Supplementary Material

We provide in this supplementary material the proofs of Section 3. Consistency of the proposed estimator relies heavily on the work of Delsol and Van Keilegom (2015) (DVK hereafter) on non-smooth semiparametric M-estimation problems, while the proof of asymptotic normality relies on a modified version of Theorem 2 of Chen et al. (2003) proposed in Birke et al. (2017). This distinction in the tools used for consistency and asymptotic normality is here to be attributed to the particular nature of our loss function. In order to show in the first place that our estimator is weakly consistent, we indeed need to rely here on the arduous work of Delsol and Van Keilegom given that the proposed loss function is no longer strictly convex, and is hence not immune to potential local minima. However, under the conditions that will allow us to establish weak consistency, the need to rely on the work of Delsol and Van Keilegom for establishing asymptotic normality in a second step vanishes in our particular context, as we may conveniently suppose that we avoid possible local minima for a sufficiently large n and for a sufficiently small neighborhood around the true parameter vector. As a consequence, the proof of asymptotic normality will only need to consider shrinking neighborhoods around the true β_τ and G_C instead of the whole spaces $\mathcal{B}$ and $\mathcal{G}$ as for consistency, hereby only requiring the application of (some version of) the work of Chen et al. (2003).

Now, as a preliminary remark and as already mentioned in Section 3, note that the following proofs are built on a crucial result of Lopez (2011) for the class $\mathcal{G}$ in (C3). This implies that our proofs have to be read as such under similar conditions as for Lopez, that is, assuming the existence of a function $g : \mathbb{R}^{d+1} \to \mathbb{R}$ such that $G_C(\cdot|\boldsymbol{X}) = G_C(\cdot|g(\boldsymbol{X}))$, or simply considering the case of a univariate covariate. For ease of reading, the proofs are written considering the latter case.

Proof of Theorem 3.1. In Theorem 1 of DVK, five high level conditions (A1)-(A5) are developed under which an M-estimator is consistent in a general semiparametric maximization problem. We therefore only need to verify the latter conditions to prove the consistency of $\widehat{\beta}_\tau$. For notational convenience with respect to the work of DVK, let us first rewrite $\widehat{\beta}_\tau$ as

$$\widehat{\beta}_\tau = \arg\max_{\beta\in\mathcal{B}} \sum_{i=1}^{n} \psi_\tau \left(\beta^\mathsf{T} \boldsymbol{X}_i; Y_i, \widehat{G}_C(\cdot|\boldsymbol{X}_i)\right), \tag{A.1}$$

where $\psi_\tau(a; y, G) = -\rho_\tau(a; y) + (1-\tau)\int_0^a G(s)\,\mathrm{d}s$ and where $\mathcal{B}$ is a compact parameter space, taken to be the neighborhood of β_τ mentioned in our assumptions. For further convenience and without any loss of generality, the proof is written considering the response variable Y to be positive. This consideration is solely done in the purpose of being coherent with the arbitrary set value of 0 in the correcting term of ψ_τ with respect to v defined in assumption (C3). For instance, one could also easily consider in the following a strictly negative variable Y, but this would possibly require the arbitrary chosen constant 0 to be replaced by any constant below v as we only wish to control the behavior of the nuisance parameter $\widehat{G}_C(\cdot|\boldsymbol{X})$ below this v.

Next, starting with condition (A1) in DVK, note that the latter is readily satisfied in our framework by construction of $\widehat{\beta}_\tau$. Furthermore, using the definition of $\mathcal{G}$ in (C3) as the space embedding the nuisance parameter G_C, and equiping the latter with the distance $d_\mathcal{G}(G_1, G_2) = \sup_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})} \sup_{y\le v} |G_1(y|\boldsymbol{x}) - G_2(y|\boldsymbol{x})|$ for any $G_1, G_2 \in \mathcal{G}$, note that (A3) in DVK is straightforwardly satisfied as well provided assumption (C5)-(i) holds. We therefore only need to verify here conditions (A2), (A4) and (A5).

