(z)$ given $X = x$ does not depend on z, x and is equal to that of λ_0 , or equivalently of T . We denote by $\mathcal{T}$ this support of λ_0 . Let us now define the structural cumulative hazard

$$\Lambda(z, t|x) := \int_0^t \lambda(z, s|x) ds = \Lambda_0(t) \exp((z^\top, x^\top)\beta_0),$$

where $\Lambda_0(t) := \int_0^t \lambda_0(s) ds$ is the cumulative baseline hazard. Let us also introduce $Q(z) := \Lambda(z, T(z)|X)$, which is the structural hazard of $T(z)$ given X evaluated at $T(z)$. The random variables $\{Q(z)\}_{z \in \mathcal{Z}}$ can be thought of as the unobserved heterogeneity of the model. We make the following Assumption.

Assumption 1. *The following holds*

- (i) *There exists a random variable Q such that $Q(z) = Q$ for all $z \in \mathcal{Z}$;*
- (ii) *The strictly positive baseline hazard λ_0 is continuous and its support $\mathcal{T}$ is a bounded interval.*

By Assumption 1 (ii), the structural cumulative hazard $\Lambda(z, \cdot|x)$ is strictly increasing on $\mathcal{T}$. Hence, we can define the inverse of $\Lambda(z, \cdot|x)$ on $\mathcal{T}$, which we denote by $\Lambda(z, \cdot|x)^{-1}$. This means that $\Lambda(z, \cdot|x)^{-1}(s)$ is the unique element t of $\mathcal{T}$ such that $\Lambda(z, t|x) = s$ (by definition the image of the cumulative hazard on $\mathcal{T}$ is $\mathbb{R}_+$ so that $\Lambda(z, \cdot|x)^{-1}(s)$ is defined on all $\mathbb{R}_+$). Because of this, we have $T(z) = \Lambda(z, \cdot|X)^{-1}(Q)$. Assumption 1 (i) then implies that for two subjects i and j such that $X_i = X_j$, $T_i(z) > T_j(z)$ implies $Q_i(z) = Q_i > Q_j = Q_j(z)$, which leads to $T_i(z') > T_j(z')$, for all $z, z' \in \mathcal{Z}$. Hence, under Assumption 1 (i) the rank in the outcome of any two subjects is the same across all potential outcomes. Because of this, Assumption 1 (i) is a rank invariance assumption as in [12]. This assumption restricts the heterogeneity of the treatment effects on the duration: the treatment can change the quantiles of the distribution

of the potential outcomes but it cannot change the rank that a subject has in this distribution. Moreover, note that the rank invariance assumption does not restrict the possible values of the structural hazard $\lambda(z, t|x)$ beyond the continuity condition in Assumption 1 (ii) and therefore, in this sense, rank invariance is not a constraint on the marginal distribution of $T(z)$ given X . It only imposes limits on the joint distribution of potential outcomes, that is the distribution of $(T(z))_{z \in \mathcal{Z}}$. This rank invariance assumption could be relaxed into a rank similarity assumption as in [12], which corresponds to assuming that $Q(z)|X, W, Z$ is distributed as a random variable Q for each $z \in \mathcal{Z}$. While all the results can be obtained under such a relaxed assumption (see [12]), we considered the more stringent assumption as it simplifies the argument and its meaning might be clearer for the reader. Remark that Assumption 1 allows us to write

$$\Lambda_0(T) \exp((Z^\top, X^\top)\beta_0) = Q. \quad (2)$$

We assume to observe a categorical instrumental variable W . To simplify the theoretical arguments, we also suppose that the instrument has the same number of modalities as the treatment. Therefore, W has support $\mathcal{W} = \{w_1, \dots, w_K\} \subset \mathbb{R}^{d_W}$. Differently from the variable Z , the instrument W does not need to consist of binary variables. Importantly, assuming that the cardinalities of the support of Z and W are equal does not limit our discussion to a specific case. First, it would be possible, but burdensome, to extend both identification and estimation results to the case where the number of modalities of W is greater or equal to the number of modalities of Z . Second, it is always possible to obtain the same number of modalities between Z and W by aggregation of modalities of W . We impose the following condition, where we denote by Q the variable $Q(X)$.

Assumption 2. $W \perp\!\!\!\perp Q|X$,

where ‘ $\perp\!\!\!\perp$ ’ stands for statistical independence. This assumption formalizes the exogeneity of the instrument W when conditioning with respect to the co-variables X . The duration T is randomly right censored by a random variable C with support in $\mathbb{R}_+$ so that we do not observe T but $Y = \min(T, C)$. The observables are (Y, δ, Z, X, W) , where $\delta = I(T \leq C)$. We impose the following standard independent censoring assumption:

Assumption 3. $T \perp\!\!\!\perp C|Z, X, W$.

2.2. Reformulation as a semiparametric IVQR model

In this section, thanks to Assumption 1, we reformulate the duration model into a semiparametric IVQR model. Let us define, for $u \in [0, 1]$,

$$\varphi(z, x, u) = \Lambda_0^{-1} \left(\frac{-\log(1-u)}{\exp((z^\top, x^\top)\beta_0)} \right). \quad (3)$$

and $U = 1 - \exp(-Q)$. By standard arguments in duration analysis, $Q|X$ follows a unit exponential distribution, so that $U|X$ follows a standard uniform

distribution. Then, inverting equation (2), we obtain that

$$T = \varphi(Z, X, U), \quad U|W, X \sim \mathcal{U}[0, 1], \quad (4)$$

where $\mathcal{U}[0, 1]$ stands for the uniform distribution on $[0, 1]$. Remark that by Assumption 1 (ii) and the inverse function theorem, $\varphi(z, x, \cdot)$ is strictly increasing and differentiable with continuous derivative on $[0, 1]$, for all $(z, x) \in \mathcal{Z} \times \mathcal{X}$. As claimed in the introduction, this is an IVQR model as in [12], where the causal quantile regression function φ follows the semiparametric model given in (3), which is generated by the proportional hazards assumption on the structural hazard. From now on, this paper will analyze directly this semiparametric IVQR model.

Notation. Finally, we introduce some additional notation. Define $V = (Z, X) \in \mathcal{V} = \mathcal{Z} \times \mathcal{X} \subset \mathbb{R}^{d_V}$, where $d_V = d_Z + d_X$. For $v = (z, x)$, we write $\varphi(v, \cdot)$ for $\varphi(z, x, \cdot)$. For $l = 1, \dots, L$, we also write $\varphi_l^x(\cdot)$ for $\varphi(z_l, x, \cdot)$ and $\varphi^x(\cdot)$ for $(\varphi_l^x(\cdot))_{l=1}^L$.

3. Identification

In this section, we provide an identification result of the parameter of interest β_0 . The identification can be obtained in two steps. First, we analyze the identification of φ . Second, we study the identification of β_0 given the first step. Note that we keep the analysis of identification brief since it does not constitute the main contribution of the paper (which is the estimation procedure).

3.1. Identification of φ

Let us first recall the standard characterization of φ as the solution to a system of nonlinear equations as in [12], that is, for every $x \in \mathcal{X}$, $\varphi^x(u)$ is a solution of the following system of equations in $(\theta_l)_{l=1}^L \in \mathbb{R}_+^L$:

$$\sum_{l=1}^L F(\theta_l, z_l | x, w_k) = u \quad \text{for } k = 1, \dots, K, u \in (0, 1), \quad (5)$$

where $F(t, z | x, w) = P(T \leq t, Z = z | X = x, W = w)$. Indeed, by the fact that $\varphi(z, x, \cdot)$ is strictly increasing and the independence between U and (W, X) , for all $k \in \{1, \dots, K\}$, it follows that:

$$\begin{aligned} \sum_{l=1}^L F(\varphi_l^x(u), z_l | x, w_k) &= \sum_{l=1}^L P(T \leq \varphi_l^x(u), Z = z_l | X = x, W = w_k) \\ &= \sum_{l=1}^L P(\varphi_l^x(U) \leq \varphi_l^x(u), Z = z_l | X = x, W = w_k) \end{aligned}$$

$$\begin{aligned}
&= \sum_{l=1}^L P(U \leq u, Z = z_l | X = x, W = w_k) \\
&= P(U \leq u | X = x, W = w_k) = u.
\end{aligned}$$

[12] give conditions under which there is a unique solution to (5). In the context of the PH model, there is an additional layer of complexity due to the fact that F is not identified everywhere because of censoring. For $z \in \mathcal{Z}, x \in \mathcal{X}$ and $w \in \mathcal{W}$, let us denote by $c_{z,x,w}$ the upper bound of the support of the distribution of C given $Z = z, X = x, W = w$. Also, given two real numbers c, d , we denote by $c \wedge d$ the value $\min(c, d)$. Moreover, let $\bar{t}$ be the upper bound of $\mathcal{T}$. Thanks to Assumption 3 and by standard “Kaplan-Meier type” arguments from the duration analysis literature, $F(\cdot, z|x, w)$ is identified on $[0, \bar{t} \wedge c_{z,x,w}]$, and $\bar{t} \wedge c_{z,x,w}$ is itself identified. Let us now introduce

$$\bar{u}^x = \inf_{z \in \mathcal{Z}, w \in \mathcal{W}} \varphi(z, x, \cdot)^{-1}(\bar{t} \wedge c_{z,x,w}),$$

where $\varphi(z, x, \cdot)^{-1}(t) = \inf\{u \in \mathbb{R}_+ : \varphi(z, x, u) \geq t\}$ is the pseudo-inverse of $\varphi(z, x, u)$. For $u \leq \bar{u}^x$, the left-hand-side of (5) will be identified at $(\theta_\ell)_{\ell=1}^L = \varphi^x(u)$ for all $x \in \mathcal{X}$, so that identification can follow along the lines of [12]. We provide below an identification result assuming that the solutions to the system (5) are unique, a property for which sufficient conditions can be found in [12].

Theorem 3.1. *Let Assumptions 1, 2 and 3 hold and assume that (5) has a unique solution in $\mathbb{R}_+^L$ for all $u \in [0, 1]$. Then, $\bar{u}^x$ is identified and $\varphi(z, x, u)$ is identified for all $z \in \mathcal{Z}, x \in \mathcal{X}, u \in [0, \bar{u}^x]$.*

In certain applied situations, such as with one-sided non-compliance, it is possible to establish straightforward and interpretable identification conditions for φ . We explore this topic in more detail in Appendix G.

3.2. Identification of β_0

The proof of identification of β_0 is constructive, meaning that we can use its rationale to estimate β_0 in the subsequent section. The proof uses the knowledge of φ to solve the problem of identifying β_0 under exogeneity.

Formally, let $\bar{u} := \inf_{x \in \mathcal{X}} \bar{u}^x$. From Theorem 3.1, φ is identified for $(z, x, u) \in \mathcal{Z} \times \mathcal{X} \times [0, \bar{u}]$. As $\bar{u}$ is not known precisely in practice, we choose a point $\bar{U}$ such that $0 < \bar{U} < \bar{u}$. We then define two random variables U_g and U_g^c such that $(U_g, U_g^c) \perp (Z, X)$. The subscript ‘g’ highlights that these random variables are generated. The variable U_g is drawn from a uniform distribution $\mathcal{U}[0, 1]$, is independent of U_g^c , and serves as the generative process. U_g^c acts as a censoring variable for U_g with support in $[0, \bar{U}]$, ensuring $\min(U_g, U_g^c)$ does not exceed $\bar{U}$, thus respecting the identification constraints. Hence, any continuous distribution on $[0, \bar{U}]$ for U_g^c is acceptable and we discuss in Appendix A preferable choices.

We then define the random variables $T_g = \varphi(Z, X, U_g)$, $C_g = \varphi(Z, X, U_g^c)$, $Y_g = T_g \wedge C_g$ and $\Delta = I(T_g \leq C_g)$. By Theorem 3.1, the distribution of

(Y_g, Δ, Z, X) is identified. As shown in Section 2, the random variable T_g follows a PH model with parameter β_0 , where exogeneity holds as U_g is independent of (Z, X) . Furthermore, the censoring mechanism is conditionally independent as $T_g \perp\!\!\!\perp C_g | Z, X$ by construction. Hence, β_0 is identifiable from the distribution of (Y_g, Δ, Z, X) under standard assumptions for the identification of the PH model (e.g., see [24]).

This argument leads us to the following theorem.

Theorem 3.2. *Let the assumptions of Theorem 3.1 hold and assume that $\bar{u} > 0$ and $E[VV^\top]$ has full rank, then β_0 is identifiable.*

For later theoretical discussions, it will be beneficial to set $\tilde{U} = \min(U_g, U_g^c)$. We will also employ the equalities $Y_g = \varphi(Z, X, \tilde{U})$ and $\Delta = I(U_g \leq U_g^c)$.

4. Estimation

In this section, we discuss the estimation of the parameter vector β_0 , given an independent and identically distributed (iid) sample $\{(Y_i, \delta_i, Z_i, X_i, W_i)\}_{i=1}^n$ of observables. We first outline a naive, but ultimately unsuitable approach to estimation before presenting the three steps of the estimation procedure mentioned in the introduction. For the sake of simplicity, we restrict ourselves to the case $d_X = 1$, so that the already involved conditions we state next do not depend on the parameter d_X . Note that this method can also be applied to higher-dimensional covariate vectors. However, it is important to exercise caution, as nonparametric techniques are susceptible to the curse of dimensionality. Therefore, it is advisable for researchers to consider using dimension reduction techniques, such as those presented in [16].

4.1. Naive approach

4.1.1. Direct substitution

A naive approach to estimation would proceed as follows. By substituting equation (3) into (5), we obtain

$$\sum_{l=1}^L F\left(\Lambda_0^{-1}\left(\frac{-\log(1-u)}{\exp((z_l^\top, x^\top)\beta_0)}\right), z_l \middle| x, w_k\right) = u \quad \text{for } k = 1, \dots, K, u \in [0, \bar{u}).$$

Thus, since for any $k = 1, \dots, K$, for any $u \in [0, \bar{u})$, and for any $x \in \mathcal{X}$ we have that $\sum_{l=1}^L F\left(\Lambda_0^{-1}\left(\frac{-\log(1-u)}{\exp((z_l^\top, x^\top)\beta_0)}\right), z_l \middle| x, w_k\right) - u = 0$, we can integrate over u , sum over k and take the expected value with respect to X in order to obtain:

$$E\left[\int_0^{\bar{u}} \sum_{k=1}^K \left[\sum_{l=1}^L F\left(\Lambda_0^{-1}\left(\frac{-\log(1-u)}{\exp((z_l^\top, X^\top)\beta_0)}\right), z_l \middle| X, w_k\right) - u\right]^2 du\right] = 0. \quad (6)$$

One could estimate F by a conditional [4]'s estimator (see Section 4.2.1 for its definition) and then obtain an empirical analog to equation (6). This empirical criterion could be minimized over Λ_0 and β_0 . As noted in the introduction, this approach would result in a highly nonsmooth and nonconvex problem.

4.1.2. Standard presmoothing methodology

The practice of estimating a (semi)parametric model based on a previously estimated nonparametric model was first proposed in the literature to enhance robustness in the estimation procedure, as discussed by [15, 1] and more recently by [23]. In these approaches, the quantity of interest is initially estimated nonparametrically, and this estimator is then used as the outcome for estimating the (semi)parametric model.

However, directly using the estimator $\hat{\varphi}$ to estimate β_0 and Λ_0 would again result in a highly non-convex problem. Specifically, this approach would require the use of equation (3) and thus estimating (Λ_0, β_0) based on the minimization of an objective function $L(\Lambda, \beta)$ of the form

$$L(\Lambda, \beta) = E \left[\int_0^{\bar{U}} \left\{ \hat{\varphi}(Z, X, u) - \Lambda^{-1} \left(\frac{-\log(1-u)}{\exp((Z^\top, X^\top)\beta)} \right) \right\}^2 du \right],$$

which presents similar difficulties as those in (6). Therefore, we take this method a step further, as explained in the following section.

4.2. Our estimation procedure

4.2.1. Step 1: Nonparametric estimator of φ

We first build a nonparametric estimator of φ . The estimator is based on (5), where F is first estimated via a conditional Beran's estimator [4].

Step 1.1: Nonparametric estimator of F under random right censoring

Remark that $F(t, z|x, w) = F(t|z, x, w)p_{z,x,w}$, where $F(t|z, x, w) = P(T \leq t|Z = z, X = x, W = w)$ and $p_{z,x,w} = P(Z = z|X = x, W = w)$. We propose to estimate $F(t|z, x, w)$ non-parametrically using the [4] estimator, given by

$$\tilde{F}(t|z, x, w) = 1 - \prod_{j=1}^n \left[1 - \frac{B_h^{z,w}(x - X_j, Z_j, W_j)}{\sum_{k=1}^n I(Y_k \geq Y_j) B_h^{z,w}(x - X_k, Z_k, W_k)} \right]^{\eta_j(t)}, \quad (7)$$

where $\eta_j(t) = I(Y_j \leq t, \delta_j = 1)$ and $B_h^{z,w}(x - X_k, Z_k, W_k)$ is a sequence of non-negative weights adding up to 1. In our case, we adopt the Nadaraya-Watson type weights, which are specified as follows:

$$B_h^{z,w}(x - X_k, Z_k, W_k) = \frac{K(\frac{x-X_k}{h})I(Z_k = z, W_k = w)}{\sum_{i=1}^n K(\frac{x-X_i}{h})I(Z_i = z, W_i = w)}, \quad (8)$$

where $K(\cdot)$ is a univariate kernel function, and $h = h_n$ is a bandwidth depending on n and converging to zero as $n \rightarrow \infty$. To ensure that the final criterion function is smooth, we smooth $\tilde{F}$ with respect to time. Consider a further kernel $\tilde{K}$ and define $H(t) = \int_{-\infty}^t \tilde{K}(u) du$, with a bandwidth $\epsilon = \epsilon_n$ depending on n and converging to zero as $n \rightarrow \infty$. We obtain

$$\hat{F}(t|z, x, w) = \int H\left(\frac{t-u}{\epsilon}\right) d\tilde{F}(u|z, x, w). \quad (9)$$

Consider now the following estimator for the quantity $p_{z,x,w}$:

$$\hat{p}_{z,x,w} = \frac{\sum_{i=1}^n I(Z_i = z, W_i = w) K\left(\frac{x-X_i}{h}\right)}{\sum_{i=1}^n I(W_i = w) K\left(\frac{x-X_i}{h}\right)}.$$

The final estimator for $F(t, z|x, w)$ is given by

$$\hat{F}(t, z|x, w) = \hat{F}(t|z, x, w) \hat{p}_{z,x,w}. \quad (10)$$

Step 1.2: Final nonparametric estimator of φ

Let $\bar{T} < \infty$ be an upper bound on $\max_{v \in \mathcal{V}} \varphi(v, \bar{U})$, let $\mathcal{F}_Z^{\bar{U}, \bar{T}} = \{f : \mathcal{Z} \times [0, \bar{U}] \rightarrow [0, \bar{T}]\}$ and let $\mathcal{F}_\uparrow^{Z, W}$ be the set of maps from $\mathbb{R}_+ \times \mathcal{Z} \times \mathcal{W}$ to $\mathbb{R}_+$ that are continuous and increasing in the first argument. Let $x \in \mathcal{X}$ be fixed. We denote by $F^x(t, z|w)$ the function $F(t, z|x, w)$. We have shown that for each $x \in \mathcal{X}$, φ^x belongs to the set of solutions to the equation:

$$A(\varphi^x, F^x) \equiv 0, \quad (11)$$

where A is an operator that associates to an element $(\tilde{\varphi}^x, \tilde{F}^x) \in \mathcal{F}_Z^{\bar{U}, \bar{T}} \times \mathcal{F}_\uparrow^{Z, W}$, a map from $[0, \bar{U}]$ to $\mathbb{R}^K$, denoted by $A(\tilde{\varphi}^x, \tilde{F}^x)$ and defined as

$$A(\tilde{\varphi}^x, \tilde{F}^x)(u) = \left(\sum_{l=1}^L \tilde{F}^x(\tilde{\varphi}_l^x(u), z_l|w_k) - u \right)_{k=1}^K,$$

where we wrote $\tilde{\varphi}_l^x(u)$ for $\tilde{\varphi}^x(z_l, u)$. Thus, we define the estimator of φ^x by

$$\hat{\varphi}^x \in \arg \min_{\theta \in \mathcal{F}_Z^{\bar{U}, \bar{T}}} \|A(\theta, \hat{F}^x)\|, \quad (12)$$

where $\|\cdot\|$ denotes the Euclidean norm.

4.2.2. Step 2: “Proxy” generation process

We now show how to use the nonparametric estimator obtained in Step 1 to generate “proxy” observations for which the treatment is exogenous. For each $i = 1, \dots, n$, we generate a random observation of (U_g, U_g^c) , namely $(U_{g,i}, U_{g,i}^c)$, from which we obtain $\tilde{U}_i = \min(U_{g,i}, U_{g,i}^c)$ and $\Delta_i = I(U_{g,i} \leq U_{g,i}^c)$. The variables U_g

and U_g^c are independent, with U_g following a uniform distribution $\mathcal{U}[0, 1]$. For the theory, we let U_g^c have a uniform distribution on the interval $[\bar{U} - \tau, \bar{U}]$ for some $\tau > 0$. Note that the theoretical results can be shown to be valid for any chosen distribution of U_g^c , provided that it possesses a continuous density with support contained in $[0, \bar{U}]$. In practical applications and simulations, for simplicity, we let $\tau \rightarrow 0$ and U_g^c is instead a degenerate random variable with a point mass at $\bar{U}$. Using the notation introduced in Section 2.2, the proxy observations are represented as $\{(\hat{Y}_{g,i}, \Delta_i, V_i)\}_{i=1}^n$, where $\hat{Y}_{g,i} = \hat{\varphi}(V_i, \bar{U}_i) = \hat{\varphi}^{X_i}(Z_i, \bar{U}_i)$, and $V_i = (Z_i, X_i)$.

4.2.3. Step 3: Partial likelihood estimator based on the “proxy” observations

The last step involves obtaining an estimator $\hat{\beta}$ for β_0 using the standard partial likelihood estimator of [14] based on the “proxy” observations obtained in Step 2. Denote by $\mathcal{B}$ a compact set that contains β_0 as an internal point. Thus, $\hat{\beta}$ is obtained as the minimizer of the standard score function for a PH model, and so

$$\hat{\beta} \in \arg \min_{\beta \in \mathcal{B}} \left\| n^{-1} \sum_{i: \Delta_i=1} V_i^\top \beta - n^{-1} \sum_{i=1}^n \log \left(n^{-1} \sum_{j: \hat{Y}_{g,j} \geq \hat{Y}_{g,i}} \exp(V_j^\top \beta) \right) \right\|. \quad (13)$$

Note that, the same procedure can indeed be applied to obtain an estimator of Λ_0 based on proxies, using standard estimation of the baseline cumulative hazard in a proportional hazards model. However, the aim of the paper concerns the estimation of the parametric component of the model as it can be easily interpreted in a causal way, and therefore we only consider the estimator defined in (13).

4.2.4. A comment on the validity of the proposed estimation method

The validity of the proposed three-step procedure, which will establish the consistency of the estimator $\hat{\beta}$, is rooted in its structure. In Step 1, the causal quantile regression function φ is recovered using the estimator $\hat{\varphi}$. Building on this, Step 2 leverages equation (3), which implies that $\varphi(Z, X, U_g)$ follows a PH model. This result enables the generation of proxy observations for this variable, where the treatment is exogenous due to equation (4). In Step 3, these proxies can therefore be used to estimate β_0 through the standard partial likelihood estimation method. The consistency of the estimator $\hat{\varphi}$ allows us to decompose $\hat{\beta}$ into the sum of an empirical process and a negligible term. This decomposition simplifies the derivation of the asymptotic properties of $\hat{\beta}$.

We observe that the influence of variations in the proxies $\hat{Y}_{g,i}$ on the variance of $\hat{\beta}$ is shaped by three key factors: the variance of the estimator $\hat{\varphi}$, the distribution of U_g^c , and the distributions of X , Z , and W . Specifically, a lower variance in $\hat{\varphi}$, reduced censoring in the proxy observations, and lower variability in X ,

Z , and W collectively contribute to a reduction in the variance of $\hat{\beta}$. However, the process of generating proxies—namely, the randomness of U_g and the fact that different samples of U_g might result in slightly different $\hat{\beta}$ values—does not affect the asymptotic variance of $\hat{\beta}$.