Starting with the identifiability condition (A2) ensuring the uniqueness of β_τ, we need to verify that for any $\epsilon > 0$, $\inf_{||\beta-\beta_\tau||>\epsilon} \mathbb{E}\left[\psi_\tau\left(\beta_\tau^\mathsf{T}\boldsymbol{X}; Y, G_C(\cdot|\boldsymbol{X})\right) - \psi_\tau\left(\beta^\mathsf{T}\boldsymbol{X}; Y, G_C(\cdot|\boldsymbol{X})\right)\right] >$

0, where $||\cdot||$ denotes the Euclidean distance. To that end, using the definition of φ_τ, we can show that

$$\begin{aligned}
&\inf_{||\beta-\beta_\tau||>\epsilon} \mathbb{E}\left[\psi_\tau\left(\beta_\tau^\mathsf{T}\boldsymbol{X};Y,G_C(\cdot|\boldsymbol{X})\right)-\psi_\tau\left(\beta^\mathsf{T}\boldsymbol{X};Y,G_C(\cdot|\boldsymbol{X})\right)\right]\\
&=\inf_{||\beta-\beta_\tau||>\epsilon} \mathbb{E}\left[\int_{\beta^\mathsf{T}\boldsymbol{X}}^{\beta_\tau^\mathsf{T}\boldsymbol{X}}\left(\mathbb{1}(Y\geq s)-(1-\tau)\bar{G}_C(s|\boldsymbol{X})\right)\mathrm{d}s\right]\\
&=\inf_{||\beta-\beta_\tau||>\epsilon} \mathbb{E}_{\boldsymbol{X}}\left[\int_{\beta^\mathsf{T}\boldsymbol{X}}^{\beta_\tau^\mathsf{T}\boldsymbol{X}}(1-G_C(s|\boldsymbol{X}))(\tau-F_{T|\boldsymbol{X}}(s|\boldsymbol{X}))\mathrm{d}s\right].
\end{aligned}$$

Under assumptions (C1), (C2) and (C4), the latter expectation is observed to be strictly positive, hereby ensuring condition (A2) is satisfied.

Next, for (A4) to hold, it suffices by Remark 1(ii) in DVK and assumption (C3) to show that the class

$$\mathcal{F}=\left\{(y,\boldsymbol{x})\mapsto\psi_\tau\left(\beta^\mathsf{T}\boldsymbol{x};y,G(\cdot|\boldsymbol{x})\right):\beta\in\mathcal{B},G\in\mathcal{G}\right\}$$

is Glivenko-Cantelli. For this, by Theorem 2.4.1 in Van der Vaart and Wellner (1996), we need to prove that for all $\epsilon>0$, the $\epsilon-$bracketing number $N_{[]}(\epsilon,\mathcal{F},L_1(P))$ of the class $\mathcal{F}$ with respect to the L_1 probability measure on $(Y,\boldsymbol{X})$ is finite. To that end, let $\psi_\tau=\psi_{\tau1}+\psi_{\tau2}+\psi_{\tau3}$, where

$$\begin{aligned}
\psi_{\tau1}\left(\beta^\mathsf{T}\boldsymbol{x};y,G(\cdot|\boldsymbol{x})\right)&=-\tau(y-\beta^\mathsf{T}\boldsymbol{x})\\
\psi_{\tau2}\left(\beta^\mathsf{T}\boldsymbol{x};y,G(\cdot|\boldsymbol{x})\right)&=(y-\beta^\mathsf{T}\boldsymbol{x})\mathbb{1}(y<\beta^\mathsf{T}\boldsymbol{x})\\
\psi_{\tau3}\left(\beta^\mathsf{T}\boldsymbol{x};y,G(\cdot|\boldsymbol{x})\right)&=(1-\tau)\int_0^{\beta^\mathsf{T}\boldsymbol{x}}G(s|\boldsymbol{x})\,\mathrm{d}s,
\end{aligned}$$

and let $\mathcal{F}_1,\mathcal{F}_2$ and $\mathcal{F}_3$ denote the classes induced by $\psi_{\tau1},\psi_{\tau2}$ and $\psi_{\tau3}$, respectively. From this decomposition, it is easy to see that

$$N_{[]}(\epsilon, \mathcal{F}, L_1(P)) \leq \prod_{j=1}^{3} N_{[]}(\epsilon, \mathcal{F}_j, L_1(P)). \tag{A.2}$$

Now, for the classes $\mathcal{F}_1$ and $\mathcal{F}_2$, suppose for simplicity and without loss of generality that all coordinates of $\boldsymbol{x}$ are positive, and define $M_\epsilon = O(\epsilon^{-2})$ pairs $(\beta_k^L, \beta_k^U), k = 1, \ldots, M_\epsilon$, that cover $\mathcal{B}$, assumed to be compact by (C1), such that $(\beta_k^{L\top}\boldsymbol{x}, \beta_k^{U\top}\boldsymbol{x})$ define brackets of length $\epsilon\tau/(1-\tau)$ for the class $\{\boldsymbol{x} \mapsto \beta^\top \boldsymbol{x} : \beta \in \mathcal{B}\}$ with respect to the L_1-norm. Then, it is straightforward that $N_{[]}(\epsilon, \mathcal{F}_j, L_1(P)) \leq K_j \epsilon^{-2}$ for some finite constants $K_j > 0, j = 1, 2$, which, combined with (A.2), suggests we only have to verify that $N_{[]}(\epsilon, \mathcal{F}_3, L_1(P))$ is bounded in order to prove that condition (A4) holds in our framework.

To that end, by Lemma 6.1 in Lopez (2011) which extends Theorem 2.7.5 in Van der Vaart and Wellner, first note that there exist $N_\epsilon \leq \exp(K_3 \epsilon^{-2/(1+\eta)})$ brackets $(\underline{G}_j, \overline{G}_j), j = 1, \ldots, N_\epsilon$, for a finite constant $K_3 > 0$ such that, under (C3), for all $G \in \mathcal{G}$, there exists $j = 1, \ldots, N_\epsilon$, for which $\underline{G}_j \leq G \leq \overline{G}_j$, and

$$\int_{\text{supp}(\boldsymbol{X})} \int_0^v \left|\overline{G}_j(s|\boldsymbol{x}) - \underline{G}_j(s|\boldsymbol{x})\right| \mathrm{d}s \, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) < \epsilon, \tag{A.3}$$

where $F_{\boldsymbol{X}}(\boldsymbol{x})$ denotes the c.d.f. of $\boldsymbol{X}$. From this result, our claim for (A4) to hold is that brackets for $\mathcal{F}_3$ are given by $(\underline{\zeta}_{jk}, \overline{\zeta}_{jk}), j = 1, \ldots, N_\epsilon$, $k = 1, \ldots, M_\epsilon$, where

$$\underline{\zeta}_{jk}(\boldsymbol{x}) = (1-\tau) \int_0^{\beta_k^{L\top}\boldsymbol{x}} \underline{G}_j(s|\boldsymbol{x}) \, \mathrm{d}s,$$
$$\overline{\zeta}_{jk}(\boldsymbol{x}) = (1-\tau) \int_0^{\beta_k^{U\top}\boldsymbol{x}} \overline{G}_j(s|\boldsymbol{x}) \, \mathrm{d}s.$$

For this claim to hold, as it is straightforward to verify that for all $\zeta \in \mathcal{F}_3$ there exist

$j = 1, \ldots, N_\epsilon$, and $k = 1, \ldots, M_\epsilon$, such that $\underline{\zeta}_{jk} \leq \zeta \leq \overline{\zeta}_{jk}$, we only need to show that

$$\int_{\mathrm{supp}(\boldsymbol{X})} \left|\overline{\zeta}_{jk}(\boldsymbol{x}) - \underline{\zeta}_{jk}(\boldsymbol{x})\right| \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) < \epsilon, \quad j = 1, \ldots, N_\epsilon,\ k = 1, \ldots, M_\epsilon.$$

To that end, developing the expressions of $\underline{\zeta}_{jk}$ and $\overline{\zeta}_{jk}$, we have that

$$\begin{aligned}
&\int_{\mathrm{supp}(\boldsymbol{X})} \left|\overline{\zeta}_{jk}(\boldsymbol{x}) - \underline{\zeta}_{jk}(\boldsymbol{x})\right| \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) \\
&= (1-\tau) \int_{\mathrm{supp}(\boldsymbol{X})} \left| \int_0^{\beta_k^{U\intercal}\boldsymbol{x}} \overline{G}_j(s|\boldsymbol{x}) \mathrm{d}s - \int_0^{\beta_k^{L\intercal}\boldsymbol{x}} \underline{G}_j(s|\boldsymbol{x}) \mathrm{d}s \right| \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}),
\end{aligned}$$