4.3. Computation

We now discuss the optimization program (12), noticing that the objective function $\|A(\theta, \hat{F}^x)\|$ might have multiple local minima. Therefore, the computation of the estimator $\hat{\varphi}^x$ can require starting the optimization algorithm at different initialization points. However, when the system of equations provided in (5) reduces to a triangular system, the estimator is easier to compute. This means that, after possibly relabeling the points in $\mathcal{Z}$ and $\mathcal{W}$, for a fixed value $u \in [0, \bar{U}]$ the system takes the form:

$$\sum_{l=1}^k F(\theta_l, z_l | x, w_k) = u \quad \text{for } k = 1, \dots, K. \quad (14)$$

This case occurs, for instance, in the presence of one-sided noncompliance, that is, Z and W are binary with support $\mathcal{Z} = \mathcal{W} = \{0, 1\}$, and $P(Z = 1 | W = 0, X) = 0$. In fact, the corresponding system takes the form:

$$\begin{cases} F(\theta_1, 0 | x, 0) = u \\ F(\theta_1, 0 | x, 1) + F(\theta_2, 1 | x, 1) = u. \end{cases}$$

Another example, in which Z and W are not binary, corresponds to the empirical application discussed in Section 7.

When the system is of the form provided in (14), the estimator has a closed-form expression based on the following inductive algorithm:

$$\begin{aligned} \hat{\varphi}(z_1, x, u) &= \hat{F}(\cdot, z_1 | x, w_1)^{-1}(u), \quad \text{and} \\ \hat{\varphi}(z_k, x, u) &= \hat{F}(\cdot, z_k | x, w_k)^{-1}\left(u - \sum_{l=1}^{k-1} \hat{F}(\hat{\varphi}(z_l, x, u), z_l | x, w_k)\right), \quad \text{for } k \geq 2, \end{aligned} \quad (15)$$

where $\hat{F}(\cdot, z | x, w)^{-1}(u) = \inf\{t \in \mathbb{R}_+ : \hat{F}(t, z | x, w) \geq u\}$ denotes the pseudo-inverse of the function $\hat{F}(\cdot, z | x, w)$. Note that, based on the assumptions of the model, $F(\cdot, z | x, w)$ is strictly monotone increasing on $[0, \bar{T}]$, for each $z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}$. In addition, by its definition, $\hat{F}(t, z | x, w)$ is monotone increasing in $t \in [0, \bar{T}]$, and this implies the inversion procedure in (15) is well-posed.

In conclusion, the computation of $\hat{\varphi}^x$ only requires the inversion of some functions, which is easily available in any programming software. For instance, in $\mathbb{R}$, the inversion of a function can be performed using `uniroot`.

We summarize the algorithm for the estimation of $\hat{\beta}$ as follows:

1. For $i = 1, \dots, n$ generate $\tilde{U}_i$ and Δ_i .
2. For $i = 1, \dots, n$ estimate $\hat{\varphi}(Z_i, X_i, \tilde{U}_i)$, possibly using the closed-form expression given in (15).
3. Obtain $\hat{\beta}$ using the standard partial likelihood estimator based on the proxy observations $\{(\hat{\varphi}(Z_i, X_i, \tilde{U}_i), \Delta_i, V_i)\}_{i=1}^n$.

We highlight that, in the estimation process of our algorithm, the semi-parametric nature of our model plays a pivotal role in enhancing efficiency. This is particularly evident in the estimation of the function φ , which necessitates computation at only a finite set of points, thereby facilitating the estimation. Such an approach greatly amplifies the practical utility of our model across a diverse range of empirical scenarios.

We note that steps 2 and 3 can additionally be repeated multiple times to obtain several estimators of $\hat{\beta}$ using different proxies. The final estimator would then be the average of these individual estimators. This approach helps to regularize any potential deviations among the proxies. However, we will not explore this further to maintain the clarity and brevity of our exposition and also because the gain in terms of precision was minimal based on our simulations.

A last note is regarding the choices of $\bar{U}$ and $\bar{T}$. $\bar{T}$ must be a time point up to which we can identify F . Hence, it must be strictly lower than

$$\underline{y} = \inf_{z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} y_{z,x,w},$$

where $y_{z,x,w}$ denotes the upper bound of the support of Y given $Z = z, X = x, W = w$. In practice, the infimum $\underline{y}$ is identified from the data and therefore can be estimated. It is also often known from domain-specific knowledge. Once $\bar{T}$ is chosen, the value of $\bar{U}$ should be selected as a point that satisfies $\hat{\varphi}(z, x, \bar{U}) \leq \bar{T}$ for all $z \in \mathcal{Z}$ and $x \in \mathcal{X}$. To compute such a value, one can start from a low value of u and then increase u until $\hat{\varphi}(z, x, u)$ reaches $\bar{T}$ for at least one value of $z \in \mathcal{Z}$ and $x \in \mathcal{X}$.

5. Asymptotic theory

In this section, we will discuss the asymptotic results of the estimator $\hat{\beta}$. First, we will consider its consistency. Then, we will discuss its asymptotic normality.

5.1. Consistency

Define the following quantities:

$$\begin{aligned} G(t|z, x, w) &= P(C \leq t | Z = z, X = x, W = w); \\ F_{Y|Z,X,W}(y|z, x, w) &= P(Y \leq y | Z = z, X = x, W = w); \\ F_{Y,1|Z,X,W}(y|z, x, w) &= P(Y \leq y, \delta = 1 | Z = z, X = x, W = w); \\ f_{Z,X,W}(z, x, w) &= \frac{\partial}{\partial x} P(Z = z, X \leq x, W = w); \end{aligned}$$

$$f_{X,W}(x, w) = \frac{\partial}{\partial x} P(X \leq x, W = w).$$

Define also

$$\begin{aligned} f(t, z|w, x) &= \frac{\partial}{\partial t} F(t, z|x, w); \\ f(t|z, w, x) &= \frac{\partial}{\partial t} F(t|z, x, w); \end{aligned}$$

and

$$\begin{aligned} \hat{f}(t, z|w, x) &= \frac{\partial}{\partial t} \hat{F}(t, z|x, w); \\ \hat{f}(t|z, w, x) &= \frac{\partial}{\partial t} \hat{F}(t|z, x, w). \end{aligned}$$

Consider the following regularity assumption, where $\alpha \in (0, 1)$ corresponds to the order of Lipschitz continuity needed for later theoretical results. A further specification of the value of α is given in the Supplementary Material.

Assumption 4. *The following holds:*

- (i) *For each $(z, w) \in \mathcal{Z} \times \mathcal{W}$, the conditional distribution functions $F(t|z, x, w)$ and $G(t|z, x, w)$ have continuous derivatives in $t \in [0, \bar{T}]$ and $x \in \mathcal{X}$ up to the third order and bounded fourth-order (included mixed derivatives).*
- (ii) *The univariate kernel functions $K(\cdot)$ and $\tilde{K}(\cdot)$ are compactly supported. The kernel $K(\cdot)$ is a continuously differentiable function of order ν satisfying $\int K(u)du = 1$, and $\int u^j K(u)du = 0$ for $j < \nu$, where $\nu \geq 4$ is an integer. The kernel $\tilde{K}(\cdot)$ is a continuously differentiable function of order π satisfying $\int \tilde{K}(u)du = 1$, and $\int u^j \tilde{K}(u)du = 0$ for $j < \pi$, where $\pi \geq 3$ is an integer. In addition, the support of $\tilde{K}$ is $[-1, 1]$. Lastly, the kernel $\tilde{K}$ and its derivatives up to order $\pi - 1$ are equal to zero at the border of the support.*
- (iii) *For some finite constant $C > 0$ the bandwidths h and ϵ satisfy $h = Cn^{-u}$ and $\epsilon = Cn^{-w}$ with*

$$\frac{1}{4\nu} < u < \min\left(\frac{1}{2}, \frac{1}{5 + 2\alpha}\right)$$

and

$$\max\left(\frac{u(1 + \alpha)}{\pi}, u, \frac{1}{4\pi}\right) < w < \min\left(\frac{1}{3 + \alpha}, 2\nu u - 1, \frac{u\nu}{\alpha}, \frac{u(\nu + 1)}{3 + \alpha}, \frac{1 - u}{5 + 2\alpha}\right)$$

- (iv) *The functions $f_{Z,X,W}(z, x, w)$, $f_{X,W}(x, w)$ are bounded away from zero on their relative support and $\inf_{t \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} f(t|z, x, w) > 0$.*

A valid example of the values for u , w , α , ν , π is, for instance, $u = 0.143$, $w = 0.144$, $\alpha = 0.001$, $\nu = 4$, $\pi = 3$. With this choice, we can also construct a valid example for $K(x)$ and $\tilde{K}(x)$. Consider the function $\bar{K}(x) = I(|x| \leq 1) \frac{693}{512} (1 - x^2)^5$. Then, it is easy to check that $K(x) = \tilde{K}(x) = \frac{3}{2} \bar{K}(x) + \frac{1}{2} x \bar{K}'(x)$

satisfies condition (ii). Statements (i) and (iv) correspond to standard assumptions about the time-to-event and censoring distributions. Statement (ii) is necessary for the proof to cancel first and second-order terms in the expansion of the involved limiting process, while statement (iii) pertains to the bandwidth condition required for higher-order kernels. Similar conditions can be found in [16]. Note that (ii) and (iii) are primarily theoretical requirements related to the asymptotic results. In practical applications, these conditions can be relaxed, and standard smoothing techniques can be applied, as shown in Section 6.

Results on consistency and rate of convergence of the estimator $\hat{F}(t, z|x, w)$ to $F(t, z|x, w)$ can be found in Appendix C. This appendix also includes the consistency of $\hat{f}(t, z|x, w)$ in t and x up to order $2 + \alpha$. These results ensure the regularity and rate of convergence of the estimator $\hat{\varphi}$, which is crucial for the convergence rate of $\hat{\beta}$. Additionally, we demonstrate that the estimators $\hat{F}(t, z|x, w)$ and $\hat{f}(t, z|x, w)$ admit an iid representation. This will be employed to derive an analogous representation for $\hat{\varphi}$, leading to the asymptotic normality of $\hat{\beta}$.

To proceed, denote by $\|\cdot\|_\infty$ the infinity norm. The argument over which the supremum is taken may vary throughout the paper, and it is specified whenever any doubt might arise. Introduce the following assumption:

Assumption 5. For all $\varepsilon > 0$, there exists $\varsigma > 0$ such that for all $x \in \mathcal{X}$, we have

$$\inf_{\theta \in \mathcal{F}_Z^{\bar{U}, \bar{T}} : \|\theta - \varphi^x\|_\infty \geq \varsigma} \|A(\theta, F^x)\|_\infty \geq \varepsilon.$$

Assumption 5 is related to the shape of the objective function A and ensures that there is a unique solution to the system of equations (5) in $[0, \bar{U}]$ and, hence, that there is a unique minimum to the program (12) when $\hat{F}$ estimates F well enough. A similar condition is Assumption (C) of [5]. In the case of one-sided noncompliance, where Z and W are binary with support $\mathcal{Z} = \mathcal{W} = \{0, 1\}$, and $P(Z = 1|W = 0, X) = 0$, we have

$$A(\theta, F^x) = \begin{pmatrix} F^x(\theta_1, 0|0) - u \\ F^x(\theta_1, 0|1) + F^x(\theta_2, 1|1) - u \end{pmatrix},$$

so that $\|A(\theta, F^x)\|_\infty \geq |F^x(\theta_1, 0|0) - u|$ and a sufficient condition for Assumption 5 is that, for all $x \in \mathcal{X}$, the derivative of $\theta_1 \in \mathbb{R} \rightarrow F^x(\theta_1, 0|0)$ is bounded from below in $[0, \bar{T}]$ by a strictly positive constant.

Theorem 5.1. Let Assumption 5 and the conditions of Theorem 3.2 hold. Then, $\hat{\beta} - \beta_0 = o_p(1)$.

5.2. Asymptotic distribution

Denote the Fréchet derivative of A in its first argument at the point $(\tilde{\varphi}^x, \tilde{F}^x) \in \mathcal{F}_Z^{\bar{U}, \bar{T}} \times \mathcal{F}_\uparrow^{Z, W}$ by

$$(\Gamma(\tilde{\varphi}^x, \tilde{F}^x)(u))_{l,k} = -\tilde{f}^x(\tilde{\varphi}_l^x(u), z_l|w_k),$$

where the subscript l, k indicates the (l, k) th entry of the matrix for $l = 1, \dots, L$, $k = 1, \dots, K$, and $\dot{f}^x$ is the derivative in the first argument of $\tilde{F}^x$. Consider the following assumption.

Assumption 6. *There exists $\varsigma > 0$ such that the lowest eigenvalue of $\Gamma(\varphi^x, F^x)(u)$ is greater than ς for all $u \in [0, \bar{U}]$, $x \in \mathcal{X}$.*

Let us give a sufficient condition for Assumption 6 in the case of one-sided noncompliance, where Z and W are binary with support $\mathcal{Z} = \mathcal{W} = \{0, 1\}$, and $P(Z = 1|W = 0, X) = 0$. In this case, the matrix $\Gamma(\varphi^x, F^x)(u)$ takes the form

$$\Gamma(\varphi^x, F^x)(u) = - \begin{pmatrix} f^x(\varphi_0^x(u), 0|0) & 0 \\ f^x(\varphi_0^x(u), 0|1) & f^x(\varphi_1^x(u), 1|1) \end{pmatrix},$$

where f^x is the derivative in the first argument of F^x . Since the eigenvalues are continuous and $\Gamma(\varphi^x, F^x)(u)$ is continuous in $u \in [0, \bar{U}]$, $x \in \mathcal{X}$ by Assumption 4(i), Assumption 6 holds if for all $u \in [0, \bar{U}]$, $x \in \mathcal{X}$, the matrix $\Gamma(\varphi^x, F^x)(u)$ has full rank (i.e. its lowest eigenvalue is strictly positive). It can be shown (see the derivations at the end of Appendix G) that

$$f^x(t, z|w) = \frac{1}{(\varphi_z^x)'(\varphi_z^x(\cdot)^{-1}(t))} \mathbb{P}(Z = z|W = w, X = x, U = \varphi_z^x(\cdot)^{-1}(t)).$$

Hence, it holds that $f^x(\varphi_0^x(u), 0|0) = \frac{1}{(\varphi_0^x)'(u)} \mathbb{P}(Z = 0|W = 0, X = x, U = u)$ and $f^x(\varphi_1^x(u), 1|1) = \frac{1}{(\varphi_1^x)'(u)} \mathbb{P}(Z = 1|W = 1, X = x, U = u)$. Therefore, $\Gamma(\varphi^x, F^x)(u)$ has full rank for every $u \in [0, \bar{U}]$, $x \in \mathcal{X}$ and Assumption (6) holds if $\mathbb{P}(Z = 1|W = 1, X = x, U = u) > 0$ for all $u \in [0, \bar{U}]$, $x \in \mathcal{X}$, which means that every subject has a strictly positive probability of being treated ($Z = 1$) when assigned to the treatment group ($W = 1$).

Theorem 5.2. *Let Assumptions 4, 5, 6 and the conditions of Theorem 3.2 hold. Then $\sqrt{n}(\hat{\beta} - \beta_0)$ converges weakly to a mean zero Normal random variable.*

Note that Theorem 5.2 does not specify the asymptotic covariance matrix of $\sqrt{n}(\hat{\beta} - \beta_0)$ since the proof is based on showing that the estimator can be written as an empirical process over a Donsker class of functions. This result ensures asymptotic normality, but to obtain an explicit expression of the covariance matrix, one would need a complete representation of the empirical process. This is possible, but, as can be easily seen from the proof, the expression would be so involved that it would lack meaningful insights into the distribution and be difficult to use for practical inference purposes. Thus, for inference purposes, a bootstrap procedure can be used, the validity of which follows from replacing the functional central limit theorem based on empirical processes over a Donsker set of functions, with the bootstrap functional central limit theorem, see Theorem 3.7.1 in [25].

6. Simulations

In this section, we evaluate the finite sample performance of our proposed method using Monte Carlo simulation. Our main objective is to compare the

performance of the partial likelihood estimator proposed by Cox, an estimator that does not take endogeneity into account, with our new estimator. We will consider several designs, of which the specifications are given in Table 1. Since our method allows for the inclusion of both discrete and continuous exogenous variables X , we will consider several univariate distributions for X : the continuous Beta design, the continuous uniform design, and the discrete Bernoulli design.

For all designs, we set the structural baseline cumulative hazard function Λ_0 to that of an exponential variable with mean 1, i.e. $\Lambda_0(s) = s$, and fix the value of $\beta_0 = (\beta_{0,z}, \beta_{0,x})^\top$. This results in the potential outcome $T(z, x)$ following an exponential distribution with a rate (reciprocal of the mean) equal to $\exp(x\beta_{0,x})$ if $z = 0$ and $\exp(\beta_{0,z} + x\beta_{0,x})$ if $z = 1$. The specifications of the design are given in Table 1. It is worth noting that the variable Z is endogenous as it depends on the unobserved heterogeneity term U , which follows a standard uniform distribution on $[0, 1]$ and which is taken independent of (W, X) . Note that, by the definition of Q in (2), it follows that $Q = \log(1 - U)$, and therefore, W is a proper instrumental variable, as it is a random variable independent of Q (also unconditionally with respect to X) while dependent on Z .

In each design, the probability of $Z = 1$ given a certain value of the variable W is as follows: $P(Z = 1|W = 0) = 0$, and $P(Z = 1|W = 1)$ is around 0.54, mimicking the empirical application considered in the next section.

TABLE 1

Distribution and coefficient specifications for each design. Note that $B(p)$ stands for the Bernoulli distribution of parameter p , $\mathcal{U}(a, b)$ represents the uniform distribution on the interval $[a, b]$, $\text{Beta}(\alpha, \beta)$ represents the Beta distribution with shape parameters α and β , and ϵ is a standard normal variable. The random variable $U \sim \mathcal{U}[0, 1]$ is the unobserved heterogeneity of the model.

Design	W	X	Z	$(\beta_{0,z}, \beta_{0,x})$
continuous Beta	$B(0.5)$	$\text{Beta}(2, 5)$	$I(W = 1)I(0.5U - W + 0.45 + X + 0.5\epsilon \geq 0)$	$(0.7, 0.3)$
continuous uniform	$B(0.5)$	$\mathcal{U}(-0.5, 0.5)$	$I(W = 1)I(0.5U - W + 0.8 + X + 0.5\epsilon \geq 0)$	$(0.7, 0.3)$
discrete Bernoulli	$B(0.5)$	$B(0.5)$	$I(W = 1)I(0.5U - W + 0.65 + 0.3X + 0.5\epsilon \geq 0)$	$(0.7, 0.7)$

We set the censoring variable, C , to be distributed as an exponential variable with rate (reciprocal of the mean) λ . We then consider different levels of censoring, specifically 20% and 40%, by setting $\lambda = 0.43$ and $\lambda = 1.15$, respectively, for the discrete Bernoulli design, $\lambda = 0.30$ and $\lambda = 0.82$, respectively, for the continuous uniform design, and $\lambda = 0.33$ and $\lambda = 0.87$, respectively, for the continuous Beta design. Finally, with $Y = \min(T, C)$ and $\delta = I(T \leq C)$ we generate an iid sample of size $n = 500$ or 1000 observations having the same distribution as (Y, δ, Z, X, W) .

We employ the Epanechnikov kernel to smooth the variable X when estimating F^x and, for each $X = x$, we select the bandwidth with a data-driven direct plug-in estimate of the Integrated Mean Squared Error-optimal band-

width (mse-dpi). This corresponds to the default bandwidth selection proposed in the R package `nprobust`, for which we refer to [9]. In Appendix H, we show that varying the kernel and bandwidth leads to similar results. Moreover, our simulation studies showed that smoothing in the argument t during the estimation procedure for F^x leads to a little deterioration of the result and significantly increases the estimation time. Therefore, we consider smoothing in the argument t to be a theoretical requirement that permits to show the asymptotic normality of the estimator, but that can be safely avoided during the method's application. Similarly, we chose to use the Epanechnikov kernel for our simulations, which is a function of order $\nu = 2$, even though Assumption 4 specifies an order of $\nu \geq 4$ for the kernel, due to its widespread acceptance in the field.

The upper bound of C is infinite, allowing for any value of $\bar{U} \in (0, 1)$ to be theoretically chosen, and so we set $\bar{U} = 0.9$. As we explain in Supplementary Material, we can choose U_g^c having a degenerate distribution with a unique point mass at $\bar{U}$.

The results given in Table 2 are based on $N = 5000$ replications. We present the average bias, standard deviation, and Mean Squared Error (MSE) for each component of $\beta = (\beta_z, \beta_x)^\top$. Additionally, we report the coverage of 95% bootstrap confidence intervals. The intervals are constructed by estimating the standard deviation of the proposed estimator based on a naive bootstrap resampling procedure with replacement, and then plugging in the standard deviation in a normal approximation of the bounds of confidence intervals. To accelerate the simulations, we employed the method proposed in [17]. To evaluate the estimator on a component-wise basis, we also include the Root Mean Squared Error (RMSE) as a measure to assess the overall power of the estimator. The value is computed following the formula $RMSE = \sqrt{N^{-1} \sum_{j=1}^N \|\hat{\beta}^{(j)} - \beta_0\|^2}$, where $\hat{\beta}^{(j)}$ is the estimation of simulation j . For comparison purposes, the same quantities are also given for the partial likelihood estimator. The results show that the proposed estimator has low bias. Additionally, as the sample size increases, the bias, and coverage of the estimator approach their theoretical values of 0 and 0.95 respectively. The performance of the estimator improves when the proportion of censored observations is lower. It can be argued that the estimator performs better under the discrete Bernoulli design, which may be attributed to the simpler structure of this design, as the exogenous component is discrete rather than continuous. As expected, the standard partial likelihood estimator is biased.

In addition to the designs just discussed, we have also considered a further design involving two continuous covariates, both uniformly distributed. This design, made especially to study situations with more than one continuous external variable, is detailed in the Supplementary Material. The results from this additional design provide further insights into the performance of our estimator under more complex variable configurations.

TABLE 2

Simulation results for each design. We include for the proposed estimator and the standard partial likelihood estimator, for each level of censoring (Cens) and for different sample sizes (n), the average of the bias (Bias), the standard deviation (Sd) and the Mean Squared Error (MSE) for each component (Comp) of $\beta = (\beta_z, \beta_x)^\top$. For each case, we also include the Root Mean Squared Error value (RMSE). For the proposed estimator, we also provide the coverage of the 95% confidence intervals (CP95) for each component of β .

Design	Cens	n	Comp	proposed estimator					partial likelihood estimator			
				Bias	Sd	MSE	RMSE	CP95	Bias	Sd	MSE	RMSE
continuous Beta	20%	500	β_z	-0.075	0.283	0.086		0.939	-0.264	0.158	0.095	
			β_x	-0.037	0.596	0.356	0.482	0.962	0.106	0.448	0.212	0.440
	40%	1000	β_z	-0.054	0.212	0.048		0.941	-0.262	0.111	0.081	
			β_x	-0.030	0.432	0.187	0.345	0.962	0.099	0.318	0.111	0.368
		500	β_z	-0.042	0.316	0.102		0.959	-0.296	0.181	0.120	
			β_x	-0.007	0.768	0.590	0.595	0.967	0.126	0.521	0.288	0.503
continuous uniform	20%	1000	β_z	-0.044	0.228	0.054		0.958	-0.293	0.128	0.102	
			β_x	0.010	0.549	0.301	0.416	0.974	0.121	0.364	0.147	0.419
		500	β_z	-0.051	0.280	0.081		0.956	-0.231	0.158	0.078	
			β_x	-0.034	0.339	0.116	0.314	0.966	0.087	0.253	0.072	0.325
	40%	1000	β_z	-0.041	0.199	0.041		0.949	-0.232	0.110	0.066	
			β_x	-0.022	0.248	0.062	0.232	0.958	0.081	0.178	0.038	0.287
		500	β_z	-0.019	0.312	0.098		0.964	-0.258	0.181	0.099	
			β_x	-0.097	0.494	0.254	0.422	0.959	0.105	0.291	0.096	0.370
discrete Bernoulli	20%	1000	β_z	-0.029	0.215	0.047		0.958	-0.259	0.126	0.083	
			β_x	-0.066	0.350	0.127	0.290	0.972	0.099	0.204	0.052	0.326
	40%	500	β_z	-0.009	0.283	0.080		0.954	-0.245	0.156	0.084	
			β_x	0.012	0.231	0.053	0.261	0.962	0.031	0.155	0.025	0.290
		1000	β_z	-0.011	0.202	0.041		0.952	-0.246	0.107	0.072	
			β_x	0.014	0.159	0.025	0.185	0.964	0.030	0.107	0.012	0.271
	40%	500	β_z	0.041	0.308	0.096		0.955	-0.270	0.180	0.105	
			β_x	-0.008	0.263	0.069	0.288	0.970	0.039	0.176	0.033	0.324
	40%	1000	β_z	0.026	0.215	0.047		0.952	-0.274	0.122	0.090	
			β_x	0.001	0.176	0.031	0.204	0.968	0.038	0.122	0.016	0.303

7. Empirical application

The Illinois Unemployment Incentive Experiment was a controlled social experiment conducted by the Illinois Department of Employment Security in 1984 to evaluate if cash bonuses reduce the duration of unemployment. New claimants for Unemployment Insurance (UI) were randomly assigned to one of three groups: the Job Search Incentive Experiment group (JSIE), the Hiring Incentive Experiment group (HIE), or the control group. The JSIE group was eligible for a \$500 bonus if they found a job of at least 30 hours per week within 11 weeks of the start of their unemployment period and held the job for 4 months. The Hiring Incentive Experiment group had the same eligibility requirements, but the \$500 bonus was given to the hiring company instead. A detailed description of the experiment can be found in [31].