where the latter expression can be bounded above by $T_1 + T_2$ where

$$\begin{aligned}
T_1 &= (1-\tau) \int_{\mathrm{supp}(\boldsymbol{X})} \left| \int_0^{\beta_k^{U\intercal}\boldsymbol{x}} \overline{G}_j(s|\boldsymbol{x}) \mathrm{d}s - \int_0^{\beta_k^{U\intercal}\boldsymbol{x}} \underline{G}_j(s|\boldsymbol{x}) \mathrm{d}s \right| \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}), \\
T_2 &= (1-\tau) \int_{\mathrm{supp}(\boldsymbol{X})} \left| \int_0^{\beta_k^{U\intercal}\boldsymbol{x}} \underline{G}_j(s|\boldsymbol{x}) \mathrm{d}s - \int_0^{\beta_k^{L\intercal}\boldsymbol{x}} \underline{G}_j(s|\boldsymbol{x}) \mathrm{d}s \right| \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}).
\end{aligned}$$

We will now show that both T_1 and T_2 can be bounded above such that their sum is bounded by ϵ. Starting with T_1, we have that

$$\begin{aligned}
T_1 &\leq (1-\tau) \int_{\mathrm{supp}(\boldsymbol{X})} \int_0^{\beta_k^{U\intercal}\boldsymbol{x}} \left|\overline{G}_j(s|\boldsymbol{x}) - \underline{G}_j(s|\boldsymbol{x})\right| \mathrm{d}s \, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) \\
&\leq (1-\tau) \int_{\mathrm{supp}(\boldsymbol{X})} \int_0^{\upsilon} \left|\overline{G}_j(s|\boldsymbol{x}) - \underline{G}_j(s|\boldsymbol{x})\right| \mathrm{d}s \, \mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x}) \leq (1-\tau)\epsilon,
\end{aligned}$$

using assumption (C4) and (A.3) for the second and last inequalities, respectively. Concentrating now on T_2, we have that

$$
\begin{aligned}
T_2 &\leq (1-\tau)\int_{\mathrm{supp}(\boldsymbol{X})}\int_{\beta_k^{L\mathsf{T}}\boldsymbol{x}}^{\beta_k^{U\mathsf{T}}\boldsymbol{x}}\left|\underline{G}_j(s|\boldsymbol{x})\right|\mathrm{d}s\,\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x})\\
&\leq (1-\tau)\int_{\mathrm{supp}(\boldsymbol{X})}\left|\beta_k^{U\mathsf{T}}\boldsymbol{x}-\beta_k^{L\mathsf{T}}\boldsymbol{x}\right|\mathrm{d}F_{\boldsymbol{X}}(\boldsymbol{x})\leq\tau\epsilon,
\end{aligned}
$$

given the brackets induced by $(\beta_k^{L\mathsf{T}}, \beta_k^{U\mathsf{T}})$ for the class $\{\boldsymbol{x}\mapsto\beta^{\mathsf{T}}\boldsymbol{x}:\beta\in\mathcal{B}\}$ with respect to the L_1-norm. This completes the proof that $N_{[]}(\epsilon,\mathcal{F}_3,L_1(P))$ is bounded. Hence, we conclude that $N_{[]}(\epsilon,\mathcal{F},L_1(P))=O(\exp(K_3\epsilon^{-2/(1+\eta)}))$, from which it follows that condition (A4) holds.

Lastly, for condition (A5), we need to establish that

$$
\lim_{d_{\mathcal{G}}(G,G_C)\to 0}\sup_{\beta\in\mathcal{B}}\left|\mathbb{E}\left[\psi_\tau\left(\beta^{\mathsf{T}}\boldsymbol{X};Y,G(\cdot|\boldsymbol{X})\right)-\psi_\tau\left(\beta^{\mathsf{T}}\boldsymbol{X};Y,G_C(\cdot|\boldsymbol{X})\right)\right]\right|=0.
$$