It is worth noting that to be part of one of the three aforementioned groups, claimants must be between 20 and 55 years old and have a valid unemployment insurance claim. The duration of unemployment T was recorded as the number of weeks in which participants received unemployment benefits, resulting in a discrete data set. However, the true unemployment duration is continuous but only observed by intervals. This creates a partial identification issue, not within the scope of this paper. The data is treated as continuous. It is important to note that unemployment insurance is granted for 26 weeks, so observations can only be made up until the end of the UI claim and are therefore right censored at 26 weeks. Approximately 40% of the observations are censored.

To clarify the connection between our model and the experiment, we examine the impact of the cash bonus in the JSIE (or HIE) experiment. Although group assignment is randomly determined, participant agreement to participate in the experiment is not independent. This is because, to participate in the program, job seekers must also read the experiment description and sign an agreement form at the start of their follow-up, and this can be influenced by the participant's motivation to find work or by other personal attributes, such as specific skills. A framework with two treatments, as indicated by the following variable $Z = (Z_1, Z_2)^\top$, will be considered:

$$Z_1 = \begin{cases} 1 & \text{if the individual is in the JSIE group and agrees to participate} \\ 0 & \text{otherwise} \end{cases}$$

$$Z_2 = \begin{cases} 1 & \text{if the individual is in the HIE group and agrees to participate} \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, $Z = (0, 0)^\top$ corresponds to the control group, and $Z = (1, 0)^\top$, $Z = (0, 1)^\top$, correspond to the JSIE and HIE group, respectively. To address selection bias we utilize the group assignment variable W as instrumental variable:

$$W = \begin{cases} 0 & \text{if the individual is assigned to the control group} \\ 1 & \text{if the individual is assigned to the JSIE group} \\ 2 & \text{if the individual is assigned to the HIE group.} \end{cases}$$

The validity of W as a proper instrumental variable follows from the fact that it determines the group assignment, directly affecting the treatment variables Z_1 and Z_2 (relevance condition). The random assignment of W ensures that it is independent of unobserved factors (exogeneity condition) influencing unemployment duration. Finally, the fact that one is eligible for treatment ($W \in \{1, 2\}$) is unlikely to affect unemployment duration by any other pathway than the treatment itself (exclusion condition).

Note that, the rank invariance assumption in the Illinois Unemployment Incentive Experiment is plausible because the bonuses offered in the JSIE and HIE groups are likely to reduce the duration of unemployment by incentivizing faster job searches. However, these incentives do not alter the intrinsic characteristics or relative employability of participants. For example, individuals with strong work histories and job search skills are likely to find employment more quickly regardless of the bonus. Thus, while the bonuses may shorten unemployment duration for all participants, they do not fundamentally change the relative order of individuals' abilities to secure a job.

We consider for this analysis a subset of 1,543 non-white women out of the 12,101 available individuals and we report in Table 3 the sample sizes for each (W, Z) combination. JSIE and HIE experiments have refusal rates of 20% and 35%, indicating significant selection biases.

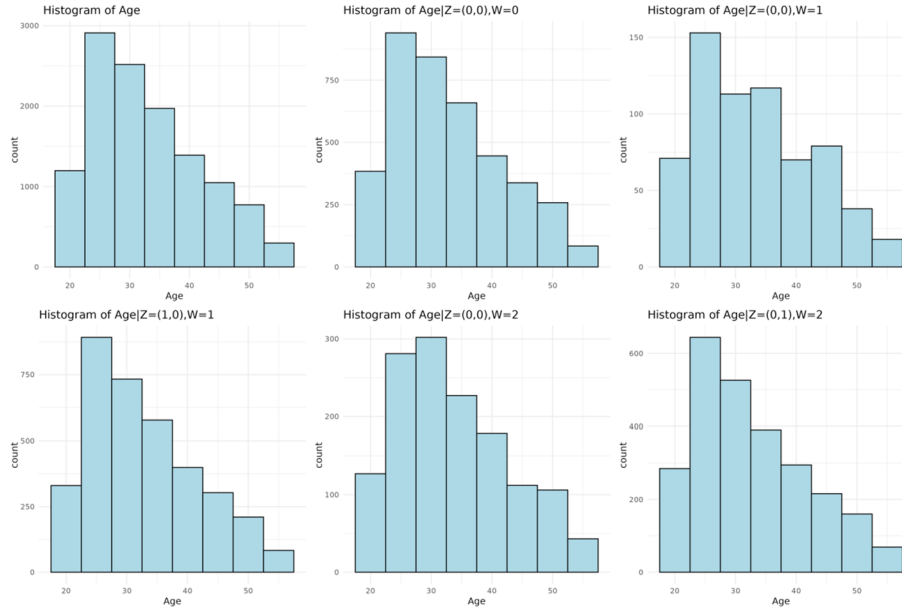


FIG 1. *Distribution of the Age variable. The histograms display the overall distribution of Age (top left) and its distribution conditional on different levels of the categorical variables Z and W . Specifically, the histograms depict the following: top row (from left to right) - overall distribution of Age, Age conditional on $Z = (0, 0)^T$ and $W = 0$, and Age conditional on $Z = (0, 0)^T$ and $W = 1$; bottom row (from left to right) - Age conditional on $Z = (1, 0)^T$ and $W = 1$, Age conditional on $Z = (0, 0)^T$ and $W = 2$, and Age conditional on $Z = (0, 1)^T$ and $W = 2$. All histograms use a consistent bin width to ensure comparability.*

TABLE 3
Number of observations for each combination of levels of the variable W and Z .

	$W = 0$	$W = 1$	$W = 2$
$Z = (0, 0)^T$	527	104	176
$Z = (1, 0)^T$	0	408	0
$Z = (0, 1)^T$	0	0	328

Furthermore, we considered an exogenous variable X representing the age of the subject, standardized to have a mean of zero and a variance of one. This standardization ensures numerical stability in the estimation procedure. The distribution of the Age variable is shown in Figure 1.

We use the proposed method to estimate the vector of coefficients. Similar to Section 6, we estimate the conditional distribution function using the Epanechnikov kernel, and the same selection method for the bandwidth using data-driven direct plug-in estimates of the Integrated Mean Squared Error-optimal bandwidth. Since the observations are censored at 26 weeks and the support of the unemployment duration T given Z, X, W should be the whole real line (some people stay unemployed forever), we chose $\bar{T} = 26$. In the scenario we are examining, where the closed-form solution of the system described in (15) is

applicable, we chose $\bar{U}$ to be the maximum value for which the closed-formula yields finite values for all $z \in \mathcal{Z}$, $x \in \mathcal{X}$, and $u \in [0, \bar{U}]$. That is, we set $\bar{U} = 0.6$. Table 4 displays the estimation results of the proposed estimator compared to the partial likelihood estimator, where the standard deviation is computed using 5000 bootstrap resamples with replications, and the confidence intervals are constructed using the estimated standard deviation and a normal approximation.

TABLE 4

Estimation results for the proposed estimator compared to the partial likelihood estimator. For each component (Comp) of $\beta = (\beta_x, \beta_{z,1}, \beta_{z,2})^\top$, we give the estimated value (Est), the standard deviation (Sd) and the confidence intervals at the 95% level (CI 0.025 and CI 0.975).

	proposed estimator				partial likelihood estimator			
	Est	Sd	CI 0.025	CI 0.975	Est	Sd	CI 0.025	CI 0.975
β_x	0.019	0.083	-0.143	0.181	-0.136	0.040	-0.215	-0.058
$\beta_{z,1}$	0.679	0.210	0.266	1.091	0.079	0.090	-0.098	0.256
$\beta_{z,2}$	0.579	0.201	0.184	0.973	0.075	0.097	-0.115	0.264

The results in Table 4 show a remarkable difference between the proposed estimator and the standard partial likelihood estimator. Specifically, the estimated values for $\beta_{z,1}$ and $\beta_{z,2}$, using the proposed estimator, are both positive and significant at the 95% confidence level. This provides empirical evidence that both treatments increase the hazard rate of finding a job, that is they reduce unemployment duration. Conversely, the values for these coefficients estimated using the standard method are close to zero and not significant at the same confidence level, leading to the conclusion that there is no evidence of any effects of the treatments on the duration.

The analysis using the partial likelihood estimator indicates that age has a significant effect on reducing the unemployment spell for non-white women, as evidenced by the negative and statistically significant estimate for β_x . This implies that being younger is beneficial in this context. However, when employing the proposed estimator, the impact of age on the unemployment duration appears to be negligible. The estimated value for β_x in this case is positive but not statistically significant, suggesting that age may not play a significant role in influencing the length of unemployment spells according to this estimation method.

In conclusion, the proposed estimator and the partial likelihood estimator yield different results leading to notably different cause-effect interpretations. It could be argued that our estimator provides a more reliable estimation of the regression coefficients compared to the partial likelihood estimator, especially when assessing the influence of the treatments on the duration outcome variable.

Appendix

The following Appendix includes the proofs of Theorems 3.1, 5.1 and 5.2, which are provided in Appendix B, D and E, respectively. The proofs make use of

the technicalities and the notations discussed in Appendix A, the results on $\hat{F}(t, z|x, w)$ presented in Appendix C and the auxiliary lemmas reported in Appendix F. Appendix G presents an interpretation of the identification assumptions discussed in Section 3, in the context of one-sided non-compliance. Appendix H contains additional simulations.

A. Technicalities and notation

A.1. Technicalities

The three steps of the estimation procedure discussed in the paper aim to clarify the generality of our method. However, in order to obtain asymptotic results on the estimator $\hat{\beta}$, it is necessary to properly formalize the fact that the estimator is based on “proxy” observations. For that, we can argue as follows. Define the space $\mathcal{F}_V^{\bar{U}, \bar{T}} = \{\phi : \mathcal{V} \times [0, \bar{U}] \rightarrow [0, \bar{T}] \mid \|\phi\|_\infty < \infty\}$, where $\|\phi\|_\infty = \sup_{v \in \mathcal{V}, u \in [0, \bar{U}]} |\phi(v, u)|$. Then, for any function $\phi \in \mathcal{F}_V^{\bar{U}, \bar{T}}$, and any continuous function $g : \mathcal{V} \mapsto \mathbb{R}^l$ for some integer $l \geq 1$, consider the quantities:

$$\begin{aligned} Q^\phi(t) &= P(\phi(V, \tilde{U}) \geq t, \Delta = 1); \\ E^\phi(g(v), t) &= E[g(V)I(\phi(V, \tilde{U}) \geq t)]; \\ E_1^\phi(g(v), t) &= E[g(V)I(\phi(V, \tilde{U}) \geq t, \Delta = 1)]. \end{aligned}$$

Consider the operator M which associates to a function $\phi \in \mathcal{F}_V^{\bar{U}, \bar{T}}$ a map from $\mathcal{B}$ to $\mathbb{R}^{d_V}$, denoted by $M(\cdot, \phi)$ and defined as

$$M(\beta, \phi) = E_1^\phi(v, 0) + \int_0^{\bar{T}} \frac{E^\phi(v \exp(v^\top \beta), s)}{E^\phi(\exp(v^\top \beta), s)} dQ^\phi(s). \quad (16)$$

The identification result of Theorem 3.2 can be equivalently stated by saying that β_0 is the unique point such that

$$\|M(\beta_0, \varphi)\| = 0. \quad (17)$$

This is because $\|M(\beta, \varphi)\| = 0$ is the first order condition of the maximization problem in β of the expected value of the PH partial likelihood based on the variables $\varphi(Z, \tilde{U}), \Delta, Z$ with respect to $(\tilde{U}, \Delta, Z)$. For any function $\phi \in \mathcal{F}_V^{\bar{U}, \bar{T}}$, and any continuous function $v \mapsto g(v)$, we define the sample analogue of the quantities $Q^\phi(t), E^\phi(g(v), t)$ and $E_1^\phi(g(v), t)$, respectively, as

$$\begin{aligned} \hat{Q}^\phi(t) &= n^{-1} \sum_{i=1}^n I(\phi(V_i, \tilde{U}_i) \geq t, \Delta_i = 1); \\ \hat{E}^\phi(g(v), t) &= n^{-1} \sum_{i=1}^n g(V_i) I(\phi(V_i, \tilde{U}_i) \geq t); \end{aligned}$$

$$\hat{E}_1^\phi(g(v), t) = n^{-1} \sum_{i=1}^n g(V_i) I(\phi(V_i, \tilde{U}_i) \geq t, \Delta_i = 1).$$

Then, we consider the sample analog of the operator M , namely M_n , defined as

$$M_n(\beta, \phi) = \hat{E}_1^\phi(v, 0) + \int_0^{\bar{T}} \frac{\hat{E}^\phi(v \exp(v^\top \beta), s)}{\hat{E}^\phi(\exp(v^\top \beta), s)} d\hat{Q}^\phi(s), \quad (18)$$

Then, the estimator $\hat{\beta}$ can be rewritten in a convenient mathematical reformulation, which underlines its dependency on the estimator of $\hat{\varphi}$:

$$\hat{\beta} \in \arg \min_{\beta \in \mathcal{B}} \|M_n(\beta, \hat{\varphi})\|. \quad (19)$$

This ensures to obtain the smallest number of censored proxy values, which improves the efficiency of the third step procedure.

We now fix an arbitrarily small value $\tau > 0$, and we suppose U_g^c has a uniform distribution on the interval $[\bar{U} - \tau, \bar{U}]$. Since $\tilde{U} = \min(U_g, U_g^c)$ and U_g has a uniform distribution on $[0, 1]$, we have:

$$P(\tilde{U} \leq u) = \begin{cases} 0, & \text{if } u < 0 \\ u, & \text{if } 0 \leq u < \bar{U} - \tau \\ (1 - \frac{\bar{U}}{\tau}) + u(\frac{1}{\tau} + \frac{\bar{U}}{\tau}) - \frac{u^2}{\tau}, & \text{if } \bar{U} - \tau \leq u < \bar{U} \\ 1, & \text{if } u \geq \bar{U} \end{cases};$$

$$P(\tilde{U} \leq u, \Delta = 1) = \begin{cases} 0, & \text{if } u < 0 \\ u, & \text{if } 0 \leq u < \bar{U} - \tau \\ -\frac{u^2}{2\tau} + \frac{u\bar{U}}{\tau} - \frac{(\bar{U}-\tau)^2}{2\tau}, & \text{if } \bar{U} - \tau \leq u < \bar{U} \\ \frac{\bar{U}^2 - (\bar{U}-\tau)^2}{2\tau}, & \text{if } u \geq \bar{U} \end{cases}.$$

Note also that the choice of the distribution for U_g^c has an impact on the distribution of $\tilde{U}$ and therefore it affects the distribution of the proxies $(\hat{\varphi}(Z, X, \tilde{U}), Z, X, \Delta)$. While the estimator $\hat{\beta}$ is consistent with any choice of the distribution of U_g^c , depending on the resulting distribution of the proxies the asymptotic variance of $\sqrt{n}(\hat{\beta} - \beta_0)$ can vary. We conjecture that the minimum of the variance is obtained by choosing a distribution for U_g^c maximizing $P(\Delta = 1)$, that is when the distribution of U_g^c is as close as possible to the probability mass point concentrated in $\bar{U}$, but it is outside the scope of this paper to prove such a result.

A.2. Notations

In the sequel, we will use the following quantities. We denote by $\Gamma_1(\beta, \phi)$ the ordinary derivative of $M(\beta, \phi)$ with respect to β , so $\Gamma_1(\beta, \phi) = \frac{\partial}{\partial \beta} M(\beta, \phi)$, that is

$$\Gamma_1(\beta, \phi) = \int_0^{\bar{T}} \left[\frac{E^\phi(vv^\top \exp(v^\top \beta), s)}{E^\phi(\exp(v^\top \beta), s)} - \left(\frac{E^\phi(v \exp(v^\top \beta), s)}{E^\phi(\exp(v^\top \beta), s)} \right)^2 \right] dQ^\phi(s).$$

We denote by $\Gamma_2(\beta, \phi)[\bar{\phi} - \phi]$ the pathwise derivative of $M(\beta, \phi)$ at ϕ in direction $[\bar{\phi} - \phi]$ when it exists, i.e.

$$\Gamma_2(\beta, \phi)[\bar{\phi} - \phi] = \lim_{\varepsilon \rightarrow 0} \frac{M(\beta, \phi + \varepsilon[\bar{\phi} - \phi]) - M(\beta, \phi)}{\varepsilon}. \quad (20)$$

Define also the function $\varrho(v, t)$ as $\varrho(v, t) = \varphi(v, \cdot)^{-1}(t)$, where $t \in [0, \bar{T}]$ and $v \in \mathcal{V}$. We denote by 1_L the vector in $\mathbb{R}^L$ with each component equal to 1. Lastly, to simplify the notation, for a function $f(v, s)$ we denote by $f'(v, s)$ and by $f''(v, s)$ the first and second derivative of $f(v, s)$ with respect to the last argument s , respectively.

B. Proof of Theorem 3.1

First, remark that, by the “Kaplan-Meier type” arguments mentioned in Section 4, the left-hand-side of equation (5) is identified for all $(\theta_\ell)_{\ell=1}^L \in \prod_{\ell=1}^L [0, \bar{t} \wedge c_{z_\ell, x}]$, where $c_{z_\ell, x} = \min_{k=1, \dots, K} c_{z_\ell, x, w_k}$.

Next, notice that

$$\bar{u}^x = \sup \left\{ u \in [0, 1] : \text{the system in (5) has a solution in } \prod_{\ell=1}^L [0, \bar{t} \wedge c_{z_\ell, x}] \right\}. \quad (21)$$

Indeed, for $u \leq \bar{u}^x$, since $\bar{u}^x = \inf_{z \in \mathcal{Z}, w \in \mathcal{W}} \varphi(z, x, \cdot)^{-1}(\bar{t} \wedge c_{z, x, w})$, we have $\varphi^x(u) \in \prod_{\ell=1}^L [0, \bar{t} \wedge c_{z_\ell, x}]$, so that (5) has a solution in $\prod_{\ell=1}^L [0, \bar{t} \wedge c_{z_\ell, x}]$ which is $\varphi^x(u)$. Next, for $u > \bar{u}^x$, $\varphi^x(u) \notin \prod_{\ell=1}^L [0, \bar{t} \wedge c_{z_\ell, x}]$, so that (5) cannot have a solution in $\prod_{\ell=1}^L [0, \bar{t} \wedge c_{z_\ell, x}]$ (otherwise it would have two solutions in $\mathbb{R}^L$ which contradicts the assumption of the theorem). Equation (21) shows that $\bar{u}^x$ is identified since its right-hand side is identified. Then, for all $u \leq \bar{u}^x$, $\varphi^x(u)$ is identified as the unique solution to (21) in $(\theta_\ell)_{\ell=1}^L \in \prod_{\ell=1}^L [0, \bar{t} \wedge c_{z_\ell, x}]$. $\square$

C. Results on $\hat{F}(t, z|x, w)$

In this section, we present results concerning the consistency and the rate of convergence of $\hat{F}(t, z|x, w)$ to $F(t, z|x, w)$. In addition, it states the consistency of $\hat{f}(t, z|x, w)$ and of its derivatives with respect to t and x up to order $2 + \alpha$. These results will be used to ensure the regularity and the rate of convergence of the estimator $\hat{\varphi}$, which is crucial for the rate of convergence of $\hat{\beta}$. Lastly, we show that the estimators $\hat{F}(t, z|x, w)$ and $\hat{f}(t, z|x, w)$ admit an iid representation.

Theorem C.1. *Under Assumptions 3 and 4, the following results hold:*

- (i) $\sup_{t \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} |\hat{F}(t, z|x, w) - F(t, z|x, w)| = O_p((\log n / (nh))^{1/2} + h^\nu + \epsilon^\pi).$
- (ii) $\sup_{t \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} |\hat{f}(t, z|x, w) - f(t, z|x, w)| = O_p((\log n / (nh\epsilon))^{1/2} + h^\nu + \epsilon^\pi).$

- (iii) $\sup_{t \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} \left| \frac{\partial}{\partial t} \hat{f}(t, z|x, w) - \frac{\partial}{\partial t} f(t, z|x, w) \right| = o_p(1).$
- (iv) $\sup_{t \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} \left| \frac{\partial^2}{\partial t^2} \hat{f}(t, z|x, w) - \frac{\partial^2}{\partial t^2} f(t, z|x, w) \right| = o_p(1).$
- (v) $\sup_{t_1, t_2 \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} \frac{\left| \frac{\partial^2}{\partial t^2} \hat{f}(t_1, z|x, w) - \frac{\partial^2}{\partial t^2} f(t_2, z|x, w) \right|}{|t_1 - t_2|^\alpha} = o_p(1).$
- (vi) $\sup_{t \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} \left| \frac{\partial}{\partial x} \hat{f}(t, z|x, w) - \frac{\partial}{\partial x} f(t, z|x, w) \right| = o_p(1).$
- (vii) $\sup_{t \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} \left| \frac{\partial^2}{\partial x^2} \hat{f}(t, z|x, w) - \frac{\partial^2}{\partial x^2} f(t, z|x, w) \right| = o_p(1).$
- (viii) $\sup_{t \in [0, \bar{T}], z \in \mathcal{Z}, x_1, x_2 \in \mathcal{X}, w \in \mathcal{W}} \frac{\left| \frac{\partial^2}{\partial x^2} \hat{f}(t, z|x_1, w) - \frac{\partial^2}{\partial x^2} f(t, z|x_2, w) \right|}{|x_1 - x_2|^\alpha} = o_p(1).$
- (ix) For $(t, z, x, w) \in [0, \bar{T}] \times \mathcal{Z} \times \mathcal{X} \times \mathcal{W}$, the quantity $\hat{F}(t, z|x, w) - F(t, z|x, w)$ can be expressed as

$$\begin{aligned} & \hat{F}(t, z|x, w) - F(t, z|x, w) \\ &= (nh)^{-1} \sum_{i=1}^n K \left(\frac{x - X_i}{h} \right) \eta^F(Y_i, \delta_i, Z_i, W_i, t, z, x, w) \\ & \quad + R_n(t, z, x, w); \end{aligned}$$

where

$$\begin{aligned} & \eta^F(Y_i, \delta_i, Z_i, W_i, t, z, x, w) \\ &= F(t|z, x, w) I(W_i = w) \frac{I(Z_i = z) - p_{z, x, w}}{f_{X, W}(x, w)} \\ & \quad + p_{z, x, w} \xi^F(Y_i, \delta_i, Z_i, W_i, t, z, x, w); \\ & \xi^F(Y_i, \delta_i, Z_i, W_i, t, z, x, w) \\ &= (1 - F(t|z, x, w)) \left[\int_0^{\min(Y_i, t)} \frac{-dF_{Y, 1|Z, X, W}(y|z, x, w)}{(1 - F_{Y|Z, X, W}(y|z, x, w))^2} \right. \\ & \quad \left. + \frac{\delta_i I(Y_i \leq t)}{1 - F_{Y|Z, X, W}(Y_i|z, x, w)} \right]; \end{aligned}$$

and

$$\sup_{t \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} |R_n(t, z, x, w)| = O_p \left((\log n / (nh))^{3/4} + h^\nu + \epsilon^\pi \right). \quad (22)$$

- (x) For $(t, z, x, w) \in [0, \bar{T}] \times \mathcal{Z} \times \mathcal{X} \times \mathcal{W}$, the quantity $\hat{f}(t, z|x, w) - f(t, z|x, w)$ can be expressed as

$$\begin{aligned} & \hat{f}(t, z|x, w) - f(t, z|x, w) \\ &= (nh)^{-1} \sum_{i=1}^n K \left(\frac{x - X_i}{h} \right) I(W_i = w) \frac{I(Z_i = z) - p_{z, x, w}}{f_{X, W}(x, w)} f(t|z, x, w) \end{aligned}$$

$$\begin{aligned}
& + (nh\epsilon)^{-1} \sum_{i=1}^n p_{z,x,w} K\left(\frac{x - X_i}{h}\right) \\
& \quad \times \int \tilde{K}(u) \xi^F(Y_i, \delta_i, Z_i, W_i, t - u\epsilon, z, x, w) du \\
& + r_n(t, z, x, w)
\end{aligned}$$

and

$$\sup_{t \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} |r_n(t, z, x, w)| = O_p((\log n / (nh\epsilon))^{3/4} + h^\nu + \epsilon^\pi).$$

Proof. The result follows from Lemma F.1 and Lemma F.2. More precisely, for (i) we can write

$$\begin{aligned}
& |\hat{F}(t, z|x, w) - F(t, z|x, w)| \\
& \leq |\hat{F}(t|z, x, w) - F(t|z, x, w)| |\hat{p}_{z,x,w} - p_{z,x,w}| \\
& \quad + F(t|z, x, w) |\hat{p}_{z,x,w} - p_{z,x,w}| + p_{z,x,w} |\hat{F}(t|z, x, w) - F(t|z, x, w)| \\
& = O_p((\log n / (nh))^{1/2} + h^\nu + \epsilon^\pi) O_p((\log n / (nh))^{1/2} + h^\nu) \\
& \quad + O_p((\log n / (nh))^{1/2} + h^\nu) + O_p((\log n / (nh))^{1/2} + h^\nu + \epsilon^\pi) \\
& = O_p((\log n / (nh))^{1/2} + h^\nu + \epsilon^\pi). \tag{23}
\end{aligned}$$

The same argument can be used for (ii) – (viii) using Lemmas F.1, F.2 and F.3. The result in (ix) and (x) are corollaries of Lemma F.4. In detail, observing that

$$\sup_{x \in \mathcal{X}, w \in \mathcal{W}} |(nh)^{-1} \sum_{i=1}^n I(W_i = w) K\left(\frac{x - X_i}{h}\right) - f_{X,W}(x, w)| \tag{24}$$

$$= O_p((\log n / (nh))^{1/2} + h^\nu), \tag{25}$$

we can write

$$\begin{aligned}
\hat{p}_{z,x,w} - p_{z,x,w} &= (nh)^{-1} \sum_{i=1}^n \frac{I(W_i = w) K\left(\frac{x - X_i}{h}\right) [I(Z_i = z) - p_{z,x,w}]}{(nh)^{-1} \sum_{i=1}^n I(W_i = w) K\left(\frac{x - X_i}{h}\right)} \\
&= (nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) I(W_i = w) \frac{I(Z_i = z) - p_{z,x,w}}{f_{X,W}(x, w)} \\
&\quad + O(h^\nu + \log n / (nh)).
\end{aligned}$$

Therefore, with a similar decomposition as in (23), we have

$$\begin{aligned}
& \hat{F}(t, z|x, w) - F(t, z|x, w) \\
& = O_p((\log n / (nh))^{1/2} + h^\nu + \epsilon^\pi) O_p((\log n / (nh))^{1/2} + h^\nu) \\
& \quad + F(t|z, x, w) (\hat{p}_{z,x,w} - p_{z,x,w}) + p_{z,x,w} (\hat{F}(t|z, x, w) - F(t|z, x, w)).
\end{aligned}$$

From this and Lemma F.4 the result in (ix) follows. With a similar argument, we also obtain (x). $\square$

D. Proof of Theorem 5.1

Proof. The proof is based on Lemma F.5 and on semiparametric Z-estimator-related theory. In particular, we refer to Theorem 1 in [10], and in the following, we show the fulfillment of conditions (1.1)-(1.5) of that theorem.