To that end, note that

$$
\begin{aligned}
&\sup_{\beta\in\mathcal{B}}\left|\mathbb{E}\left[\psi_\tau\left(\beta^{\mathsf{T}}\boldsymbol{X};Y,G(\cdot|\boldsymbol{X})\right)-\psi_\tau\left(\beta^{\mathsf{T}}\boldsymbol{X};Y,G_C(\cdot|\boldsymbol{X})\right)\right]\right|\\
&\leq(1-\tau)\sup_{\beta\in\mathcal{B}}\mathbb{E}_{\boldsymbol{X}}\left[\int_0^{\beta^{\mathsf{T}}\boldsymbol{X}}\left|G(s|\boldsymbol{X})-G_C(s|\boldsymbol{X})\right|\mathrm{d}s\right].
\end{aligned}
$$

Under assumption (C4), this expression can then in turn be bounded above by

$$
(1-\tau)\int_0^{\upsilon}\sup_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})}\left|G(s|\boldsymbol{x})-G_C(s|\boldsymbol{x})\right|\mathrm{d}s\leq(1-\tau)\,\upsilon\sup_{\boldsymbol{x}\in\mathrm{supp}(\boldsymbol{X})}\sup_{y\leq\upsilon}\left|G(y|\boldsymbol{x})-G_C(y|\boldsymbol{x})\right|,
$$

which converges to 0 when $d_{\mathcal{G}}(G,G_C)\to 0$, provided assumption (C3) holds. This completes the proof that (A5) holds in our framework. Hence the assumptions of Theorem 1 in DVK are met, from which the weak consistency of $\widehat{\beta}_\tau$ follows. $\square$

Before developing the proof of Theorem 3.2, we now illustrate in the following Lemma how one could replace the general condition (C5)-(ii) in our assumptions by appropriate bandwidth and kernel conditions when considering the particular case of Beran's estimator described in (2.8).

Lemma 1. *Suppose conditions (C2) and (C3) hold, and that the following assumptions for Beran's estimator in* (2.8) *hold as well:*

(C6) The bandwidth h_n satisfies $h_n = O(n^{-\nu})$, for $1/4 < \nu < 1/3$.

(C7) The kernel function $K(\cdot) \geq 0$ is compactly supported and Lipschitz continuous of order 1. Furthermore, $\int K(u)\,\mathrm{d}u = 1$, $\int uK(u)\,\mathrm{d}u = 0$ and $\int K^2(u)\,\mathrm{d}u < \infty$.

Then, uniformly in $\beta \in \mathcal{B}$,

$$\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\left(\widehat{G}_C(\beta^\top\boldsymbol{X}|\boldsymbol{X}) - G_C(\beta^\top\boldsymbol{X}|\boldsymbol{X})\right)\right] = n^{-1}\sum_{i=1}^{n}\boldsymbol{X}_i\xi(Y_i,\Delta_i,\beta^\top\boldsymbol{X}_i|\boldsymbol{X}_i) + o_{\mathbb{P}}(n^{-1/2}),$$

where ξ takes the following form for a response variable still considered to be positive:

$$\xi(Y_i,\Delta_i,t|\boldsymbol{x}) = (1 - G_C(t|\boldsymbol{x}))\Big[\int_0^{Y_i\wedge t}\frac{-\mathrm{d}H_0(s|\boldsymbol{x})}{\{1-H(s|\boldsymbol{x})\}^2} + \frac{(1-\Delta_i)\mathbb{1}(Y_i\leq t)}{\{1-H(Y_i|\boldsymbol{x})\}}\Big], \tag{A.4}$$

where $H(t|\boldsymbol{x}) = \mathbb{P}(Y \leq t|\boldsymbol{X} = \boldsymbol{x})$ and $H_0(t|\boldsymbol{x}) = \mathbb{P}(Y \leq t, \Delta = 0|\boldsymbol{X} = \boldsymbol{x})$.

Proof of Lemma 1. Using the i.i.d. expansion of $\widehat{G}_C(c|\boldsymbol{x})$ uniformly in c and $\boldsymbol{x}$, under conditions (C2), (C3), (C6) and (C7), we have (see e.g. Gonzalez-Manteiga and Cadarso-

Suarez (1994) or Van Keilegom and Veraverbeke (1997)):

$$
\begin{aligned}
&\mathbb{E}_{\boldsymbol{X}}\big[\boldsymbol{X}\big(\widehat{G}_C(\beta^\top\boldsymbol{X}|\boldsymbol{X}) - G_C(\beta^\top\boldsymbol{X}|\boldsymbol{X})\big)\big] \\
&\qquad = (nh_n)^{-1}\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\sum_{i=1}^{n}\frac{K\left(\frac{\boldsymbol{X}-\boldsymbol{X}_i}{h_n}\right)}{(nh_n)^{-1}\sum_{j=1}^{n}K\left(\frac{\boldsymbol{X}-\boldsymbol{X}_j}{h_n}\right)}\xi(Y_i,\Delta_i,\beta^\top\boldsymbol{X}|\boldsymbol{X})\right] + o_{\mathbb{P}}(n^{-1/2}) \\
&\qquad = (nh_n)^{-1}\sum_{i=1}^{n}\int_{-\infty}^{+\infty}\boldsymbol{x}K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_i}{h_n}\right)\xi(Y_i,\Delta_i,\beta^\top\boldsymbol{x}|\boldsymbol{x})\mathrm{d}\boldsymbol{x} + o_{\mathbb{P}}(n^{-1/2}),
\end{aligned}
$$

where ξ is described in (A.4). Observing that the function $\xi(Y_i, \Delta_i, t|\boldsymbol{x})$ involves an indicator function with respect to t, standard change of variables and a modulus of continuity argument of the empirical distribution function (see Theorem 2.14 in Stute (1982)) combined with assumption (C7) then lead to the desired result. $\square$

Proof of Theorem 3.2. Given that $\widehat{\beta}_\tau$ is shown to be weakly consistent in Theorem 3.1 and that $\widehat{G}_C$ is assumed by (C5)-(i) to be a uniformly consistent estimator of G_C, to prove that our proposed estimator is asymptotically normally distributed we now restrict the spaces $\mathcal{B}$ and $\mathcal{G}$ to shrinking neighborhoods around the true β_τ and G_C in order to avoid possible local minima. That is, we define the spaces $\mathcal{B}_\delta = \{\beta \in \mathcal{B} : ||\beta - \beta_\tau|| \leq \delta_n\}$ and $\mathcal{G}_\delta = \{G \in \mathcal{G} : d_{\mathcal{G}}(G, G_C) \leq \delta_n\}$ for some $\delta_n = o(1)$. In this context, we rely for our proof on the work of Birke et al. (2017) which slightly adapts Theorem 2 of Chen et al. (2003).

We therefore need to verify conditions (C.1)-(C.6) of Proposition 2 in Birke et al. (2017) in order to establish the asymptotic normality of our proposed estimator. To that end, let us first define $M_n(\beta, G) = n^{-1}\sum_{i=1}^{n} m(Y_i, \boldsymbol{X}_i, \beta, G)$, where

$$
m(Y_i, \boldsymbol{X}_i, \beta, G) = \boldsymbol{X}_i\left((1-\tau)(1 - G(\beta^\top\boldsymbol{X}_i|\boldsymbol{X}_i)) - \mathbb{1}(Y_i > \beta^\top\boldsymbol{X}_i)\right).
$$

Furthermore, let

$$
\begin{aligned}
M(\beta, G) &= \mathbb{E}\left[m(Y, \boldsymbol{X}, \beta, G)\right] \\
&= \mathbb{E}_{\boldsymbol{X}}\Big[\boldsymbol{X}\left\{(1-\tau)(1-G(\beta^{\mathsf{T}}\boldsymbol{X}|\boldsymbol{X})) - (1-F_{T|\boldsymbol{X}}(\beta^{\mathsf{T}}\boldsymbol{X}|\boldsymbol{X}))(1-G_C(\beta^{\mathsf{T}}\boldsymbol{X}|\boldsymbol{X}))\right\}\Big],
\end{aligned}
$$

and observe that $M(\beta_\tau, G_C) = 0$.