- (1.1) The result follows from the definition of $\hat{\beta}$.
- (1.2) The result is a consequence of the identification of β_0 , given in Theorem 3.2 and the continuity of the map $\beta \mapsto M(\beta, \varphi)$.
- (1.3) Note that $P(\tilde{U} \leq u)$ and $P(\tilde{U} \leq u, \Delta = 1)$ are continuous in u . In addition, the functions $g : \mathcal{V} \times \mathcal{B} \rightarrow \mathbb{R}$ and $g_1 : \mathcal{V} \times \mathcal{B} \rightarrow \mathbb{R}^{d_V}$ defined as $g(v, \beta) = \exp(v^\top \beta)$ and $g_1(v, \beta) = v \exp(v^\top \beta)$ are Lipschitz in β uniformly in v , since $\mathcal{V}$ and $\mathcal{B}$ are compact. Thus, the result easily follows.
- (1.4) The result follows from Lemma F.5.
- (1.5) The result is a consequence of the proof of Theorem 3 in [10], which is shown to be valid in point (2.5) under Theorem 5.2 below. $\square$

E. Proof of Theorem 5.2

Proof. Fix a value $\alpha \in (0, 1)$. We are interested in showing that $\hat{f}(t, z|x, w)$ is Lipschitz continuous up to order $2 + \alpha$. Any value of $\alpha \in (0, 1)$ is suitable, taking into account that a higher value of α restricts the conditions on the bandwidths h and ϵ . Now, for a function $f : \mathcal{Z} \times \mathcal{X} \times [0, \bar{U}] \rightarrow \mathbb{R}$ define the norm $\|\cdot\|_\Phi$ as

$$\begin{aligned} \|f\|_\Phi = \sum_{z \in \mathcal{Z}} \Big\{ & \sup_{u \in [0, \bar{U}], x \in \mathcal{X}} |f(z, x, u)| \\ & + \sup_{u \in [0, \bar{U}], x \in \mathcal{X}} \left| \frac{\partial}{\partial u} f(z, x, u) \right| \\ & + \sup_{u \in [0, \bar{U}], x \in \mathcal{X}} \left| \frac{\partial^2}{\partial u^2} f(z, x, u) \right| \\ & + \sup_{u \in [0, \bar{U}], x \in \mathcal{X}} \left| \frac{\partial}{\partial x} f(z, x, u) \right| \\ & + \sup_{u \in [0, \bar{U}], x \in \mathcal{X}} \left| \frac{\partial^2}{\partial x^2} f(z, x, u) \right| \\ & + \sup_{u_1, u_2 \in [0, \bar{U}], x \in \mathcal{X}} \frac{\left| \frac{\partial^2}{\partial u^2} f(z, x, u_1) - \frac{\partial^2}{\partial u^2} f(z, x, u_2) \right|}{|u_1 - u_2|^\alpha} \\ & + \sup_{x_1, x_2 \in \mathcal{X}, u \in [0, \bar{U}]} \frac{\left| \frac{\partial^2}{\partial x^2} f(z, x_1, u) - \frac{\partial^2}{\partial x^2} f(z, x_2, u) \right|}{|x_1 - x_2|^\alpha} \Big\}. \end{aligned}$$

With the abuse of notation of denoting the restriction of $\varphi(v, u)$ on $\mathcal{V} \times [0, \bar{U}]$ again as φ , under our regularity assumptions, for any $\alpha \in (0, 1)$, there exists a constant m_0 such that $\|\varphi\|_\Phi < m_0$. This result follows from the implicit

function theorem and our regularity assumptions 5 and 6. Therefore, $\varphi \in \Phi$, where $\Phi = \{f : \mathcal{Z} \times \mathcal{X} \times [0, \bar{U}] \rightarrow \mathbb{R} \mid f(z, x, \cdot) \text{ and } f(z, \cdot, u) \text{ are twice differentiable and such that } \|f\|_{\Phi} \leq m_0\}$. The norm $\|\cdot\|_{\Phi}$ is defined specifically to establish the set Φ . It ensures that the necessary level of regularity for the functions involved in the theorem is met. This is crucial for achieving the Donsker property results of the given class of functions. The set Φ is endowed with the norm $\|\phi\|_1 = \sup_{v \in \mathcal{V}, u \in [0, \bar{U}]} |\phi(v, u)| + \sup_{v \in \mathcal{V}, u \in [0, \bar{U}]} |\phi'(v, u)|$ for any $\phi \in \Phi$. The operator M and its sample analog M_n are considered as operators on the space $(\Phi, \|\cdot\|_1)$.

The proof utilizes lemmas provided in the Online Appendix and Z-estimator-related theory. In particular, we refer to Theorem 2 of [10] and, in the following, we show the fulfillment of conditions (2.1)-(2.6) of that theorem.

- (2.1) The result follows by definition of $\hat{\beta}$.
- (2.2) It is easy to see that $\Gamma_1(\beta, \varphi)$ is continuous in β and that $\Gamma_1(\beta_0, \varphi)$ is negative definite, and so it has full rank.
- (2.3) First, note that, thanks to our regularity conditions, the functions $\varphi'(v, u)$, $\varphi''(v, u)$, $\varrho(v, t)$ and $\varrho'(v, t)$ exist. Assumption 1 (ii) also implies that, on $\mathcal{V} \times [0, \bar{U}]$, the function $\varphi'(v, u)$ admits a minimum strictly greater than zero. This, together with the definition of Φ which ensures that for any $\phi \in \Phi$, the functions $\phi(v, u)$ and $\phi'(v, u)$ are uniformly bounded by the constant m_0 , we obtain that there exists $\epsilon = \epsilon(m_0) > 0$ such that for any $\phi \in \Phi$ the function $\varphi + t\phi$ is invertible for every $|t| < \epsilon$. Denoting the measure induced on $\mathcal{V}$ by the probability of the variable V by $\mu(v)$, we can use Lemma F.8 and the boundedness of the involved quantities to justify the following equalities:

$$\begin{aligned}
 & \lim_{\varepsilon \rightarrow 0} \frac{E^{\varphi + \varepsilon\phi}(g(v), t) - E^{\varphi}(g(v), t)}{\varepsilon} \\
 &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{\mathcal{V}} g(v) \left[P((\varphi + \varepsilon\phi)(V, \tilde{U}) \geq t | V = v) \right. \\
 & \quad \left. - P(\varphi(V, \tilde{U}) \geq t | V = v) \right] d\mu(v) \\
 &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{\mathcal{V}} g(v) \left[P((\varphi + \varepsilon\phi)(v, \tilde{U}) \geq t | V = v) \right. \\
 & \quad \left. - P(\varphi(v, \tilde{U}) \geq t | V = v) \right] d\mu(v) \\
 &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{\mathcal{V}} g(v) \left[P((\varphi + \varepsilon\phi)(v, \tilde{U}) \geq t) \right. \\
 & \quad \left. - P(\varphi(v, \tilde{U}) \geq t) \right] d\mu(v) \\
 &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \int_{\mathcal{V}} g(v) \left[P(\tilde{U} \geq (\varphi + \varepsilon\phi)(v, \cdot)^{-1}(t)) \right. \\
 & \quad \left. - P(\tilde{U} \geq \varphi(v, \cdot)^{-1}(t)) \right] d\mu(v)
 \end{aligned}$$

$$= - \int_{\mathcal{V}} g(v) \left\{ I(\varrho(v, t) < \bar{U} - \delta) + I(\bar{U} - \delta \leq \varrho(v, t) \leq \bar{U}) \right. \\ \left. \times \frac{1 + \bar{U} - 2\varrho(v, t)}{\delta} \right\} \times \frac{\phi(v, \varrho(v, t))}{\varphi'(v, \varrho(v, t))} d\mu(v),$$

where we used the expression of the distribution of $\tilde{U}$. With a similar argument, we also obtain

$$\lim_{\varepsilon \rightarrow 0} \frac{E_1^{\varphi+\varepsilon\phi}(g(v), t) - E_1^{\varphi}(g(v), t)}{\varepsilon} \\ = - \int_{\mathcal{V}} g(v) \left\{ I(\varrho(v, t) < \bar{U} - \delta) + I(\bar{U} - \delta \leq \varrho(v, t) \leq \bar{U}) \frac{\bar{U} - \varrho(v, t)}{\delta} \right\} \\ \times \frac{\phi(v, \varrho(v, t))}{\varphi'(v, \varrho(v, t))} d\mu(v).$$

The quantity $\frac{dQ^{\phi}(t)}{dt}$ can be written as

$$\frac{dQ^{\phi}(t)}{dt} = \frac{\partial}{\partial t} \int_{\mathcal{V}} P(\tilde{U} \geq \phi(v, \cdot)^{-1}(t), \Delta = 1) d\mu(v) \\ = \int_{\mathcal{V}} \left\{ I(\phi(v, \cdot)^{-1}(t) < \bar{U} - \delta) \right. \\ \left. + I(\bar{U} - \delta \leq \phi(v, \cdot)^{-1}(t) \leq \bar{U}) \frac{\bar{U} - \phi(v, \cdot)^{-1}(t)}{\delta} \right\} \\ \times \frac{\partial}{\partial t} \phi(v, \cdot)^{-1}(t) d\mu(v).$$

Therefore, Lemma F.9 and again Lemma F.8 yield

$$\lim_{\varepsilon \rightarrow 0} \frac{\frac{dQ^{\varphi+\varepsilon\phi}(t)}{dt} - \frac{dQ^{\phi}(t)}{dt}}{\varepsilon} \\ = \int_{\mathcal{V}} \left\{ I(\varrho(v, t) < \bar{U} - \tau) + I(\bar{U} - \tau \leq \varrho(v, t) \leq \bar{U}) \frac{\bar{U}}{\tau} \right\} \\ \times \frac{\phi(v, \varrho(v, t)) \varrho'(v, t) \varphi''(v, \varrho(v, t)) - \phi'(v, \varrho(v, t))}{(\varphi'(v, \varrho(v, t)))^2} d\mu(v) \\ + \int_{\mathcal{V}} \left\{ I(\bar{U} - \tau \leq \varrho(v, t) \leq \bar{U}) \frac{\varrho(v, t)'}{\tau} \right\} \frac{\phi(v, \varrho(v, t))}{\varphi'(v, \varrho(v, t))} d\mu(v).$$

Then, we can define Then, we can define

$$e^{\varphi}(g, t)[\phi] = - \int_{\mathcal{V}} g(v) \left\{ I(\varrho(v, t) < \bar{U} - \tau) + I(\bar{U} - \tau \leq \varrho(v, t) \leq \bar{U}) \right. \\ \left. \times \frac{1 + \bar{U} - 2\varrho(v, t)}{\tau} \right\} \frac{\phi(v, \varrho(v, t))}{\varphi'(v, \varrho(v, t))} d\mu(v); \\ e_1^{\varphi}(g, t)[\phi] = - \int_{\mathcal{V}} g(v) \left\{ I(\varrho(v, t) < \bar{U} - \tau) \right.$$

$$\begin{aligned}
& + I(\bar{U} - \tau \leq \varrho(v, t) \leq \bar{U}) \frac{\bar{U} - \varrho(v, t)}{\tau} \Big\} \\
& \times \frac{\phi(v, \varrho(v, t))}{\varphi'(v, \varrho(v, t))} d\mu(v); \\
q^\varphi(t)[\phi] &= \int_{\mathcal{V}} \left\{ I(\varrho(v, t) < \bar{U} - \tau) + I(\bar{U} - \tau \leq \varrho(v, t) \leq \bar{U}) \frac{\bar{U}}{\tau} \right\} \\
& \times \frac{\phi(v, \varrho(v, t)) \varrho'(v, t) \varphi''(v, \varrho(v, t)) - \phi'(v, \varrho(v, t))}{(\varphi'(v, \varrho(v, t)))^2} d\mu(v) \\
& + \int_{\mathcal{V}} \left\{ I(\bar{U} - \tau \leq \varrho(v, t) \leq \bar{U}) \frac{\varrho(v, t)'}{\tau} \right\} \frac{\phi(v, \varrho(v, t))}{\varphi'(v, \varrho(v, t))} d\mu(v); \\
q^\varphi(t) &= \int_{\mathcal{V}} \left\{ I(\varrho(v, t) < \bar{U} - \tau) + I(\bar{U} - \tau \leq \varrho(v, t) \leq \bar{U}) \frac{\bar{U} - \varrho(v, t)}{\tau} \right\} \\
& \times \frac{\varrho(v, t)}{\frac{\partial}{\partial u} \varphi(v, \varrho(v, t))} d\mu(v).
\end{aligned}$$

After some computation, it can be shown that

$$\begin{aligned}
\Gamma_2(\beta, \varphi)[\phi] &= e_1(v, 0)[\phi] \\
& - \int_0^{\bar{T}} \frac{\{e^\varphi(v \exp(v^\top \beta), t)[\phi] q(t) + E^\varphi(v \exp(v^\top \beta), t) q^\varphi(t)[\phi]\}}{E^\varphi(\exp(v^\top \beta), t)^2} \quad (26) \\
& \times E^\varphi(\exp(v^\top \beta), t) dt
\end{aligned}$$

$$+ \int_0^{\bar{T}} \frac{E^\varphi(v \exp(v^\top \beta), t) q^\varphi(t) e^\varphi(\exp(v^\top \beta), t)[\phi]}{E^\varphi(\exp(v^\top \beta), t)^2} dt. \quad (27)$$

Therefore, by performing a routine Taylor expansion of $M(\beta, \phi)$ around φ , we can verify the desired condition due to the fact that the quantities $e^\varphi(g, t)[\phi]$, $e_1^\varphi(g, t)[\phi]$ and $q^\varphi(t)[\phi]$ are linear with respect to $(\phi(v, u), \phi'(v, u))$ and the fact that $\Gamma_2(\beta, \varphi)[\phi]$ is linear with respect to $(e^\varphi(g, t)[\phi], e_1^\varphi(g, t)[\phi], q^\varphi(t)[\phi])$. This shows that condition (2.3)(i) of [10] is satisfied. Condition (2.3)(ii) is instead implied by the differentiability in β of $\Gamma_2(\beta, \varphi)[\phi]$.

(2.4) The result follows from the Implicit Function Theorem. We now explain the argument in detail. Consider the map $\hat{g} : \mathbb{R} \times \mathbb{R} \times \mathbb{R}^L \rightarrow \mathbb{R}^K$, defined as

$$\hat{g} : (u, x, \theta) \mapsto (\hat{g}_1(u, x, \theta), \dots, \hat{g}_K(u, x, \theta))^\top, \quad (28)$$

where $\hat{g}_k(u, x, \theta) = \sum_{l=1}^L \hat{F}^x(\theta_l, z_l | w_k) - u$. By definition, $\hat{\varphi}$ satisfies $\hat{g}(u, x, \hat{\varphi}^x(u)) = 0$, for each $u \in [0, \bar{U}]$, $x \in \mathcal{X}$. Theorem C.1 implies $\sup_{u \in [0, \bar{U}], x \in \mathcal{X}} \|\Gamma(\hat{\varphi}^x, \hat{F}^x)(u) - \Gamma(\varphi^x, F^x)\| = o_p(1)$. Therefore, with probability tending to 1, $\Gamma(\hat{\varphi}^x, \hat{F}^x)(u)$ is invertible, because of Assumption 6.

Thus, the Implicit Function Theorem implies

$$\begin{aligned}\frac{\partial}{\partial u}\hat{\varphi}^x(u) &= -\Gamma(\hat{\varphi}^x, \hat{F}^x)(u)^{-1}\left(\frac{\partial \hat{g}_0}{\partial u}(u, x, \hat{\varphi}^x(u)), \dots, \frac{\partial \hat{g}_K}{\partial u}(u, x, \hat{\varphi}^x(u))\right)^\top \\ &= \Gamma(\hat{\varphi}^x, \hat{F}^x)(u)^{-1}1_K.\end{aligned}\quad (29)$$

Using the same argument as in Equation (41), we obtain

$$\begin{aligned}\frac{\partial}{\partial u}\hat{\varphi}^x - \frac{\partial}{\partial u}\varphi^x &= \left(\Gamma(\hat{\varphi}^x, \hat{F}^x)^{-1} - \Gamma(\varphi^x, F^x)^{-1}\right)1_K \\ &= \left(\Gamma(\hat{\varphi}^x, F^x)^{-1} - \Gamma(\varphi^x, F^x)^{-1} + O_p((\log n/(nh\epsilon))^{1/2} + h^\nu + \epsilon^\pi)\right)1_K\end{aligned}\quad (30)$$

Using the boundedness of the second derivative in t of $F(t, z|x, w)$ uniformly in $x \in \mathcal{X}, z \in \mathcal{Z}, w \in \mathcal{W}, t \in [0, \bar{T}]$, Lemma F.7, and the differentiability of the operator that associates to a matrix M its inverse M^{-1} , it holds $\sup_{u \in [0, \bar{U}], x \in \mathcal{X}} \|\Gamma(\hat{\varphi}^x, F^x)^{-1} - \Gamma(\varphi^x, F^x)^{-1}\| = o_p(n^{-1/4})$. This results, together with Lemma F.7 implies $\|\hat{\varphi} - \varphi\|_1 = o_p(n^{-1/4})$. The regularity of $\hat{F}(t, z|x, w)$ in both the arguments t, x implies, with a similar argument, that also $\hat{\varphi}$ has the same regularity in u, x and so $\hat{\varphi} \in \Phi$ with probability tending to 1.

(2.5) For $\gamma > 0$, define $\Phi_\gamma = \{\phi \in \Phi : \|\phi - \varphi\|_1 < \gamma\}$. We first show that the family

$$\mathcal{F} = \{(v, u) \mapsto I(\phi(v, u) \geq t) | \phi \in \Phi_\gamma, t \in [0, \bar{T}]\},$$

is Donsker. Using Theorem 2.5.6 of [25], it is sufficient to show that

$$\int_0^\infty \sqrt{\log N_{[]}(\bar{\epsilon}, \mathcal{F}, L_2(P))} d\bar{\epsilon} < \infty,$$

where $N_{[]}$ is the bracketing number, P is the probability measure corresponding to the joint distribution of $(V, \tilde{U})$, and $L_2(P)$ is the L_2 -norm. Fix $\bar{\epsilon} > 0$. In Corollary 2.7.2 of [25], it is stated that

$$m = N_{[]}(\bar{\epsilon}^2, \Phi_\gamma, L_2(P)) \leq \exp(K\bar{\epsilon}^{-4/(2+\alpha)}).$$

Let $\phi_1^L \leq \phi_1^T, \dots, \phi_m^L \leq \phi_m^T$ be the functions defining the m brackets for Φ_γ . Thus, for each fixed t and $\phi \in \Phi_\gamma$, we have

$$I(\phi_i^L(v, u) \geq t) \leq I(\phi(v, u) \geq t) \leq I(\phi_i^T(v, u) \geq t),$$

for some $i = 1, \dots, m$. Define $F_i^L(t) = P(\phi_i^L(V, \tilde{U}) \geq t)$, and let t_{ij}^L for $j = 1, \dots, O(\bar{\epsilon}^{-2})$, partition the line in segments having F_i^L -probability less than or equal to a fraction of $\bar{\epsilon}^2$. Similarly, define $F_i^T(t) = P(\phi_i^T(V, \tilde{U}) \geq$

t), and let t_{ij}^T for $j = 1, \dots, O(\bar{\epsilon}^{-2})$, partition the line in segments having F_i^T -probability less than or equal to a fraction of $\bar{\epsilon}^2$. Now define the following bracket for t :

$$t_{ij_1}^L \leq t \leq t_{ij_2}^T,$$

where $t_{ij_1}^L$ is the largest of the t_{ij}^L with the property of being less than or equal to t and $t_{ij_2}^T$ is the smallest of the t_{ij}^T with the property of being greater than or equal to t . We will now show that the brackets for $\mathcal{F}$ are given by

$$I(\phi_i^L(v, u) \geq t_{ij_2}^T) \leq I(\phi(v, u) \geq t) \leq I(\phi_i^T(v, u) \geq t_{ij_1}^L).$$

Note that, taking eventually a smaller $\gamma > 0$, for $\phi_1, \phi_2 \in \Phi_\gamma$, there exist $\phi_1(v, \cdot)^{-1}$ and $\phi_2(v, \cdot)^{-1}$, which we can denote by $\rho_1(v, t)$ and $\rho_2(v, t)$, respectively. In addition, by the regularity assumptions on φ , it is easy to show that there exists a constant Θ , such that, for any integer $k \geq 1$, we have

$$\|\rho_1 - \rho_2\|_{P,k} \leq \|\rho_1 - \rho_2\|_\infty \leq \Theta \|\phi_1 - \phi_2\|_1, \quad (31)$$

where $\|\cdot\|_{P,k}$ is the $L_k(P)$ norm. This also implies that there exists a constant K such that

$$\begin{aligned} & \|I(\phi_i^L(V, \tilde{U}) \geq t_{ij_2}^T) - I(\phi_i^T(V, \tilde{U}) \geq t_{ij_1}^L)\|_{P,2}^2 \\ &= P(\phi_i^T(V, \tilde{U}) \geq t_{ij_1}^L) - P(\phi_i^L(V, \tilde{U}) \geq t_{ij_2}^T) \\ &= F_i^T(t_{ij_1}^L) - F_i^L(t_{ij_2}^T) \\ &\leq F_i^T(t) - F_i^L(t) + K\epsilon^2. \end{aligned}$$

Now, denote by $f_{\tilde{U}}(u)$ the quantity $f_{\tilde{U}}(u) = dP(\tilde{U} \leq u)/du$, which exists a.e., and by $\rho_i^L(v, t)$ and $\rho_i^T(v, t)$ the functions $\phi_i^L(v, \cdot)^{-1}$ and $\phi_i^T(v, \cdot)^{-1}$, respectively. Then, using a Taylor expansion, we obtain

$$\begin{aligned} F_i^T(t) - F_i^L(t) &= \int_{\mathcal{V}} \left(P(\tilde{U} \geq \rho_i^T(v, t)) - P(\tilde{U} \geq \rho_i^L(v, t)) \right) d\mu(v) \\ &= \int_{\mathcal{V}} f_{\tilde{U}}(\xi(v, t)) (\rho_i^T(v, t) - \rho_i^L(v, t)) d\mu(v), \end{aligned} \quad (32)$$

where $\xi(v, t)$ is a point between $\rho_i^T(v, t)$ and $\rho_i^L(v, t)$. Therefore, for a suitable constant J bounding $f_{\tilde{U}}$, we obtain

$$\begin{aligned} F_i^T(t) - F_i^L(t) &\leq J \|\rho_i^T(v, t) - \rho_i^L(v, t)\|_{P,1} \\ &\leq J \|\rho_i^T(v, t) - \rho_i^L(v, t)\|_{P,2} \\ &\leq J\Theta\epsilon^2. \end{aligned}$$

Hence for the class $\mathcal{F}$ and for each $\bar{\epsilon} > 0$, we have at most

$O(\bar{\epsilon}^{-2} \exp(K\bar{\epsilon}^{-4/(2+\alpha)}))$ brackets in total. However, for $\bar{\epsilon} > 1$, one bracket suffices. So, it holds

$$\int_0^\infty \sqrt{\log N_{[]}(\bar{\epsilon}, \mathcal{F}, L_2(P))} d\bar{\epsilon} < \infty.$$

This shows that the class $\mathcal{F}$ is Donsker. Together with Theorem 2.10.6 in [25], this result also implies that the following families of functions are Donsker:

$$\begin{aligned} \mathcal{F}_1 &= \{\mathcal{V} \times [0, \bar{U}] \ni (v, u) \mapsto \exp(v^\top \beta) I(\phi(v, u) \geq t) \\ &\quad | t \in [0, \bar{T}], \beta \in \mathcal{B}, \phi \in \Phi_\gamma\}; \\ \mathcal{F}_2 &= \{\mathcal{V} \times [0, \bar{U}] \times \{0, 1\} \ni (v, u, \Delta) \mapsto I(\phi(v, u) \geq t) I(\Delta = 1) \\ &\quad | t \in [0, \bar{T}], \phi \in \Phi_\gamma\}; \\ \mathcal{F}_3 &= \{\mathcal{V} \times [0, \bar{U}] \times \{0, 1\} \ni (v, u, \Delta) \mapsto v \exp(v^\top \beta) I(\phi(v, u) \geq t) I(\Delta = 1) \\ &\quad | t \in [0, \bar{T}], \beta \in \mathcal{B}, \phi \in \Phi_\gamma\}. \end{aligned}$$

Now, consider $\mathcal{F}_1$. What remains to show is that if β_n is a sequence in n converging to β_0 (with respect to the Euclidean norm) and ϕ_n is a sequence in n converging to φ (with respect to $\|\cdot\|_1$) then

$$\begin{aligned} E \left\{ \left[\left(\exp(V^\top \beta_n) I(\phi_n(V, \tilde{U}) \geq t) - \exp(V^\top \beta) I(\varphi(V, \tilde{U}) \geq t) \right) \right. \right. \\ \left. \left. - E \left(\exp(V^\top \beta_n) I(\phi_n(V, \tilde{U}) \geq t) - \exp(V^\top \beta) I(\varphi(V, \tilde{U}) \geq t) \right) \right] \right\}^2 \rightarrow 0 \end{aligned}$$

This result easily follows from the regularity of the distribution of $\tilde{U}$, the compactness of the support of V , the independence between V and $\tilde{U}$, the definition of the norm $\|\cdot\|_1$ and the Lipschitz continuity of $v \mapsto \exp(v^\top \beta)$ in β uniformly in v . More precisely, write

$$\begin{aligned} &E \left\{ \left[\left(\exp(V^\top \beta_n) I(\phi_n(V, \tilde{U}) \geq t) - \exp(V^\top \beta) I(\varphi(V, \tilde{U}) \geq t) \right) \right. \right. \\ &\quad \left. \left. - E \left(\exp(V^\top \beta_n) I(\phi_n(V, \tilde{U}) \geq t) - \exp(V^\top \beta) I(\varphi(V, \tilde{U}) \geq t) \right) \right] \right\}^2 \\ &= I_1 + I_2, \end{aligned}$$

where

$$\begin{aligned} I_1 &= E[\exp(2V^\top \beta_n) I(\phi_n(V, \tilde{U}) \geq t)] \\ &\quad + E[\exp(2V^\top \beta_0) I(\varphi(V, \tilde{U}) \geq t)] \\ &\quad - 2E[\exp(V^\top \beta_n) I(\phi_n(V, \tilde{U}) \geq t) \exp(V^\top \beta_0) I(\varphi(V, \tilde{U}) \geq t)]; \end{aligned}$$

and

$$I_2 = \left(E[(V^\top \beta_n) P(\phi_n(V, \tilde{U}))] - E[(V^\top \beta_0) P(\varphi(V, \tilde{U}))] \right)^2.$$

Now, using similar arguments as before, we have

$$\begin{aligned} \sup_{v \in \mathcal{V}} |P(\phi_n(v, \tilde{U}) \geq t) - P(\varphi(v, \tilde{U}) \geq t)| \\ = \sup_{v \in \mathcal{V}} |P(\tilde{U} \geq \rho_n(v, t)) - P(\tilde{U} \geq \varrho(v, t))| \\ \leq K \|\phi_n - \varphi\|_1 \end{aligned}$$

for a suitable constant K , where $\rho_n(v, t) = \phi_n(v, \cdot)^{-1}(t)$. Also,

$$\sup_{v \in \mathcal{V}} |\exp(v^\top \beta_n) - \exp(v^\top \beta_0)| \leq K \|\beta_n - \beta_0\|,$$

for a further constant K . Therefore, it is easy to see that $I_2 \rightarrow 0$. Using the independence between V and $\tilde{U}$, similar arguments also show that $I_1 \rightarrow 0$. Similar results can be shown for the families $\mathcal{F}_2$ and $\mathcal{F}_3$. Thus, from Corollary 2.3.12 in [25], we have that, for any sequence $\varsigma_n = o(1)$, it holds

$$\begin{aligned} \sup_{s \in [0, \bar{T}], \|\beta - \beta_0\| \leq \varsigma_n, \|\phi - \varphi\|_1 \leq \varsigma_n} |\hat{Q}^\phi(s) - Q^\phi(s) - \hat{Q}^\varphi(s) + Q^\varphi(s)| \quad (33) \\ = o_p(n^{-1/2}); \end{aligned}$$

$$\begin{aligned} \sup_{s \in [0, \bar{T}], \|\beta - \beta_0\| \leq \varsigma_n, \|\phi - \varphi\|_1 \leq \varsigma_n} \|\hat{E}^\phi(g(\beta, v), s) \\ - E^\phi(g(\beta, v), s) - \hat{E}^\varphi(g(\beta_0, v), s) + E^\varphi(g(\beta_0, v), s)\| = o_p(n^{-1/2}); \quad (34) \end{aligned}$$

$$\begin{aligned} \sup_{s \in [0, \bar{T}], \|\beta - \beta_0\| \leq \varsigma_n, \|\phi - \varphi\|_1 \leq \varsigma_n} \|\hat{E}_1^\phi(g(\beta, v), s) \\ - E_1^\phi(g(\beta, v), s) - \hat{E}_1^\varphi(g(\beta_0, v), s) + E_1^\varphi(g(\beta_0, v), s)\| = o_p(n^{-1/2}). \end{aligned}$$

Now, with a similar decomposition as the one in [24] (page 97), we can write

$$\begin{aligned} M_n(\beta, \phi) - M(\beta, \varphi) - (M_n(\beta, \varphi) - M(\beta_0, \varphi)) \\ = \sum_{j=1}^6 D_{jn}(\beta, \phi) + \sum_{j=1}^8 R_{jn}(\beta, \phi), \end{aligned}$$

where $D_{jn}(\beta, \phi) = C_{jn}(\beta, \phi) - C_{jn}(\beta_0, \varphi)$,

$$\begin{aligned} C_{1n}(\beta, \phi) &= \hat{E}_1^\phi(v, 0) - E_1^\phi(v, 0); \\ C_{2n}(\beta, \phi) &= -[\hat{Q}^\phi(0) - Q^\phi(0)]E^\phi(v \exp(v^\top \beta), 0)/E^\phi(\exp(v^\top \beta), 0); \\ C_{3n}(\beta, \phi) &= \int_0^{\bar{T}} [(\hat{Q}^\phi(s) - Q^\phi(s))/E^\phi(\exp(v^\top \beta), s)] dE^\phi(v \exp(v^\top \beta), s); \\ C_{4n}(\beta, \phi) &= \int_0^{\bar{T}} [(\hat{Q}^\phi(s) - Q^\phi(s))E^\phi(v \exp(v^\top \beta), s)/E^\phi(\exp(v^\top \beta), s)^2] \end{aligned}$$

$$\begin{aligned}
& dE^\phi(\exp(v^\top \beta), s); \\
C_{5n}(\beta, \phi) &= \int_0^{\bar{T}} [(\hat{E}^\phi(v \exp(v^\top \beta), s) - E^\phi(v \exp(v^\top \beta), s)) \\
&\quad \times 1/E^\phi(\exp(v^\top \beta), s)] dQ^\phi(s); \\
C_{6n}(\beta, \phi) &= \int_0^{\bar{T}} [(\hat{E}^\phi(\exp(v^\top \beta), s) - E^\phi(\exp(v^\top \beta), s)) \\
&\quad \times E^\phi(v \exp(v^\top \beta), s)/E^\phi(\exp(v^\top \beta), s)^2] dQ^\phi(s);
\end{aligned}$$

and

$$\begin{aligned}
R_{1n}(\beta, \phi) &= \int_0^{\bar{T}} [(\hat{E}^\phi(v \exp(v^\top \beta), s) - E^\phi(v \exp(v^\top \beta), s)) \\
&\quad \times 1/\hat{E}^\phi(\exp(v^\top \beta), s)] d(\hat{Q}^\phi(s) - Q^\phi(s) - \hat{Q}^\varphi(s) + Q^\varphi(s)); \\
R_{2n}(\beta, \phi) &= \int_0^{\bar{T}} \left\{ [(\hat{E}^\phi(v \exp(v^\top \beta), s) - E^\phi(v \exp(v^\top \beta), s)) \right. \\
&\quad \times 1/\hat{E}^\phi(\exp(v^\top \beta), s)] \\
&\quad - [(\hat{E}^\varphi(v \exp(v^\top \beta_0), s) - E^\varphi(v \exp(v^\top \beta_0), s)) \\
&\quad \times 1/\hat{E}^\varphi(\exp(v^\top \beta), s)] \left. \right\} d(\hat{Q}^\varphi(s) - Q^\varphi(s)); \\
R_{3n}(\beta, \phi) &= \int_0^{\bar{T}} [(\hat{E}^\phi(v \exp(v^\top \beta), s) - E^\phi(v \exp(v^\top \beta), s)) \\
&\quad \times (\hat{E}^\phi(\exp(v^\top \beta), s) - E^\phi(\exp(v^\top \beta), s)) \\
&\quad \times (\hat{E}^\phi(\exp(v^\top \beta), s)E^\phi(\exp(v^\top \beta), s))^{-1}] d(Q^\phi(s) - Q^\varphi(s)); \\
R_{4n}(\beta, \phi) &= \int_0^{\bar{T}} \left\{ [(\hat{E}^\phi(v \exp(v^\top \beta), s) - E^\phi(v \exp(v^\top \beta), s)) \right. \\
&\quad \times (\hat{E}^\phi(\exp(v^\top \beta), s) - E^\phi(\exp(v^\top \beta), s)) \\
&\quad \times (\hat{E}^\phi(\exp(v^\top \beta), s)E^\phi(\exp(v^\top \beta), s))^{-1}] \\
&\quad - [(\hat{E}^\varphi(v \exp(v^\top \beta_0), s) - E^\varphi(v \exp(v^\top \beta_0), s)) \\
&\quad \times (\hat{E}^\varphi(\exp(v^\top \beta_0), s) - E^\varphi(\exp(v^\top \beta_0), s)) \\
&\quad \times (\hat{E}^\varphi(\exp(v^\top \beta_0), s)E^\varphi(\exp(v^\top \beta_0), s))^{-1}] \left. \right\} dQ^\varphi(s); \\
R_{5n}(\beta, \phi) &= - \int_0^{\bar{T}} [(\hat{E}^\phi(\exp(v^\top \beta), s) - E^\phi(\exp(v^\top \beta), s)) \\
&\quad \times E^\phi(v \exp(v^\top \beta), s)/(\hat{E}^\phi(\exp(v^\top \beta), s)E^\phi(\exp(v^\top \beta), s))] \\
&\quad d(\hat{Q}^\phi(s) - Q^\phi(s) - \hat{Q}^\varphi(s) + Q^\varphi(s)); \\
R_{6n}(\beta, \phi) &= - \int_0^{\bar{T}} \left\{ [(\hat{E}^\phi(\exp(v^\top \beta), s) - E^\phi(\exp(v^\top \beta), s)) \right. \\
&\quad \times E^\phi(v \exp(v^\top \beta), s)/(\hat{E}^\phi(\exp(v^\top \beta), s)E^\phi(\exp(v^\top \beta), s))]
\end{aligned}$$

$$\begin{aligned}
& - [(\hat{E}^\varphi(\exp(v^\top \beta_0), s) - E^\varphi(\exp(v^\top \beta_0), s))E^\varphi(v \exp(v^\top \beta_0), s) \\
& \times (\hat{E}^\varphi(\exp(v^\top \beta_0), s)E^\varphi(\exp(v^\top \beta_0), s))^{-1}] \} d(\hat{Q}^\varphi(s) - Q^\varphi(s)); \\
R_{7n}(\beta, \phi) &= \int_0^{\bar{T}} [(\hat{E}^\phi(\exp(v^\top \beta), s) - E^\phi(\exp(v^\top \beta), s))^2 E^\phi(v \exp(v^\top \beta), s) \\
& \times (\hat{E}^\phi(\exp(v^\top \beta), s)E^\phi(\exp(v^\top \beta), s)^2)^{-1}] d(Q^\phi(s) - Q^\varphi(s)); \\
R_{8n}(\beta, \phi) &= \\
& \int_0^{\bar{T}} \left\{ [(\hat{E}^\phi(\exp(v^\top \beta), s) - E^\phi(\exp(v^\top \beta), s))^2 E^\phi(v \exp(v^\top \beta), s) \right. \\
& \times (\hat{E}^\phi(\exp(v^\top \beta), s)E^\phi(\exp(v^\top \beta), s)^2)^{-1}] \\
& - [(\hat{E}^\varphi(\exp(v^\top \beta_0), s) - E^\varphi(\exp(v^\top \beta_0), s))^2 E^\varphi(v \exp(v^\top \beta_0), s) \\
& \times (\hat{E}^\varphi(\exp(v^\top \beta_0), s)E^\varphi(\exp(v^\top \beta_0), s)^2)^{-1}] \} dQ^\varphi(s).
\end{aligned}$$

This decomposition, and the equalities in (33) imply the assertion.

(2.6) The statement can be reduced to showing that $n^{1/2}\{M_n(\beta_0, \varphi) + \Gamma_2(\beta_0, \varphi)[\hat{\varphi} - \varphi]\}$ can be expressed as a sum of iid terms of the form

$$\begin{aligned}
& n^{1/2}\{M_n(\beta_0, \varphi) + \Gamma_2(\beta_0, \varphi)[\hat{\varphi} - \varphi] \\
& = n^{-1/2} \sum_{i=1}^n \eta(Y_i, \delta_i, Z_i, W_i, X_i, \Delta_i, \tilde{U}_i) + o_p(1).
\end{aligned}$$

The fact that $n^{1/2}M_n(\beta_0, \varphi)$ admits such representation is thanks to the decomposition given in (2.5). We now show that the same holds for the term $n^{1/2}\Gamma_2(\beta_0, \varphi)[\hat{\varphi} - \varphi]$, whose expression is given in (26).

Define $p(z, x) = P(Z = z|X = x)$ and denote by $\nu(x)$ the density of X . First, consider $e_1(v, 0)[\hat{\varphi} - \varphi]$, that is

$$\begin{aligned}
& e_1(v, 0)[\hat{\varphi} - \varphi] \\
& = - \int_{\mathcal{X}} \sum_{l=1}^L \frac{(z_l^\top, x^\top)^\top}{\frac{\partial}{\partial u} \varphi(z_l, x, 0)} (\hat{\varphi}(z_l, x, 0) - \varphi(z_l, x, 0)) p(z_l, x) \nu(x) dx.
\end{aligned}$$

Denote $\Gamma(\varphi^x, F^x)_l^{-1}$ the l th row the matrix $\Gamma(\varphi^x, F^x)^{-1}$. Given two vectors v and w denote by vw the inner product, defined by $v^\top w$. We can write

$$\begin{aligned}
& e_1(v, 0)[\hat{\varphi} - \varphi] \\
& = - \int_{\mathcal{X}} \sum_{l=1}^L \frac{(z_l^\top, x^\top)^\top}{\frac{\partial}{\partial u} \varphi(z_l, x, 0)} \left((nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \right. \\
& \quad \times \left. (\Gamma(\varphi^x, F^x)^{-1})_l \left(\sum_{k=1}^L \eta^F(Y_i, \delta_i, Z_i, W_i, \varphi(z_l, x, 0), z_l, x, w_k) \right) \right)_{k=1}^K \Bigg)
\end{aligned}$$

$$\begin{aligned} & \times p(z_l, x)\nu(x)dx + o_p(n^{-1/2}) \\ & = (nh)^{-1} \int_{\mathcal{X}} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) b(Y_i, \delta_i, Z_i, W_i, x) dx + o_p(n^{-1/2}), \end{aligned}$$

where

$$\begin{aligned} b(Y_i, \delta_i, Z_i, W_i, x) &= - \sum_{l=1}^L \frac{(z_l^\top, x^\top)^\top}{\frac{\partial}{\partial u} \varphi(z_l, x, 0)} \left((\Gamma(\varphi^x, F^x)^{-1})_l \right. \\ & \times \left. \left(\sum_{l=1}^L \eta^F(Y_i, \delta_i, Z_i, W_i, \varphi(z_l, x, 0), z_l, x, w_k) \right)_{k=1}^K \right) p(z_l, x)\nu(x). \end{aligned}$$

Thus, by standard change of variable and Taylor expansion, we obtain

$$\begin{aligned} & e_1(v, 0)[\hat{\varphi} - \varphi] \\ &= n^{-1} \sum_{i=1}^n \int_{\mathcal{X}} b(Y_i, \delta_i, Z_i, W_i, hx + X_i) K(x) dx \\ &= n^{-1} \sum_{i=1}^n \int_{\mathcal{X}} [b(Y_i, \delta_i, Z_i, W_i, X_i) \\ & \quad + \sum_{j=1}^{\nu-1} (hx)^j \frac{\partial^j}{\partial x^j} b(Y_i, \delta_i, Z_i, W_i, X_i) + O_p(h^\nu)] K(x) dx \\ &= n^{-1} \sum_{i=1}^n b(Y_i, \delta_i, Z_i, W_i, X_i) + o_p(n^{-1/2}), \end{aligned}$$

where in the last equality we used that $O_p(h^\nu) = o_p(n^{-1/2})$. Now, write $q^\varphi(t)[\hat{\varphi} - \varphi] = I_1 + I_2$, where

$$\begin{aligned} I_1 &= \int_{\mathcal{V}} g_0(v, t) (\hat{\varphi}(v, \rho(v, t)) - \varphi(v, \rho(v, t))) d\mu(v); \\ I_2 &= \int_{\mathcal{V}} g_1(v, t) \left(\frac{\partial}{\partial u} \hat{\varphi}(v, \rho(v, t)) - \frac{\partial}{\partial u} \varphi(v, \rho(v, t)) \right) d\mu(v); \end{aligned}$$

and

$$\begin{aligned} g_0(v, t) &= \left\{ I(\varrho(v, t) < \bar{U} - \delta) + I(\bar{U} - \delta \leq \varrho(v, t) \leq \bar{U}) \frac{\bar{U}}{\delta} \right\} \\ & \quad \times \frac{\varrho'(v, t) \varphi''(v, \varrho(v, t))}{\varphi'(v, \varrho(v, t))} \\ & \quad + \left\{ I(\bar{U} - \delta \leq \varrho(v, t) \leq \bar{U}) \frac{\varrho(v, t)'}{\delta} \right\} \frac{1}{(\varphi'(v, \varrho(v, t)))^2} \\ g_1(v, t) &= \left\{ I(\varrho(v, t) < \bar{U} - \delta) + I(\bar{U} - \delta \leq \varrho(v, t) \leq \bar{U}) \frac{\bar{U}}{\delta} \right\} \end{aligned}$$

$$\times \frac{1}{(\varphi'(v, \varrho(v, t)))^2}$$

Using a similar argument as for $e_1(v, 0)[\hat{\varphi} - \varphi]$ we again can show that there exists a function $\tilde{b}(y, \delta, z, w, x)$ such that

$$I_1 = n^{-1} \sum_{i=1}^n \tilde{b}(Y_i, \delta_i, Z_i, W_i, X_i) + o_p(n^{-1/2}).$$

We now focus on I_2 , and we write it as $I_2 = I_{21} + I_{22} + o_p(n^{-1/2})$, where

$$\begin{aligned} I_{21} &= \int_{\mathcal{X}} \sum_{l=1}^L g_0(z_l, x, t) \left[(nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \right. \\ &\quad \times \sum_{j=1}^L \kappa_j^l(x, \varrho(z_l, x, t)) \eta_j(Y_i, \delta_i, Z_i, W_i, x, \varrho(z_l, x, t)) \\ &\quad + (nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \sum_{j=1}^L \sum_{k=1}^K \zeta_{jk}^l(x, \varrho(z_l, x, t)) I(W_i = w_k) \\ &\quad \times \frac{I(Z_i = z_l) - p_{z_l, x, w_k}}{f_{X, W}(x, w_k)} f(t|z_l, x, w_k) \Big] \\ &\quad \times p(z_l, x) \nu(x) dx \\ I_{22} &= \int_{\mathcal{X}} \sum_{l=1}^L g_1(z_l, x, t) \left[(nh\epsilon)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \right. \\ &\quad \times \int \tilde{K}'(s) \sum_{j=1}^L \sum_{k=1}^K \zeta_{jk}^l(x, \varrho(z_l, x, t)) \xi^F(Y_i, \delta_i, Z_i, W_i, t - s\epsilon, z_l, x, w_k) ds \Big] \\ &\quad \times p(z_l, x) \nu(x) dx. \end{aligned}$$

So, again with a similar argument as before, I_{21} admits such representation, and we now discuss I_{22} . Theorem 3 (b) of [27] implies that

$$\begin{aligned} &(nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \xi^F(Y_i, \delta_i, Z_i, W_i, t - s\epsilon, z_l, x, w) \\ &= (nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \xi^F(Y_i, \delta_i, Z_i, W_i, t, z, x, w) \\ &\quad + O_p((\epsilon \log n / (nh))^{1/2} + \epsilon^{1/2} h^\nu). \end{aligned}$$

Since $\int \tilde{K}'(s) ds = 0$, $\int s \tilde{K}'(s) ds = -1$ and $\int s^j \tilde{K}'(s) ds = 0$ for $j = 2, \dots, \pi$, we obtain

$$I_{22} = \int_{\mathcal{X}} \sum_{l=1}^L g_1(z_l, x, t) \left[(nh\epsilon)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \right.$$

$$\begin{aligned}
& \times \int \tilde{K}'(s) \sum_{j=1}^L \sum_{k=1}^K \zeta_{jk}^l(x, \varrho(z_l, x, t)) F(t - \epsilon s, z|x, w) ds \Big] \\
& \times p(z_l, x) \nu(x) dx + O_p((\log n / (nh\epsilon))^{1/2} + \epsilon^{-1/2} h^\nu) \\
& = \int_{\mathcal{X}} \sum_{l=1}^L g_1(z_l, x, t) \\
& \times \left[(nh\epsilon)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \right. \\
& \times \int \tilde{K}'(s) \sum_{j=1}^L \sum_{k=1}^K \zeta_{jk}^l(x, \varrho(z_l, x, t)) (-s\epsilon f(t, z_l|x, w_k) \\
& \quad \left. + \frac{1}{2} s^2 \epsilon^2 f'(t, z_l|x, w_k) - \frac{1}{6} s^3 \epsilon^3 f''(t, z_l|x, w_k) + \dots + O(\epsilon^{\pi+1})) ds \right] \\
& \times p(z_l, x) \nu(x) dx + O_p((\log n / (nh\epsilon))^{1/2} + \epsilon^{-1/2} h^\nu) \\
& = \int_{\mathcal{X}} \sum_{l=1}^L g_1(z_l, x, t) \\
& \times \left[(nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \sum_{j=1}^L \sum_{k=1}^K \zeta_{jk}^l(x, \varrho(z_l, x, t)) (f(t, z_l|x, w_k) \right. \\
& \quad \left. + O(\epsilon^\pi)) \right] \times p(z_l, x) \nu(x) dx \\
& \quad + O_p((\log n / (nh\epsilon))^{1/2} + \epsilon^{-1/2} h^\nu)
\end{aligned}$$

Note that $O_p((\log n / (nh\epsilon))^{1/2} + \epsilon^{-1/2} h^\nu) = o_p(n^{-1/2})$ since $w < 2\nu u - 1$. All the other terms in $\Gamma_2(\beta_0, \varphi)[\hat{\varphi} - \varphi]$ can follow the same arguments, from which the assertion follows. $\square$

F. Auxiliary lemmas

F.1. Lemmas for Theorem C.1

Lemma F.1. Under Assumption 4, we have

- (i) $\sup_{t \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} |\hat{F}(t|z, x, w) - F(t|z, x, w)| = O_p((\log n / (nh))^{1/2} + h^\nu + \epsilon^\pi).$
- (ii) $\sup_{t \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} |\hat{f}(t|z, x, w) - f(t|z, x, w)| = O_p((\log n / (nh\epsilon))^{1/2} + h^\nu + \epsilon^\pi).$
- (iii) $\sup_{t \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} |\frac{\partial}{\partial t} \hat{f}(t|z, x, w) - \frac{\partial}{\partial t} f(t|z, x, w)| = o_p(1).$
- (iv) $\sup_{t \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} |\frac{\partial^2}{\partial t^2} \hat{f}(t|z, x, w) - \frac{\partial^2}{\partial t^2} f(t|z, x, w)| = o_p(1).$

(iv)

$$\sup_{t_1, t_2 \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} \left| \frac{\partial^2}{\partial t^2} (\hat{f}(t_1|z, x, w) - \hat{f}(t_2|z, x, w) - f(t_1|z, x, w) + f(t_2|z, x, w)) \right| / |t_1 - t_2|^\alpha = o_p(1).$$

Proof. First, Proposition 1 in [2] implies

$$\sup_{t \in [0, \bar{T}]} \sup_{z \in \mathcal{Z}, w \in \mathcal{W}, x \in \mathcal{X}} |\tilde{F}(t|z, x, w) - F(t|z, x, w)| = O_p((\log n/(nh))^{1/2} + h^\nu). \quad (35)$$

The result in (i) then follows by integration by parts and the following decomposition:

$$\begin{aligned} \hat{F}(t|z, x, w) - F(t|z, x, w) &= \int H\left(\frac{t-u}{\epsilon}\right) d\tilde{F}(u|z, x, w) - F(t|z, x, w) \\ &= \int \tilde{F}(t - \epsilon u|z, x, w) \tilde{K}(u) du - F(t|z, x, w) \\ &= \int (\tilde{F}(t - \epsilon u|z, x, w) - F(t - \epsilon u|z, x, w)) \tilde{K}(u) du \\ &\quad + \int (F(t - \epsilon u|z, x, w) - F(t|z, x, w)) \tilde{K}(u) du. \end{aligned}$$

Specifically, we obtain that

$$\begin{aligned} &\left| \int (\tilde{F}(t - \epsilon u|z, x, w) - F(t - \epsilon u|z, x, w)) \tilde{K}(u) du \right| \\ &\leq \int O_p((\log n/(nh))^{1/2} + h^\nu) \tilde{K}(u) du \\ &= O_p((\log n/(nh))^{1/2} + h^\nu). \end{aligned}$$

Using Assumption 4, the Taylor expansion $F(t - \epsilon u|z, x, w) - F(t|z, x, w) = -\epsilon u f(t|z, x, w) + \frac{1}{2} \epsilon^2 u^2 \frac{\partial}{\partial t} f(t|z, x, w) + \cdots + O(\epsilon^\pi)$ and the boundedness of $f(t|z, x, w)$ and its derivatives in the argument t , we have

$$\int (F(t - \epsilon u|z, x, w) - F(t|z, x, w)) \tilde{K}(u) du = O(\epsilon^\pi).$$

The result in (ii) corresponds to the rate of convergence of the term (T_3) in the proof of Lemma 2 in the Supplementary Material of [16]. More precisely, write $\hat{f}(t|z, x, w) - f(t|z, x, w) = I_1 + I_2$, where

$$I_1 = \epsilon^{-1} \int \tilde{K}\left(\frac{t-s}{\epsilon}\right) d(\tilde{F}(s|z, x, w) - F(s|z, x, w));$$

$$I_2 = \epsilon^{-1} \int \tilde{K} \left(\frac{t-s}{\epsilon} \right) dF(s|z, x, w) - f(t|z, x, w).$$

Now, integrating by parts and using standard change of variables, we can write $I_1 = I_{11} + I_{12} + I_{13}$, where

$$\begin{aligned} I_{11} &= \epsilon^{-1} \int (\tilde{F}(t - u\epsilon|z, x, w) \\ &\quad - E\tilde{F}(t - u\epsilon|z, x, w) - \tilde{F}(t|z, x, w) + E\tilde{F}(t|z, x, w)) \tilde{K}'(u) du; \\ I_{12} &= \epsilon^{-1} \int (E\tilde{F}(t - u\epsilon|z, x, w) \\ &\quad - F(t - u\epsilon|z, x, w) - E\tilde{F}(t|z, x, w) + F(t|z, x, w)) \tilde{K}'(u) du; \\ I_{13} &= \epsilon^{-1} (\tilde{F}(t|z, x, w) - F(t|z, x, w)) \int \tilde{K}'(u) du. \end{aligned}$$

Now, I_{13} is straightforwardly equal to 0, because of Assumption 4. Following the proof of Theorem 3(b) of [27] it can be shown that there exists a finite constant M_1 such that

$$\begin{aligned} I_{11} &\leq M_1 \epsilon^{-1} \sup_{t, s \in [0, \bar{T}], |t-s| \leq \epsilon, z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} |\tilde{F}(t|z, x, w) - E\tilde{F}(t|z, x, w) \\ &\quad - \tilde{F}(s|z, x, w) + E\tilde{F}(s|z, x, w)| \\ &= O_p((\log n / (nh\epsilon))^{1/2}). \end{aligned}$$

For the term I_{12} , it is possible to follow the work of [28] with consideration here of higher-order kernels. We have that

$$I_{12} \leq \epsilon^{-1} \int (b(t - u\epsilon|z, x, w)h^\nu + o(h^\nu) + O(n^{-1}) - b(t|z, x, w)h^\nu) \tilde{K}'(u) du,$$

where the function b is in equation (3.3) in [28], and where their Taylor expansion term $o(h^\nu)$ can be replaced by $O(h^{\nu+1})$. Hence, under Lipschitz continuous requirements for the function $b(t|z, x, w)$ with respect to t induced by Assumption 4, we obtain $I_{12} = \epsilon^{-1} O(\epsilon h^\nu + h^{\nu+1} + n^{-1}) = O(h^\nu)$, due to the assumptions on the bandwidths.

This argument shows that $I_1 = O_p((\log n / (nh\epsilon))^{1/2} + h^\nu)$. For the term I_2 , we can write

$$I_2 = \epsilon^{-1} \int \tilde{K} \left(\frac{t-s}{\epsilon} \right) (f(s|z, x, w) - f(t|z, x, w)) ds,$$

and apply standard Taylor expansion and change of variables to show that $I_2 = O(\epsilon^\pi)$.

For the result in (iii), we can use a similar argument, where we will have $\epsilon^{-2} \tilde{K}''(u)$ in place of $\epsilon^{-1} \tilde{K}'(u)$, obtaining a rate of convergence for $\frac{\partial}{\partial t} \hat{f}(t|z, x, w) - \frac{\partial}{\partial t} f(t|z, x, w)$ equal to

$$O_p((\log n)^{1/2} n^{-1/2} h^{-1/2} \epsilon^{-3/2}) + \epsilon^{-2} O_p(h^{\nu+1}) + O_p(h^\nu) + \epsilon^{-2} O_p(n^{-1}) + O(\epsilon^\pi).$$

The same applies for the result in (iv) with rate of convergence

$$O_p((\log n)^{1/2} n^{-1/2} h^{-1/2} \epsilon^{-5/2}) + \epsilon^{-3} O_p(h^{\nu+1}) + O_p(h^\nu) + \epsilon^{-3} O_p(n^{-1}) + O(\epsilon^\pi).$$

The result in (v) follows by the almost sure behavior of the modulus of continuity of $\frac{d^2}{dt^2} \hat{f}(t|z, x, w) - \frac{d^2}{dt^2} f(t|z, x, w)$, with a similar proof as the one given in Lemma 2.11 in [26]. Thus

$$\begin{aligned} & \sup_{t_1, t_2 \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} \left| \frac{\partial^2}{\partial t^2} (\hat{f}(t_1|z, x, w) \right. \\ & \quad \left. - \hat{f}(t_2|z, x, w) - f(t_1|z, x, w) + f(t_2|z, x, w)) \right| / |t_1 - t_2|^\alpha \\ &= O_p((\log n)^{1/2} n^{-1/2} h^{-1/2} \epsilon^{-5/2-\alpha}) + \epsilon^{-3-\alpha} O_p(h^{\nu+1}) \\ & \quad + \epsilon^{-\alpha} O_p(h^\nu) + \epsilon^{-3-\alpha} O_p(n^{-1}) + O(\epsilon^\pi). \end{aligned} \quad \square$$

Lemma F.2. Under Assumption 4, we have

$$\begin{aligned} (i) \quad & \sup_{z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} |\hat{p}_{z,x,w} - p_{z,x,w}| = O_p((\log n/(nh))^{1/2} + h^\nu). \\ (ii) \quad & \sup_{z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} \left| \frac{\partial}{\partial x} \hat{p}_{z,x,w} - \frac{\partial}{\partial x} p_{z,x,w} \right| = o_p(1). \\ (iii) \quad & \sup_{z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} \left| \frac{\partial^2}{\partial x^2} \hat{p}_{z,x,w} - \frac{\partial^2}{\partial x^2} p_{z,x,w} \right| = o_p(1). \\ (iv) \quad & \sup_{z \in \mathcal{Z}, x_1, x_2 \in \mathcal{X}, w \in \mathcal{W}} \frac{\left| \frac{\partial^2}{\partial x^2} (\hat{p}_{z,x_1,w} - \hat{p}_{z,x_2,w} - p_{z,x_1,w} + p_{z,x_2,w}) \right|}{|x_1 - x_2|^\alpha} = o_p(1). \end{aligned}$$

Proof. The quantity $\hat{p}_{z,x,w}$ corresponds to an adjusted Nadaraya-Watson estimator designed for estimating a conditional probability of a discrete outcome variable. The outcome variable, in this case, is the variable Z , the conditioning variable is X and the weight for the i th observation is equal to the indicator of whether W_i equals w or not. The result in (i) then follows from [18]. The results in (ii) – (iii) can be proved similarly as in Proposition 5.3 of [26], obtaining

$$\sup_{z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} \left| \frac{\partial}{\partial x} \hat{p}_{z,x,w} - \frac{\partial}{\partial x} p_{z,x,w} \right| = O_p((\log n/(nh^3))^{1/2} + h^\nu),$$

and

$$\sup_{z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} \left| \frac{\partial^2}{\partial x^2} \hat{p}_{z,x,w} - \frac{\partial^2}{\partial x^2} p_{z,x,w} \right| = O_p((\log n/(nh^5))^{1/2} + h^\nu).$$

The result in (iv) follows from Lemma 2.11 in [26], and we obtain

$$\begin{aligned} & \sup_{z \in \mathcal{Z}, x_1, x_2 \in \mathcal{X}, w \in \mathcal{W}} \frac{\left| \frac{\partial^2}{\partial x^2} (\hat{p}_{z,x_1,w} - \hat{p}_{z,x_2,w} - p_{z,x_1,w} + p_{z,x_2,w}) \right|}{|x_1 - x_2|^\alpha} \\ &= O_p((\log n)^{1/2} (nh^{5+2\alpha})^{-1/2} + h^\nu) \end{aligned} \quad \square$$

Lemma F.3. Under Assumption 5, we have

$$\begin{aligned}
(i) \quad & \sup_{t \in [0, \bar{T}], x \in \mathcal{X}, z \in \mathcal{Z}, w \in \mathcal{W}} \left| \frac{\partial}{\partial x} \hat{F}(t|z, x, w) - \frac{\partial}{\partial x} F(t|z, x, w) \right| = o_p(1). \\
(ii) \quad & \sup_{t \in [0, \bar{T}], x \in \mathcal{X}, z \in \mathcal{Z}, w \in \mathcal{W}} \left| \frac{\partial^2}{\partial x^2} \hat{F}(t|z, x, w) - \frac{\partial^2}{\partial x^2} F(t|z, x, w) \right| = o_p(1). \\
(iii) \quad &
\end{aligned}$$

$$\begin{aligned}
& \sup_{t \in [0, \bar{T}], x_1, x_2 \in \mathcal{X}, z \in \mathcal{Z}, w \in \mathcal{W}} \left| \frac{\partial^2}{\partial x^2} (\hat{F}(t|z, x_1, w) \right. \\
& \quad \left. - \hat{F}(t|z, x_2, w) - F(t|z, x_1, w) + F(t|z, x_2, w)) \right| / |x_1 - x_2|^\alpha = o_p(1).
\end{aligned}$$

Proof. The results in (i) and (iii) follow from Proposition 5.5 in [26]:

$$\begin{aligned}
& \sup_{t \in [0, \bar{T}], x \in \mathcal{X}, z \in \mathcal{Z}, w \in \mathcal{W}} \left| \frac{\partial}{\partial x} \hat{F}(t|z, x, w) - \frac{\partial}{\partial x} F(t|z, x, w) \right| \\
& \quad = O_p((\log n / (nh^3))^{1/2} + h^\nu + \epsilon^\pi),
\end{aligned}$$

and

$$\begin{aligned}
& \sup_{t \in [0, \bar{T}], x \in \mathcal{X}, z \in \mathcal{Z}, w \in \mathcal{W}} \left| \frac{\partial^2}{\partial x^2} \hat{F}(t|z, x, w) - \frac{\partial^2}{\partial x^2} F(t|z, x, w) \right| \\
& \quad = O_p((\log n / (nh^5))^{1/2} + h^\nu + \epsilon^\pi h^{-1}).
\end{aligned}$$

The result in (iii) follows from Proposition 5.6 in [26],

$$\begin{aligned}
& \sup_{t \in [0, \bar{T}], x_1, x_2 \in \mathcal{X}, z \in \mathcal{Z}, w \in \mathcal{W}} \left| \frac{\partial^2}{\partial x^2} (\hat{F}(t|z, x_1, w) \right. \\
& \quad \left. - \hat{F}(t|z, x_2, w) - F(t|z, x_1, w) + F(t|z, x_2, w)) \right| / |x_1 - x_2|^\alpha \\
& \quad = O_p((\log n / (nh^{5+2\alpha}))^{1/2} + h^\nu + \epsilon^\pi h^{-1-\alpha}). \quad \square
\end{aligned}$$

Lemma F.4. Under Assumption 4, we have

$$\begin{aligned}
(i) \quad & \hat{F}(t|z, x, w) - F(t|z, x, w) \\
& = (nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \xi^F(Y_i, \delta_i, Z_i, W_i, t, z, x, w) \\
& \quad + R_n(t, z, x, w),
\end{aligned}$$

where

$$\sup_{t \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} |R_n(t, z, x, w)| = O_p\left((\log n / (nh))^{3/4} + h^\nu + \epsilon^\pi\right). \quad (36)$$

(ii)

$$\hat{f}(t|z, x, w) - f(t|z, x, w)$$

$$\begin{aligned}
&= (nh\epsilon)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \int \tilde{K}'(u) \xi^F(Y_i, \delta_i, Z_i, W_i, t - u\epsilon, z, x, w) du \\
&\quad + r_n(t, z, x, w),
\end{aligned}$$

where

$$\sup_{t \in [0, \bar{T}], z \in \mathcal{Z}, x \in \mathcal{X}, w \in \mathcal{W}} |r_n(t, z, x, w)| = O_p((\log n / (nh\epsilon))^{3/4} + h^\nu + \epsilon^\pi).$$

Proof. The result in (i) is an easy adaptation of Lemma 1 in the Supplementary Material of [16]. The result in (ii) follows from (i) and the fact that

$$\begin{aligned}
\hat{f}(t|z, x, w) - f(t|z, x, w) &= \epsilon^{-1} \int \tilde{K}\left(\frac{t-s}{\epsilon}\right) d(\tilde{F}(s|z, x, w) - F(s|z, x, w)) \\
&\quad + \epsilon^{-1} \int \tilde{K}\left(\frac{t-s}{\epsilon}\right) dF(s|z, x, w) - f(t|z, x, w).
\end{aligned}$$

More precisely, rewrite $\hat{f}(t|z, x, w) - f(t|z, x, w)$ as

$$\begin{aligned}
&\hat{f}(t|z, x, w) - f(t|z, x, w) \\
&= \epsilon^{-2} \int \tilde{K}'\left(\frac{t-s}{\epsilon}\right) (\tilde{F}(s|z, x, w) - F(s|z, x, w)) ds \\
&\quad + \epsilon^{-1} \int \tilde{K}\left(\frac{t-s}{\epsilon}\right) (f(s|z, x, w) - f(t|z, x, w)) ds \\
&= \epsilon^{-2} \int \tilde{K}'\left(\frac{t-s}{\epsilon}\right) (\hat{F}(s|z, x, w) - F(s|z, x, w)) ds \\
&\quad + \epsilon^{-1} \int \tilde{K}\left(\frac{t-s}{\epsilon}\right) (f(s|z, x, w) - f(t|z, x, w)) ds \\
&\quad + O_p((\log n / (nh))^{3/4} + h^\nu + \epsilon^3) \\
&= (nh\epsilon)^{-1} \sum_{i=1}^n \epsilon^{-1} \int \tilde{K}'\left(\frac{t-s}{\epsilon}\right) K\left(\frac{x - X_i}{h}\right) \xi^F(Y_i, \delta_i, Z_i, W_i, s, z, x, w) ds \\
&\quad + O_p((\log n / (nh))^{3/4} + h^\nu + \epsilon^3) \\
&\quad + \int \tilde{K}(u) (f(t - \epsilon u|z, x, w) - f(t|z, x, w)) du. \tag{37}
\end{aligned}$$

Since

$$\begin{aligned}
f(t - \epsilon u|z, x, w) &= f(t|z, x, w) - u\epsilon \frac{\partial f(t|z, x, w)}{\partial t} \\
&\quad + \frac{1}{2} u^2 \epsilon^2 \frac{\partial^2 f(t|z, x, w)}{\partial t^2} + \dots + O(\epsilon^\pi),
\end{aligned}$$

and because of Assumption 4, we have

$$\int \tilde{K}(u) (f(t - \epsilon u|z, x, w) - f(t|z, x, w)) du = O(\epsilon^\pi).$$

Therefore, by standard change of variable, equation (37) leads to the result in (ii). $\square$

F.2. Lemma for Theorem 5.1

Lemma F.5. *Under Assumptions 4 and 5, it holds $\|\hat{\varphi} - \varphi\|_\infty = o_p(1)$.*

Proof. Theorem C.1 implies that

$$\sup_{x \in \mathcal{X}} \|A(\hat{\varphi}^x, \hat{F}^x) - A(\hat{\varphi}^x, F^x)\|_\infty = o_p(n^{-1/4}).$$

Now, the proof of the result we are interested in is similar to the one of Theorem 4.1 in [5]. More precisely,

$$\begin{aligned} \|A(\hat{\varphi}^x, F^x)\|_\infty &\leq \|A(\hat{\varphi}^x, \hat{F}^x)\|_\infty + \|A(\hat{\varphi}^x, \hat{F}^x) - A(\hat{\varphi}^x, F^x)\|_\infty \\ &\leq \|A(\hat{\varphi}^x, \hat{F}^x)\|_\infty + \sup_{x \in \mathcal{X}} \|A(\hat{\varphi}^x, \hat{F}^x) - A(\hat{\varphi}^x, F^x)\|_\infty \\ &= \|A(\hat{\varphi}^x, \hat{F}^x)\|_\infty + o_p(n^{-1/4}). \end{aligned} \quad (38)$$

By definition,

$$\begin{aligned} \|A(\hat{\varphi}^x, \hat{F}^x)\|_\infty &= \inf_{\theta \in \mathcal{F}_Z^{\bar{U}, \bar{T}}} \|A(\theta, \hat{F}^x)\|_\infty \\ &\leq \inf_{\theta \in \mathcal{F}_Z^{\bar{U}, \bar{T}}} \|A(\theta, F^x)\|_\infty + \sup_{\theta \in \mathcal{F}_Z^{\bar{U}, \bar{T}}} \|A(\theta, \hat{F}^x) - A(\theta, F^x)\|_\infty \\ &\leq \sup_{x \in \mathcal{X}} \inf_{\theta \in \mathcal{F}_Z^{\bar{U}, \bar{T}}} \|A(\theta, F^x)\|_\infty + \sup_{x \in \mathcal{X}} \sup_{\theta \in \mathcal{F}_Z^{\bar{U}, \bar{T}}} \|A(\theta, \hat{F}^x) - A(\theta, F^x)\|_\infty \\ &= 0 + o_p(n^{-1/4}), \end{aligned}$$

where in the last equality we used Assumption 5.

This argument and (38) imply that $\sup_{x \in \mathcal{X}} \|A(\hat{\varphi}^x, F^x)\|_\infty = o_p(n^{-1/4})$, and so $\|\hat{\varphi}^x - \varphi^x\|_\infty = o_p(1)$ uniformly in $x \in \mathcal{X}$ by Assumption 5. $\square$

F.3. Lemmas for Theorem 5.2

Lemma F.6. *Under Assumption 4 and 6, there exist $\varsigma, c > 0$ such that for any $\theta \in \mathcal{F}_Z^{\bar{U}, \bar{T}}$ such that $\|\theta - \varphi^x\|_\infty \leq \varsigma$, it holds $\|\theta - \varphi^x\|_\infty \leq c\|A(\theta, F^x)\|_\infty$, where ς, c do not depend on $x \in \mathcal{X}$.*

Proof. Since $F(\cdot, z|x, w)$ is twice differentiable, also $A(\cdot, F^x)$ is twice differentiable in the first argument. Since $F(\cdot, z|x, w)$ has a bounded second derivative, uniformly in z, w, x , also the second derivative of $A(\cdot, F^x)(u)$ is bounded in $[0, \bar{T}]$ by a constant M_1 which does not depend on x . Using the Taylor expansion, for $\theta \in \mathcal{F}_Z^{\bar{U}, \bar{T}}$, it holds

$$A(\theta, F^x)(u) \leq A(\varphi^x, F^x)(u) + \Gamma(\varphi^x, F^x)(u)(\theta(u) - \varphi^x(u))$$

$$+ M_1 \|\theta(u) - \varphi^x(u)\|^2.$$

This implies that

$$\begin{aligned} \|A(\theta, F^x)(u) - A(\varphi^x, F^x)(u) - \Gamma(\varphi^x, F^x)(u)(\theta(u) - \varphi^x(u))\| \\ \leq M_1 \|\theta(u) - \varphi^x(u)\|^2 \\ \leq M_1 \|\theta - \varphi^x\|_\infty^2. \end{aligned}$$

The previous inequality yields

$$\|A(\theta, F^x) - A(\varphi^x, F^x) - \Gamma(\varphi^x, F^x)(\theta - \varphi^x)\|_\infty \leq M_1 \|\theta - \varphi^x\|_\infty^2.$$

Since $A(\varphi^x, F^x) = 0$, we have

$$\|\Gamma(\varphi^x, F^x)(\theta - \varphi^x)\|_\infty \leq \|A(\theta, F^x)\|_\infty + M_1 \|\theta - \varphi^x\|_\infty^2.$$

By Assumption 6, there exists a constant $M_2 > 0$, which does not depend on x or $u \in [0, \bar{U}]$ such that

$$M_2 \|\theta - \varphi^x\|_\infty - M_1 \|\theta - \varphi^x\|_\infty^2 \leq \|A(\theta, F^x)\|_\infty.$$

Choosing $\varsigma > 0$ small enough, we obtain that when $\|\theta - \varphi^x\|_\infty \leq \varsigma$, then

$$\frac{M_2}{2} \|\theta - \varphi^x\|_\infty \leq \|A(\theta, F^x)\|_\infty,$$

which finishes the proof. $\square$

Lemma F.7. Under Assumptions 4, 5 and 6, we have $\|\hat{\varphi} - \varphi\|_\infty = o_p(n^{-1/4})$.

Proof. Similarly as in the proof of Lemma F.5, we obtain

$$\sup_{x \in \mathcal{X}} \|A(\hat{\varphi}^x, F^x)\|_\infty = o_p(n^{-1/4}),$$

and $\|\hat{\varphi}^x - \varphi^x\|_\infty = o_p(1)$ uniformly in $x \in \mathcal{X}$. Hence, by Lemma F.6, we have $\|\hat{\varphi}^x - \varphi^x\|_\infty \leq c \|A(\hat{\varphi}^x, F^x)\|_\infty$ with probability approaching to 1, where c does not depend on x . We obtain,

$$\sup_{x \in \mathcal{X}} \|\hat{\varphi}^x - \varphi^x\|_\infty \leq c \sup_{x \in \mathcal{X}} \|A(\hat{\varphi}^x, F^x)\|_\infty = o_p(n^{-1/4}),$$

which is the assertion. $\square$

Lemma F.8. Given a function $f(x)$, denote by $f^{-1}(y)$ its inverse, such that $f^{-1}(f(x)) = x$. Denote by T the operator that associates to a function $f(x)$ its inverse $T(f)(y) = f^{-1}(y)$. Denote by $\Gamma(f)[h]$ the Gâteaux derivative of the operator T at f in direction h , defined as

$$\Gamma(f)[h](y) = \lim_{t \rightarrow 0} \frac{T(f + th)(y) - T(f)(y)}{t}.$$

Then, the Gâteaux derivative of the operator T takes the following form:

$$\Gamma(f)[h](y) = -\frac{h(f^{-1}(y))}{f'(f^{-1}(y))},$$

where $f'(x) = \frac{d}{dx}f(x)$.

Proof. Solving for x the equation $y = f(x) + th(x)$, we have $x = f^{-1}(y - th(x)) = f^{-1}(y) - th(f^{-1}(y)) \frac{d}{dy}f^{-1}(y) + O(t^2)$, where $O(t^2)$ denotes terms of order t^2 or higher. Therefore,

$$T(f + th)(y) = f^{-1}(y) - th(f^{-1}(y)) \frac{1}{f'(f^{-1}(y))} + O(t^2).$$

The assertion then follows. $\square$

Lemma F.9. Given a function $f(x)$, denote by $f^{-1}(y)$ its inverse, as in Lemma F.8. Denote by T the operator that associates to a function $f(x)$ the derivative of its inverse $T(f)(y) = \frac{d}{dy}f^{-1}(y)$, and denote by $\Gamma(f)[h]$ the Gâteaux derivative of T , defined as in Lemma F.8. Then, $\Gamma(f)[h]$ takes the following form:

$$\Gamma(f)[h](y) = \frac{h(f^{-1}(y)) \frac{d}{dy}f^{-1}(y) f''(f^{-1}(y)) - h'(f^{-1}(y))}{f'(f^{-1}(y))^2},$$

where $f'(x) = \frac{d}{dx}f(x)$ and $f''(x) = \frac{d^2}{dx^2}f(x)$.

Proof. In the proof of Lemma F.8 we have shown that

$$(f + th)^{-1}(y) = f^{-1}(y) - th(f^{-1}(y)) \frac{d}{dy}f^{-1}(y) + O(t^2).$$

Therefore,

$$\begin{aligned} & T(f + th) \\ &= \frac{d}{dy}(f + th)^{-1}(y) \\ &= \frac{1}{(f' + th')((f + th)^{-1}(y))} \\ &= \left\{ f'(f^{-1}(y) - th(f^{-1}(y)) \frac{d}{dy}f^{-1}(y) + O(t^2)) \right. \\ &\quad \left. + th'(f^{-1}(y) - th(f^{-1}(y)) \frac{d}{dy}f^{-1}(y) + O(t^2)) \right\}^{-1} \\ &= \frac{1}{f'(f^{-1}(y)) - th(f^{-1}(y)) \frac{d}{dy}f^{-1}(y) f''(f^{-1}(y)) + th'(f^{-1}(y)) + O(t^2)} \\ &= \frac{1}{f'(f^{-1}(y))} \left\{ 1 - \frac{t}{f'(f^{-1}(y))} [h'(f^{-1}(y)) \right. \\ &\quad \left. - h(f^{-1}(y)) \frac{d}{dy}f^{-1}(y) f''(f^{-1}(y))] + O(t^2) \right\}. \end{aligned}$$

The assertion then follows. $\square$

Lemma F.10. *Under Assumption 6, we have*

(i)

$$\begin{aligned}\hat{\varphi}^x(u) - \varphi^x(u) &= (nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \eta^\varphi(Y_i, \delta_i, Z_i, W_i, x, u) \\ &\quad + S_n(u, x);\end{aligned}$$

where

$$\begin{aligned}\eta^\varphi(Y_i, \delta_i, Z_i, W_i, x, u) &= \Gamma(\varphi^x, F^x)^{-1} \\ &\quad \left(\sum_{l=1}^L \eta^F(Y_i, \delta_i, Z_i, W_i, \varphi(z_l, x, u), z_l, x, w_k) \right)_{k=1}^K.\end{aligned}$$

and

$$\sup_{u \in [0, \bar{U}], x \in \mathcal{X}} |S_n(u, x)| = o_p(n^{-1/2}). \quad (39)$$

(ii) Denoting by η_j^φ the j th component of η^φ , we have

$$\begin{aligned}&\frac{\partial}{\partial u} \hat{\varphi}^x(u) - \frac{\partial}{\partial u} \varphi^x(u) \\ &= (nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \left(\sum_{j=1}^L \kappa_j^l(x, u) \eta_j(Y_i, \delta_i, Z_i, W_i, x, u) \right)_{l=1}^L \\ &\quad + (nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \left(\sum_{j=1}^L \sum_{k=1}^K \zeta_{jk}^l(x, u) \right. \\ &\quad \times \left. I(W_i = w_k) \frac{I(Z_i = z_l) - p_{z_l, x, w_k}}{f_{X, W}(x, w_k)} f(\varphi(z_l, x, u) | z_l, x, w_k) \right)_{l=1}^L \\ &\quad + (nh\epsilon)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \int \tilde{K}'(s) \\ &\quad \times \left(\sum_{j=1}^L \sum_{k=1}^K \zeta_{jk}^l(x, u) \xi^F(Y_i, \delta_i, Z_i, W_i, \varphi(z_l, x, u) - s\epsilon, z_l, x, w_k) \right)_{l=1}^L ds \\ &\quad + s_n(u, x)\end{aligned}$$

for suitable functions $\kappa_j^l(x, u)$ and $\zeta_{jk}^l(x, u)$, varying $j, l = 1, \dots, L$ and $k = 1, \dots, K$, and where

$$\sup_{u \in [0, \bar{U}], x \in \mathcal{X}} |s_n(u, x)| = o_p(n^{-1/2}). \quad (40)$$

Proof. Theorem C.1 provides a representation of the quantity $\hat{F}(t, z|x, w) - F(t, z|x, w)$, and similar arguments to those used in the proof of Theorem 4.2 in [5] yield the first equality. Specifically, the first-order conditions of the optimization program (12) are

$$\Gamma(\hat{\varphi}^x, \hat{F}^x)^\top A(\hat{\varphi}^x, \hat{F}^x) = 0,$$

which leads to

$$\Gamma(\hat{\varphi}^x, \hat{F}^x)^\top A(\varphi^x, \hat{F}^x) + \Gamma(\hat{\varphi}^x, \hat{F}^x)^\top \Gamma(\varphi^x, \hat{F}^x)(\hat{\varphi}^x - \varphi^x) + \Gamma(\hat{\varphi}^x, \hat{F}^x)^\top R^x = 0,$$

where $R^x = A(\hat{\varphi}^x, \hat{F}^x) - A(\varphi^x, \hat{F}^x) - \Gamma(\varphi^x, \hat{F}^x)(\hat{\varphi}^x - \varphi^x)$. Thus, we have

$$\begin{aligned} \hat{\varphi}^x - \varphi^x &= -[\Gamma(\hat{\varphi}^x, \hat{F}^x)^\top \Gamma(\varphi^x, \hat{F}^x)]^{-1} \Gamma(\hat{\varphi}^x, \hat{F}^x)^\top A(\varphi^x, \hat{F}^x) \\ &\quad - [\Gamma(\hat{\varphi}^x, \hat{F}^x)^\top \Gamma(\varphi^x, \hat{F}^x)]^{-1} \Gamma(\hat{\varphi}^x, \hat{F}^x)^\top R^x \\ &= -\Gamma(\varphi^x, \hat{F}^x)^{-1} A(\varphi^x, \hat{F}^x) - \Gamma(\varphi^x, \hat{F}^x)^{-1} R^x. \end{aligned}$$

Now, Theorem C.1 implies that $\|A(\varphi^x, \hat{F}^x)\|_\infty = O_p((\log n/(nh))^{1/2} + h^\nu + \epsilon^\pi)$ and that $\Gamma(\varphi^x, \hat{F}^x) = \Gamma(\varphi^x, F^x) + O_p((\log n/(nh\epsilon))^{1/2} + h^\nu + \epsilon^\pi)$. Under Assumption 6, we have

$$\begin{aligned} \Gamma(\varphi^x, \hat{F}^x)^{-1} &= (\Gamma(\varphi^x, F^x) + O_p((\log n/(nh\epsilon))^{1/2} + h^\nu + \epsilon^\pi))^{-1} \\ &= (\Gamma(\varphi^x, F^x)(I + \Gamma(\varphi^x, F^x)^{-1} O_p((\log n/(nh\epsilon))^{1/2} + h^\nu + \epsilon^\pi)))^{-1} \\ &= (I + \Gamma(\varphi^x, F^x)^{-1} O_p((\log n/(nh\epsilon))^{1/2} + h^\nu + \epsilon^\pi))^{-1} \Gamma(\varphi^x, F^x)^{-1} \\ &= (I - \Gamma(\varphi^x, F^x)^{-1} O_p((\log n/(nh\epsilon))^{1/2} + h^\nu + \epsilon^\pi)) \Gamma(\varphi^x, F^x)^{-1} \\ &= \Gamma(\varphi^x, F^x)^{-1} - (\Gamma(\varphi^x, F^x)^{-1})^2 O_p((\log n/(nh\epsilon))^{1/2} + h^\nu + \epsilon^\pi) \\ &= \Gamma(\varphi^x, F^x)^{-1} + O_p((\log n/(nh\epsilon))^{1/2} + h^\nu + \epsilon^\pi) \end{aligned} \quad (41)$$

where the last equality is due to the boundedness of the involved quantities. Thus, we obtain

$$\begin{aligned} \Gamma(\varphi^x, \hat{F}^x)^{-1} A(\varphi^x, \hat{F}^x) &= [\Gamma(\varphi^x, \hat{F}^x)^{-1} - \Gamma(\varphi^x, F^x)^{-1}] A(\varphi^x, \hat{F}^x) + \Gamma(\varphi^x, F^x)^{-1} A(\varphi^x, \hat{F}^x) \\ &= O_p((\log n/(nh\epsilon))^{1/2} + h^\nu + \epsilon^\pi) O_p((\log n/(nh))^{1/2} + h^\nu + \epsilon^\pi) \\ &\quad + \Gamma(\varphi^x, F^x)^{-1} A(\varphi^x, \hat{F}^x) \\ &= \Gamma(\varphi^x, F^x)^{-1} A(\varphi^x, \hat{F}^x) + o_p(n^{-1/2}) \end{aligned}$$

The proof of Theorem 5.1 implies that $R^x = o_p(n^{-1/2})$. Since, for every $k = 1, \dots, K$, it holds $u = \sum_{l=1}^L F^x(\varphi_l^x(u), z_l|w_k)$, we obtain

$$\begin{aligned} \hat{\varphi}^x(u) - \varphi^x(u) &= \Gamma(\varphi^x, F^x)^{-1} A(\varphi^x, \hat{F}^x) + o_p(n^{-1/2}) \end{aligned}$$

$$\begin{aligned}
&= \Gamma(\varphi^x, F^x)^{-1} \left(\sum_{l=1}^L \hat{F}^x(\varphi_l^x(u), z_l|w_k) - F^x(\varphi_l^x(u), z_l|w_k) \right)_{k=1}^K + o_p(n^{-1/2}) \\
&= (nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \Gamma(\varphi^x, F^x)^{-1} \\
&\quad \left(\sum_{l=1}^L \eta^F(Y_i, \delta_i, Z_i, W_i, \varphi(z_l, x, u), z_l, x, w_k) \right)_{k=1}^K + o_p(n^{-1/2}).
\end{aligned}$$

We now show the result in (ii). To simplify the argument we assume $L = K = 2$, but it will be clear that the result holds for any integer $L \geq 1$. Denote by 1_L the vector in $\mathbb{R}^L$ with all entrances equals to 1. We have shown in equation (29) that

$$\begin{aligned}
\frac{\partial}{\partial u} \hat{\varphi}^x - \frac{\partial}{\partial u} \varphi^x &= \left(\Gamma(\hat{\varphi}^x, \hat{F}^x)^{-1} - \Gamma(\varphi^x, F^x)^{-1} \right) 1_K \\
&= \left(\frac{1}{\hat{f}^x(\hat{\varphi}_1^x, z_1|w_1) \hat{f}^x(\hat{\varphi}_2^x, z_2|w_2) - \hat{f}^x(\hat{\varphi}_2^x, z_2|w_1) \hat{f}^x(\hat{\varphi}_1^x, z_1|w_2)} \right. \\
&\quad \times \begin{bmatrix} \hat{f}^x(\hat{\varphi}_2^x, z_2|w_2) & -\hat{f}^x(\hat{\varphi}_2^x, z_2|w_1) \\ -\hat{f}^x(\hat{\varphi}_1^x, z_1|w_2) & \hat{f}^x(\hat{\varphi}_1^x, z_1|w_1) \end{bmatrix} \\
&\quad \left. - \frac{1}{f^x(\varphi_1^x, z_1|w_1) f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1) f^x(\varphi_1^x, z_1|w_2)} \right. \\
&\quad \times \begin{bmatrix} f^x(\varphi_2^x, z_2|w_2) & -f^x(\varphi_2^x, z_2|w_1) \\ -f^x(\varphi_1^x, z_1|w_2) & f^x(\varphi_1^x, z_1|w_1) \end{bmatrix} \left. \right) 1_K
\end{aligned}$$

We now show that the aimed iid representation holds for the first component of $\frac{\partial}{\partial u} \hat{\varphi}^x - \frac{\partial}{\partial u} \varphi^x$, that is $\frac{\partial}{\partial u} \hat{\varphi}_1^x - \frac{\partial}{\partial u} \varphi_1^x$, and clearly the same argument can be used for the second component. So, write

$$\begin{aligned}
\frac{\partial}{\partial u} \hat{\varphi}_1^x - \frac{\partial}{\partial u} \varphi_1^x &= \frac{\hat{f}^x(\hat{\varphi}_2^x, z_2|w_2) - \hat{f}^x(\hat{\varphi}_2^x, z_2|w_1)}{\hat{f}^x(\hat{\varphi}_1^x, z_1|w_1) \hat{f}^x(\hat{\varphi}_2^x, z_2|w_2) - \hat{f}^x(\hat{\varphi}_2^x, z_2|w_1) \hat{f}^x(\hat{\varphi}_1^x, z_1|w_2)} \\
&\quad - \frac{f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1)}{f^x(\varphi_1^x, z_1|w_1) f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1) f^x(\varphi_1^x, z_1|w_2)} \\
&= I_1 + I_2,
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= \frac{\hat{f}^x(\hat{\varphi}_2^x, z_2|w_2) - \hat{f}^x(\hat{\varphi}_2^x, z_2|w_1)}{\hat{f}^x(\hat{\varphi}_1^x, z_1|w_1) \hat{f}^x(\hat{\varphi}_2^x, z_2|w_2) - \hat{f}^x(\hat{\varphi}_2^x, z_2|w_1) \hat{f}^x(\hat{\varphi}_1^x, z_1|w_2)} \\
&\quad - \frac{\hat{f}^x(\varphi_2^x, z_2|w_2) - \hat{f}^x(\varphi_2^x, z_2|w_1)}{\hat{f}^x(\varphi_1^x, z_1|w_1) \hat{f}^x(\varphi_2^x, z_2|w_2) - \hat{f}^x(\varphi_2^x, z_2|w_1) \hat{f}^x(\varphi_1^x, z_1|w_2)}; \\
I_2 &= \frac{\hat{f}^x(\varphi_2^x, z_2|w_2) - \hat{f}^x(\varphi_2^x, z_2|w_1)}{\hat{f}^x(\varphi_1^x, z_1|w_1) \hat{f}^x(\varphi_2^x, z_2|w_2) - \hat{f}^x(\varphi_2^x, z_2|w_1) \hat{f}^x(\varphi_1^x, z_1|w_2)}
\end{aligned}$$

$$- \frac{f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1)}{f^x(\varphi_1^x, z_1|w_1)f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1)f^x(\varphi_1^x, z_1|w_2)}.$$

We first start with I_1 , and using standard Taylor expansion, we can write $I_1 = I_{11} + I_{12}$, where

$$\begin{aligned} I_{11} &= \left((\hat{f}^x(\varphi_1^x, z_1|w_1)\hat{f}^x(\varphi_2^x, z_2|w_2) - \hat{f}^x(\varphi_2^x, z_2|w_1)\hat{f}^x(\varphi_1^x, z_1|w_2))^{-2} \right. \\ &\quad \times \left(\frac{d}{dt}\hat{f}^x(\varphi_2^x, z_2|w_2) - \frac{d}{dt}\hat{f}^x(\varphi_2^x, z_2|w_1) \right) \\ &\quad \times (\hat{f}^x(\varphi_1^x, z_1|w_1)\hat{f}^x(\varphi_2^x, z_2|w_2) - \hat{f}^x(\varphi_2^x, z_2|w_1)\hat{f}^x(\varphi_1^x, z_1|w_2)) \\ &\quad - (\hat{f}^x(\varphi_1^x, z_1|w_1)\hat{f}^x(\varphi_2^x, z_2|w_2) - \hat{f}^x(\varphi_2^x, z_2|w_1)\hat{f}^x(\varphi_1^x, z_1|w_2))^{-2} \\ &\quad \times (\hat{f}^x(\varphi_2^x, z_2|w_2) - \hat{f}^x(\varphi_2^x, z_2|w_1)) \\ &\quad \times (\hat{f}^x(\varphi_1^x, z_1|w_1)\frac{d}{dt}\hat{f}^x(\varphi_2^x, z_2|w_2) - \frac{d}{dt}\hat{f}^x(\varphi_2^x, z_2|w_1)\hat{f}^x(\varphi_1^x, z_1|w_2)) \Big) \\ &\quad \times (\hat{\varphi}_2^x - \varphi_2^x) + O_p(\|\hat{\varphi}_2^x - \varphi_2^x\|_\infty^2); \\ I_{12} &= -(\hat{f}^x(\varphi_1^x, z_1|w_1)\hat{f}^x(\varphi_2^x, z_2|w_2) - \hat{f}^x(\varphi_2^x, z_2|w_1)\hat{f}^x(\varphi_1^x, z_1|w_2))^{-2} \\ &\quad \times (\hat{f}^x(\varphi_2^x, z_2|w_2) - \hat{f}^x(\varphi_2^x, z_2|w_1)) \\ &\quad \times \left(\frac{d}{dt}\hat{f}^x(\varphi_1^x, z_1|w_1)\hat{f}^x(\varphi_2^x, z_2|w_2) - \hat{f}^x(\varphi_2^x, z_2|w_1)\frac{d}{dt}\hat{f}^x(\varphi_1^x, z_1|w_2) \right) \\ &\quad \times (\hat{\varphi}_1^x - \varphi_1^x) + O_p(\|\hat{\varphi}_1^x - \varphi_1^x\|_\infty^2) \end{aligned}$$

Now, $I_{12} = I_{121} + I_{122}$, with

$$\begin{aligned} I_{121} &= \left(-(\hat{f}^x(\varphi_1^x, z_1|w_1)\hat{f}^x(\varphi_2^x, z_2|w_2) - \hat{f}^x(\varphi_2^x, z_2|w_1)\hat{f}^x(\varphi_1^x, z_1|w_2))^{-2} \right. \\ &\quad \times (\hat{f}^x(\varphi_2^x, z_2|w_2) - \hat{f}^x(\varphi_2^x, z_2|w_1)) \\ &\quad \times \left(\frac{d}{dt}\hat{f}^x(\varphi_1^x, z_1|w_1)\hat{f}^x(\varphi_2^x, z_2|w_2) - \hat{f}^x(\varphi_2^x, z_2|w_1)\frac{d}{dt}\hat{f}^x(\varphi_1^x, z_1|w_2) \right) \\ &\quad + (f^x(\varphi_1^x, z_1|w_1)f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1)f^x(\varphi_1^x, z_1|w_2))^{-2} \\ &\quad \times (f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1)) \\ &\quad \times \left(\frac{d}{dt}f^x(\varphi_1^x, z_1|w_1)f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1)\frac{d}{dt}f^x(\varphi_1^x, z_1|w_2) \right) \Big) \\ &\quad \times (\hat{\varphi}_1^x - \varphi_1^x); \\ I_{122} &= -(f^x(\varphi_1^x, z_1|w_1)f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1)f^x(\varphi_1^x, z_1|w_2))^{-2} \\ &\quad \times (f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1)) \\ &\quad \times \left(\frac{d}{dt}f^x(\varphi_1^x, z_1|w_1)f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1)\frac{d}{dt}f^x(\varphi_1^x, z_1|w_2) \right) \\ &\quad \times (\hat{\varphi}_1^x - \varphi_1^x) + O_p(\|\hat{\varphi}_1^x - \varphi_1^x\|_\infty^2). \end{aligned}$$

It is easy to show that $I_{121} = o_p(n^{-1/2})$ because $\frac{\partial}{\partial t}\hat{f}^x$ is bounded with probability tending to 1, $\hat{f}^x - f^x = O_p(n^{-1/4})$ and $\hat{\varphi}_1^x - \varphi_1^x = o_p(n^{-1/4})$. Denote by

η_l^φ the l th functional component of η^φ for $l = 1, \dots, L$. It follows that

$$\begin{aligned} I_{12} &= o_p(n^{-1/2}) + I_{122} \\ &= -\left(f^x(\varphi_1^x, z_1|w_1)f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1)f^x(\varphi_1^x, z_1|w_2)\right)^{-2} \\ &\quad \times \left(f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1)\right) \\ &\quad \times \left(\frac{d}{dt}f^x(\varphi_1^x, z_1|w_1)f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1)\frac{d}{dt}f^x(\varphi_1^x, z_1|w_2)\right) \\ &\quad \times (nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \eta_1^\varphi(Y_i, \delta_i, Z_i, W_i, x, u) + o_p(n^{-1/2}). \end{aligned}$$

A similar argument can be used for I_{11} .

We now focus on I_2 , which, by standard decomposition argument, can be written as $I_2 = I_{21} + I_{22} + o_p(n^{-1/2})$, where

$$\begin{aligned} I_{21} &= \frac{f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1) - \hat{f}^x(\varphi_2^x, z_2|w_2) - \hat{f}^x(\varphi_2^x, z_2|w_1)}{f^x(\varphi_1^x, z_1|w_1)f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1)f^x(\varphi_1^x, z_1|w_2)} \\ I_{22} &= \frac{(f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1))}{\left(f^x(\varphi_1^x, z_1|w_1)f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1)f^x(\varphi_1^x, z_1|w_2)\right)^2} \\ &\quad \times \left(f^x(\varphi_1^x, z_1|w_1)f^x(\varphi_2^x, z_2|w_2) - f^x(\varphi_2^x, z_2|w_1)f^x(\varphi_1^x, z_1|w_2)\right. \\ &\quad \left.- \hat{f}^x(\varphi_1^x, z_1|w_1)\hat{f}^x(\varphi_2^x, z_2|w_2) + \hat{f}^x(\varphi_2^x, z_2|w_1)\hat{f}^x(\varphi_1^x, z_1|w_2)\right) \end{aligned}$$

This equality and standard decomposition arguments similar to the ones used before, imply that it is possible to use the iid representation of $\hat{f}^x - f^x$ to obtain a representation for I_{21} and I_{22} . This means that there exist four functions $\zeta_{lk}^0(x, u)$ for $l, k = 1, 2$ such that

$$\begin{aligned} I_2 &= \sum_{l,k=1}^2 \zeta_{lk}^0(x, u) \times \left((nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)\right. \\ &\quad \left.I(W_i = w_k) \frac{I(Z_i = z_l) - p_{z_l, x, w_k}}{f_{X,W}(x, w_k)} f(\varphi(z_l, x, u)|z_l, x, w_k)\right) \\ &\quad + (nh\epsilon)^{-1} \sum_{i=1}^n p_{z_l, x, w_k} K\left(\frac{x - X_i}{h}\right) \\ &\quad \times \int \tilde{K}(s) \xi^F(Y_i, \delta_i, Z_i, W_i, \varphi(z_l, x, u) - s\epsilon, z_l, x, w_k) ds \\ &\quad + o_p(n^{-1/2}). \end{aligned}$$

To summarize, there exist some differentiable functions $\kappa_j^l(x, u)$ and $\zeta_{jk}^l(x, u)$, for $j, l = 1, \dots, L$, $k = 1, \dots, K$ such that

$$\frac{\partial}{\partial u} \hat{\varphi}^x(u) - \frac{\partial}{\partial u} \varphi^x(u)$$

$$\begin{aligned}
&= (nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \left(\sum_{j=1}^L \kappa_j^l(x, u) \eta_j(Y_i, \delta_i, Z_i, W_i, x, u) \right)_{l=1}^L \\
&+ (nh)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \left(\sum_{j=1}^L \sum_{k=1}^K \zeta_{jk}^l(x, u) I(W_i = w_k) \right. \\
&\quad \times \left. \frac{I(Z_i = z_l) - p_{z_l, x, w_k}}{f_{X, W}(x, w_k)} f(\varphi(z_l, x, u) | z_l, x, w_k) \right)_{l=1}^L \\
&+ (nh\epsilon)^{-1} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) \\
&\quad \times \int \tilde{K}'(s) \left(\sum_{j=1}^L \sum_{k=1}^K \zeta_{jk}^l(x, u) \xi^F(Y_i, \delta_i, Z_i, W_i, \varphi(z_l, x, u) - s\epsilon, z_l, x, w_k) \right)_{l=1}^L ds \\
&+ o_p(n^{-1/2}).
\end{aligned}$$

This corresponds to the result in (ii). $\square$

G. Simpler identification conditions in randomized experiments with non-compliance

In this section, we provide the reader with conditions under which (5) has a unique solution similar to the ones discussed in [12] or in [7]. We focus on the simpler case where both the treatment Z and the instrument W are binary in the presence of a randomized experiment with non-compliance, a scenario that frequently arises in empirical applications.

We consider $X = x$ and $u \in [0, 1]$ to be fixed. Let $f_{T|Z, W}^x(t|z, w)$ be the density of T given $Z = z, W = w, X = x$ and $\nu, \underline{f} > 0$ be small constants. We define the set $\mathcal{L}$ as the closed rectangle of vectors $(t_0, t_1) \in \mathbb{R}^2$ satisfying the following conditions:

- (i) $\mathbb{P}(T \leq t_Z | W = w_1, X = x) \in [u - \nu, u + \nu]$ for each $w_1 \in \{0, 1\}$;
- (ii) $f_{T|Z, W}^x(t|z, w) > \underline{f}$ is bounded from below for all $z, w \in \{0, 1\}$ such that $\mathbb{P}(Z = z | W = w, X = x) > 0$.

We use the notation $t_Z = (1 - Z)t_0 + Zt_1$. We want to show that there exists a unique $t = (t_0, t_1)^\top \in \mathcal{L}$ such that $\Pi^x(t) = 0$, where

$$\Pi^x(t) = \begin{bmatrix} \mathbb{P}(T \leq t_Z | W = 0, X = x) - u \\ \mathbb{P}(T \leq t_Z | W = 1, X = x) - u \end{bmatrix}.$$

The Jacobian of $\Pi^x(t)$ with respect to t , denoted by $\Pi^x(t)'$, takes the following form:

$$\Pi^x(t)' = \begin{bmatrix} f^x(t_0, 0|0) & f^x(t_1, 1|0) \\ f^x(t_0, 0|1) & f^x(t_1, 1|1) \end{bmatrix}.$$

where $f^x(t, z|w) = f_{T|Z,W}^x(t|z, w)\mathbb{P}(Z = z|W = w, X = x)$. We make the following Assumption:

- (B) (i) $f_{T|Z,W}^x(\varphi(z, x, u)|z, w) > \underline{f}$ for all $z, w \in \{0, 1\}, x \in \mathcal{X}$ such that $\mathbb{P}(Z = z|W = w, X = x) > 0$;
(ii) $\Pi^x(t)'$ has full rank for all $t \in \mathcal{L}, x \in \mathcal{X}$.

Assumption (B) guarantees that $(\varphi(0, x, u), \varphi(1, x, u))^\top$ belongs to $\mathcal{L}$ and that the system in (5) has a unique solution. We refer to Theorem 3.2 of [7] or Section 2.4 of [12] for a detailed proof.

Assumption (B)(i) means that for $X = x$ and for all populations $Z = z, W = w$ that appear in the data, there are some subjects with $U = u$ in this population. Assumption (B)(ii) is a full rank condition, which is standard in econometrics. Similarly as in [12], we can provide the following interpretation of (B)(ii). The matrix $\Pi^x(t)'$ has full rank for all $t \in \mathcal{L}, x \in \mathcal{X}$ if and only if

$$\frac{f^x(t_0, 0|0)}{f^x(t_0, 0|1)} > \frac{f^x(t_1, 1|0)}{f^x(t_1, 1|1)}, \text{ for all } t \in \mathcal{L}, x \in \mathcal{X} \quad (42)$$

(or the same property with $<$ instead of $>$). This inequality can be interpreted as a monotone likelihood ratio condition: for all $t \in \mathcal{L}, x \in \mathcal{X}$, the instrument increases (or decreases) the likelihood ratio in (42).

Let us also consider the special case of randomized experiments with one-sided non-compliance where $\mathbb{P}(Z = 1|W = 0, X = x) = 0$. Notice that

$$\begin{aligned} F^x(s, z|w) &= \mathbb{P}(T \leq s|Z = z, W = w)\mathbb{P}(Z = z|W = w, X = x) \\ &= \mathbb{P}(U \leq \varphi_z^x(\cdot)^{-1}(s)|Z = z, X = x, W = w)\mathbb{P}(Z = z|W = w, X = x). \end{aligned}$$

By the implicit function theorem and the chain rule this yields

$$f^x(s, z|w) = \frac{\mathbb{P}(Z = z|W = w, X = x)}{(\varphi_z^x)'(\varphi_z^x(\cdot)^{-1}(s))} f_{U|Z,W,X}(\varphi_z^x(\cdot)^{-1}(s)|z, w, x),$$

where $f_{U|Z,W,X}(u|z, w, x)$ is the density of the distribution of U given $Z = z, W = w, X = x$ evaluated at u . By the chain rule and the fact that U follows a uniform distribution given W, X , we obtain

$$f^x(s, z|w) = \frac{1}{(\varphi_z^x)'(\varphi_z^x(\cdot)^{-1}(s))} \mathbb{P}(Z = z|W = w, X = x, U = \varphi_z^x(\cdot)^{-1}(s)).$$

Hence, it can be seen that under one-sided non-compliance, Assumption (B) holds as long as $\mathbb{P}(Z = 1|W = w, X = x, U = v) > \xi$ for some $\xi > 0$, for all $v \in [0, 1], x \in \mathcal{X}$. This condition means that every subject has at least a probability ξ of being treated $Z = 1$ if they are assigned to the treatment group $W = 1$.

H. Further simulations

H.1. Choice of the kernel and bandwidth

In this section, we conduct the same simulation as discussed in Section 6, varying the choice of kernel and bandwidth in the smoothing procedure of the continuous covariate X for the estimator $\hat{F}(t, z | x, w)$. Specifically, we explore two different bandwidth choices: the second-generation Direct Plug-in implementation of MSE-optimal bandwidth (mse-dpi), which is the default method in the R package `nprobust`, and the second-generation Coverage Error optimal bandwidth (ce-dpi). We also consider two different kernels: the Epanechnikov kernel and the uniform kernel.

The results for the Epanechnikov kernel with mse-dpi bandwidth choice are reported in the main text in Table 2. The results for the other combinations—uniform kernel with mse-dpi, Epanechnikov kernel with ce-dpi, and uniform kernel with ce-dpi—are presented in Tables 5, 6, and 7, respectively. The results across these different combinations are comparable.

TABLE 5

Simulation results for Beta and continuous design, with uniform kernel and mse-dpi choice of bandwidth. We include for the proposed estimator and the standard partial likelihood estimator, for each level of censoring (Cens) and for different sample sizes (n), the average of the bias (Bias), the standard deviation (Sd) and the Mean Squared Error (MSE) for each component (Comp) of $\beta = (\beta_z, \beta_x)^\top$. For each case, we also include the Root Mean Squared Error value (RMSE). For the proposed estimator, we also provide the coverage of the 95% confidence intervals (CP95) for each component of β .

Design	Cens	n	Comp	proposed estimator					partial likelihood estimator			
				Bias	Sd	MSE	RMSE	CP95	Bias	Sd	MSE	RMSE
continuous Beta	20%	500	β_z	-0.100	0.262	0.078	0.504	0.928	-0.265	0.155	0.094	0.439
			β_x	0.016	0.654	0.428		0.973	0.103	0.451	0.214	
		1000	β_z	-0.067	0.197	0.043	0.353	0.919	-0.262	0.108	0.080	0.367
			β_x	-0.010	0.446	0.199		0.970	0.087	0.327	0.115	
	40%	500	β_z	-0.043	0.300	0.092	0.657	0.958	-0.295	0.182	0.120	0.509
			β_x	0.052	0.856	0.735		0.962	0.123	0.525	0.291	
		1000	β_z	-0.044	0.217	0.049	0.448	0.959	-0.295	0.127	0.103	0.421
			β_x	0.009	0.580	0.336		0.971	0.115	0.371	0.151	
continuous uniform	20%	500	β_z	-0.073	0.270	0.078	0.324	0.962	-0.230	0.162	0.079	0.326
			β_x	-0.005	0.368	0.136		0.974	0.084	0.256	0.073	
		1000	β_z	-0.055	0.185	0.037	0.234	0.950	-0.231	0.110	0.066	0.288
			β_x	0.014	0.255	0.065		0.968	0.081	0.181	0.039	
	40%	500	β_z	-0.031	0.307	0.095	0.441	0.980	-0.257	0.185	0.100	0.373
			β_x	-0.080	0.500	0.256		0.940	0.102	0.297	0.099	
		1000	β_z	-0.034	0.196	0.040	0.300	0.960	-0.262	0.125	0.084	0.328
			β_x	-0.057	0.365	0.136		0.978	0.103	0.204	0.052	

TABLE 6

Simulation results for Beta and continuous design, with Epanechnikov kernel and ce-dpi choice of bandwidth. We include for the proposed estimator and the standard partial likelihood estimator, for each level of censoring (Cens) and for different sample sizes (n), the average of the bias (Bias), the standard deviation (Sd) and the Mean Squared Error (MSE) for each component (Comp) of $\beta = (\beta_z, \beta_x)^\top$. For each case, we also include the Root Mean Squared Error value (RMSE). For the proposed estimator, we also provide the coverage of the 95% confidence intervals (CP95) for each component of β .

Design	Cens	n	Comp	proposed estimator					partial likelihood estimator			
				Bias	Sd	MSE	RMSE	CP95	Bias	Sd	MSE	RMSE
continuous Beta	20%	500	β_z	-0.095	0.269	0.081	0.506	0.939	-0.265	0.155	0.094	0.439
			β_x	0.019	0.663	0.439		0.969	0.103	0.451	0.214	
		1000	β_z	-0.070	0.198	0.044	0.360	0.918	-0.262	0.108	0.080	0.367
			β_x	-0.005	0.446	0.199		0.967	0.087	0.327	0.115	
	40%	500	β_z	-0.045	0.306	0.096	0.690	0.959	-0.295	0.182	0.120	0.509
			β_x	0.073	0.873	0.768		0.943	0.123	0.525	0.291	
		1000	β_z	-0.047	0.218	0.050	0.463	0.957	-0.295	0.127	0.103	0.421
			β_x	0.036	0.594	0.355		0.963	0.115	0.371	0.151	
continuous uniform	20%	500	β_z	-0.074	0.270	0.078	0.323	0.950	-0.230	0.162	0.079	0.326
			β_x	0.006	0.361	0.130		0.974	0.084	0.256	0.073	
		1000	β_z	-0.062	0.180	0.036	0.234	0.940	-0.231	0.110	0.066	0.288
			β_x	0.020	0.257	0.066		0.968	0.081	0.181	0.039	
	40%	500	β_z	-0.037	0.312	0.099	0.458	0.974	-0.257	0.185	0.100	0.373
			β_x	-0.063	0.516	0.271		0.934	0.102	0.297	0.099	
		1000	β_z	-0.038	0.198	0.041	0.310	0.964	-0.262	0.125	0.084	0.328
			β_x	-0.052	0.386	0.152		0.980	0.103	0.204	0.052	

TABLE 7

Simulation results for Beta and continuous design, with uniform kernel and ce-dpi choice of bandwidth. We include for the proposed estimator and the standard partial likelihood estimator, for each level of censoring (Cens) and for different sample sizes (n), the average of the bias (Bias), the standard deviation (Sd) and the Mean Squared Error (MSE) for each component (Comp) of $\beta = (\beta_z, \beta_x)^\top$. For each case, we also include the Root Mean Squared Error value (RMSE). For the proposed estimator, we also provide the coverage of the 95% confidence intervals (CP95) for each component of β .

Design	Cens	n	Comp	proposed estimator					partial likelihood estimator			
				Bias	Sd	MSE	RMSE	CP95	Bias	Sd	MSE	RMSE
continuous Beta	20%	500	β_z	-0.098	0.270	0.082	0.511	0.945	-0.265	0.155	0.094	0.439
			β_x	0.022	0.661	0.438		0.971	0.103	0.451	0.214	
		1000	β_z	-0.074	0.200	0.045	0.364	0.919	-0.262	0.108	0.080	0.367
			β_x	-0.002	0.447	0.199		0.968	0.087	0.327	0.115	
	40%	500	β_z	-0.040	0.304	0.094	0.695	0.962	-0.295	0.182	0.120	0.509
			β_x	0.093	0.877	0.777		0.954	0.123	0.525	0.291	
		1000	β_z	-0.044	0.219	0.050	0.463	0.963	-0.295	0.127	0.103	0.421
			β_x	0.052	0.594	0.356		0.961	0.115	0.371	0.151	
continuous uniform	20%	500	β_z	-0.077	0.268	0.078	0.325	0.948	-0.230	0.162	0.079	0.326
			β_x	0.003	0.366	0.134		0.970	0.084	0.256	0.073	
		1000	β_z	-0.067	0.180	0.037	0.235	0.938	-0.231	0.110	0.066	0.288
			β_x	0.019	0.258	0.067		0.966	0.081	0.181	0.039	
	40%	500	β_z	-0.031	0.306	0.095	0.455	0.978	-0.257	0.185	0.100	0.373
			β_x	-0.068	0.520	0.274		0.934	0.102	0.297	0.099	
		1000	β_z	-0.037	0.199	0.041	0.311	0.972	-0.262	0.125	0.084	0.328
			β_x	-0.056	0.384	0.151		0.972	0.103	0.204	0.052	

H.2. An example of higher dimensional covariate

This section presents an additional design that includes two continuous covariates. The table below provides detailed specifications for this design, similar to those in the main text. The specification of the simulation remains unchanged, including the baseline cumulative hazard function Λ_0 , and the number of replications $N = 5000$. Thus, we refer to the main text for these details.

Regarding the value of the rate λ of the exponential distribution of the censoring variable, we set $\lambda = 0.41, 1.14$ to achieve 20% and 40% censoring rates, respectively.

The results of this additional design are in line with the previous simulations, indicating that, differently from the partial likelihood estimator, the proposed estimator presents unbiased results and maintains good coverage probability. This consistency across different designs and censoring levels further validates the robustness of our estimator, especially in complex scenarios involving multiple continuous covariates. The detailed outcomes of this extended analysis substantiate the estimator's applicability and effectiveness in a broader range of statistical modeling contexts.

TABLE 8

Specifications for the additional design with two continuous covariates. Here, $\mathcal{U}(a, b)$ represents the uniform distribution over the interval $[a, b]$. The expression for Z is structured to reflect the relationship with the two covariates and the instrumental variable W . The coefficients $(\beta_{0,z}, \beta_{0,x_1}, \beta_{0,x_2})$ correspond to the impact of Z , X_1 , and X_2 respectively.

Design	W	X_1	X_2	Z	$(\beta_{0,z}, \beta_{0,x_1}, \beta_{0,x_2})$
Continuous uniform-uniform	$B(0.5)$	$\mathcal{U}(0, 1)$	$\mathcal{U}(0, 1)$	$I(W = 1)I(1 - W + 0.5U + 0.5\epsilon \geq 0)$	$(0.9, 0.2, 0.2)$

TABLE 9

Simulation results for design. We include for the proposed estimator and the standard partial likelihood estimator, for each level of censoring (Cens) and for different sample sizes (n), the average of the bias (Bias), the standard deviation (Sd) and the Mean Squared Error (MSE) for each component (Comp) of $\beta = (\beta_z, \beta_{x_1}, \beta_{x_2})^\top$. For each case, we also include the Root Mean Squared Error value (RMSE). For the proposed estimator, we also provide the coverage of the 95% confidence intervals (CP95) for each component of β .

Design	Cens	n	Comp	proposed estimator					partial likelihood estimator			
				Bias	Sd	MSE	RMSE	CP95	Bias	Sd	MSE	RMSE
Design Continuous uniform-uniform	20%	500	β_z	-0.068	0.241	0.063		0.945	-0.199	0.149	0.062	
			β_{x_1}	-0.048	0.334	0.114	0.388	0.958	0.003	0.245	0.060	0.335
			β_{x_2}	-0.053	0.339	0.118		0.967	-0.000	0.253	0.064	
		1000	β_z	-0.052	0.177	0.034		0.937	-0.200	0.106	0.051	
			β_{x_1}	-0.039	0.233	0.056	0.275	0.961	-0.001	0.176	0.031	0.275
			β_{x_2}	-0.042	0.239	0.059		0.962	0.001	0.173	0.030	
	40%	500	β_z	-0.025	0.270	0.073		0.965	-0.222	0.169	0.078	
			β_{x_1}	-0.086	0.472	0.230	0.541	0.955	0.002	0.288	0.083	0.384
			β_{x_2}	-0.091	0.462	0.221		0.951	-0.002	0.292	0.085	
		1000	β_z	-0.031	0.196	0.039		0.963	-0.222	0.120	0.064	
			β_{x_1}	-0.072	0.348	0.126	0.401	0.950	0.002	0.205	0.042	0.311
			β_{x_2}	-0.077	0.354	0.131		0.950	0.004	0.203	0.041	

Acknowledgments

The authors thank Gerda Claeskens, Elia Lapenta and Juan Carlos Pardo-Fernández for comments that improved the paper. Jad Beyhum undertook most

of this work while employed by CREST, ENSAI.

Funding

Ingrid Van Keilegom acknowledges support from the FWO and F.R.S.-FNRS under the Excellence of Science (EOS) programme, project ASTeRISK (grant No. 40007517).

Lorenzo Tedesco acknowledges support from the National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.3, under Call for Tender No. 341 of 15/03/2022, funded by the Italian Ministry of University and Research and the European Union's NextGenerationEU initiative. The project was awarded under Grant No. PE 00000018, Concession Decree No. 1558 of 11/10/2022, CUP F83C22001720001, as part of the Growing Resilient Inclusive and Sustainable (GRINS) initiative.

References

- [1] Akritas, M. On the use of nonparametric regression techniques for fitting parametric regression models. *Biometrics*. pp. 1342-1362 (1996) [MR1422081](#)
- [2] Akritas, M. & Van Keilegom, I. Non-parametric estimation of the residual distribution. *Scandinavian Journal Of Statistics*. **28**, 549-567 (2001) [MR1858417](#)
- [3] Angrist, J., Imbens, G. & Rubin, D. Identification of causal effects using instrumental variables. *Journal Of The American Statistical Association*. **91**, 444-455 (1996)
- [4] Beran, R. Nonparametric regression with randomly censored survival data. (technical report, University of California, Berkeley, 1981)
- [5] Beyhum, J., Florens, J. & Van Keilegom, I. Nonparametric instrumental regression with right censored duration outcomes. *Journal Of Business & Economic Statistics*. **40**, 1034-1045 (2022) [MR4439270](#)
- [6] Beyhum, J., Centorrino, S., Florens, J. & Van Keilegom, I. Instrumental variable estimation of dynamic treatment effects on a duration outcome. *Journal Of Business & Economic Statistics*. **42**, 732-742 (2024) [MR4729003](#)
- [7] Beyhum, J., Tedesco, L. & Van Keilegom, I. Instrumental variable quantile regression under random right censoring. *The Econometrics Journal*. **27**, 21-36 (2024) [MR4712657](#)
- [8] Bijwaard, G. & Ridder, G. Correcting for selective compliance in a re-employment bonus experiment. *Journal Of Econometrics*. **125**, 77-111 (2005) [MR2143373](#)
- [9] Calonico, S., Cattaneo, M. & Farrell, M. nprobust: Nonparametric kernel-based estimation and robust bias-corrected inference. *ArXiv Preprint arXiv:1906.00198*. (2019)
- [10] Chen, X., Linton, O. & Van Keilegom, I. Estimation of semiparametric models when the criterion function is not smooth. *Econometrica*. **71**, 1591-1608 (2003) [MR2000259](#)

- [11] Chen, S. Sequential estimation of censored quantile regression models. *Journal Of Econometrics*. **207**, 30-52 (2018) [MR3856759](#)
- [12] Chernozhukov, V. & Hansen, C. An IV model of quantile treatment effects. *Econometrica*. **73**, 245-261 (2005) [MR2115636](#)
- [13] Chernozhukov, V. & Hansen, C. Instrumental quantile regression inference for structural and treatment effect models. *Journal Of Econometrics*. **132**, 491-525 (2006) [MR2323990](#)
- [14] Cox, D. Regression models and life-tables. *Journal Of The Royal Statistical Society: Series B (Methodological)*. **34**, 187-202 (1972) [MR0341758](#)
- [15] Cristobal, J., Roca, P. & Manteiga, W. A class of linear regression parameter estimators constructed by nonparametric estimation. *The Annals Of Statistics*. **15**, 603-609 (1987) [MR0888428](#)
- [16] De Backer, M., El Ghouch, A. & Van Keilegom, I. Linear censored quantile regression: A novel minimum-distance approach. *Scandinavian Journal Of Statistics*. **47**, 1275-1306 (2020) [MR4178194](#)
- [17] Giacomini, R., Politis, D. & White, H. A warp-speed method for conducting Monte Carlo experiments involving bootstrap estimators. *Econometric Theory*. **29**, 567-589 (2013) [MR3064050](#)
- [18] Hall, P., Wolff, R. & Yao, Q. Methods for estimating a conditional distribution function. *Journal Of The American Statistical Association*. **94**, 154-163 (1999) [MR1689221](#)
- [19] Hong, H. & Tamer, E. Inference in censored models with endogenous regressors. *Econometrica*. **71**, 905-932 (2003) [MR1983231](#)
- [20] Kaido, H. & Wüthrich, K. Decentralization estimators for instrumental variable quantile regression models. *Quantitative Economics*. **12**, 443-475 (2021) [MR4325591](#)
- [21] Kaplan, D. & Sun, Y. Smoothed estimating equations for instrumental variables quantile regression. *Econometric Theory*. **33**, 105-157 (2017) [MR3574862](#)
- [22] Koenker, R., Chernozhukov, V., He, X. & Peng, L. Handbook of quantile regression.(2017). [MR3728340](#)
- [23] Musta, E., Patilea, V. & Van Keilegom, I. A presmoothing approach for estimation in the semiparametric Cox mixture cure model. *Bernoulli*. **28**, 2689-2715 (2022) [MR4474559](#)
- [24] Tsiatis, A. A large sample study of Cox's regression model. *The Annals Of Statistics*. **9**, 93-108 (1981) [MR0600535](#)
- [25] Vaart, A. & Wellner, J. Weak Convergence and Empirical Processes: With Applications to Statistics. (Springer Cham, 2023) [MR4628026](#)
- [26] Van Keilegom, I. Nonparametric estimation of the conditional distribution in regression with censored data. (UHasselt Diepenbeek, 1998)
- [27] Van Keilegom, I. & Veraverbeke, N. Uniform strong convergence results for the conditional Kaplan-Meier estimator and its quantiles. *Communications In Statistics-Theory And Methods*. **25**, 2251-2265 (1996) [MR1411556](#)
- [28] Van Keilegom, I. & Veraverbeke, N. Estimation and bootstrap with censored data in fixed design nonparametric regression. *Annals Of The Institute Of Statistical Mathematics*. **49** pp. 467-491 (1997) [MR1482368](#)

- [29] Wang, Q. & Chen, S. Moment estimation for censored quantile regression. *Econometric Reviews*. **40**, 815-829 (2021) [MR4296871](#)
- [30] Wang, L., Tchetgen Tchetgen, E., Martinussen, T. & Vansteelandt, S. Instrumental variable estimation of the causal hazard ratio. *Biometrics*. **79**, 539-550 (2023) [MR4606299](#)
- [31] Woodbury, S. & Spiegelman, R. Bonuses to workers and employers to reduce unemployment: Randomized trials in Illinois. *The American Economic Review*. pp. 513-530 (1987)
- [32] Wüthrich, K. A closed-form estimator for quantile treatment effects with endogeneity. *Journal Of Econometrics*. **210**, 219-235 (2019) [MR3958404](#)
- [33] Wüthrich, K. A comparison of two quantile models with endogeneity. *Journal Of Business & Economic Statistics*. **38**, 443-456 (2020) [MR4080995](#)