We now verify the conditions of Proposition 2 in Birke et al.. First, note that (C.1) trivially holds by construction of our estimator. Next, for $\beta \in \mathcal{B}_\delta$ let $\Gamma_1(\beta, G_C)$ denote the ordinary derivative of $M(\beta, G_C)$ with respect to β, that is,

$$
\begin{aligned}
\Gamma_1(\beta, G_C) &:= \frac{\partial M(\beta, G_C)}{\partial \beta} \\
&= \mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\boldsymbol{X}^{\mathsf{T}}\left\{f_{T|\boldsymbol{X}}(\beta^{\mathsf{T}}\boldsymbol{X}|\boldsymbol{X})(1-G_C(\beta^{\mathsf{T}}\boldsymbol{X}|\boldsymbol{X})) + g_C(\beta^{\mathsf{T}}\boldsymbol{X}|\boldsymbol{X})(\tau - F_{T|\boldsymbol{X}}(\beta^{\mathsf{T}}\boldsymbol{X}|\boldsymbol{X}))\right\}\right],
\end{aligned}
$$

where $g_C(\cdot|\boldsymbol{x})$ denotes the density of C conditionally on $\boldsymbol{X} = \boldsymbol{x}$. Under assumptions (C1)-(C4), $\Gamma_1(\beta, G_C)$ is then observed to be continuous and of full rank at β_τ. Hence, condition (C.2) is satisfied in our framework.

For condition (C.3), define first for all $\beta \in \mathcal{B}_\delta$ the functional derivative of $M(\beta, G)$ at G_C in the direction $[G - G_C]$ as

$$
\begin{aligned}
\Gamma_2(\beta, G_C)[G - G_C] &:= \lim_{\eta \to 0} \frac{1}{\eta}\left[M(\beta, G_C + \eta(G - G_C)) - M(\beta, G_C)\right] \\
&= (1-\tau)\mathbb{E}_{\boldsymbol{X}}\left[\boldsymbol{X}\left(G_C(\beta^{\mathsf{T}}\boldsymbol{X}|\boldsymbol{X}) - G(\beta^{\mathsf{T}}\boldsymbol{X}|\boldsymbol{X})\right)\right].
\end{aligned}
$$

We then observe that for all $(\beta, G) \in \mathcal{B}_\delta \times \mathcal{G}_\delta$, $M(\beta, G)$ is linear in G since $M(\beta, G) - M(\beta, G_C) - \Gamma_2(\beta, G_C)[G - G_C] = 0$. This verifies the first part of (C.3). For the second part, we have to show that $||\Gamma_2(\beta, G_C)[\widehat{G}_C - G_C] - \Gamma_2(\beta_\tau, G_C)[\widehat{G}_C - G_C]|| = O_{\mathbb{P}}(n^{-1/2})$. To prove this, using assumption (C5)-(ii), we have uniformly in $\beta \in \mathcal{B}_\delta$ that

$$\Gamma_2(\beta, G_C)[\widehat{G}_C - G_C] = n^{-1}\sum_{i=1}^{n} \boldsymbol{\phi}(Y_i, \Delta_i, \beta^\mathsf{T}\boldsymbol{X}_i, \boldsymbol{X}_i) + o_{\mathbb{P}}(n^{-1/2}),$$

where $\boldsymbol{\phi}(Y_i, \Delta_i, t, \boldsymbol{x}) = -(1-\tau)\,\boldsymbol{x}\,\xi(Y_i, \Delta_i, t|\boldsymbol{x})$. An application of the central limit theorem then implies that $n^{1/2}\Gamma_2(\beta, G_C)[\widehat{G}_C - G_C] \xrightarrow{\mathcal{L}} \mathcal{N}(0, \boldsymbol{V})$ where $\boldsymbol{V}$ is finite under assumptions (C1) and (C5)-(ii). This, in turn, implies that condition (C.3) in Birke et al. is satisfied in our context.

Next, condition (C.4) in Birke et al. is readily satisfied in our context by assumption (C5). To establish that condition (C.5) holds as well, we need to verify the conditions of Theorem 3 in Chen et al. (2003). To that end, let $m = m_c + m_{lc}$, where

$$\begin{aligned} m_c(Y, \boldsymbol{X}, \beta, G) &= \boldsymbol{X}(1-\tau)(1 - G(\beta^\mathsf{T}\boldsymbol{X}|\boldsymbol{X})) \\ m_{lc}(Y, \boldsymbol{X}, \beta, G) &= -\boldsymbol{X}\mathbb{1}(Y > \beta^\mathsf{T}\boldsymbol{X}). \end{aligned}$$

Then, condition (3.1) is easily observed to hold for some $s_{1j}, s_j \in (0, 1]$ and $r = 2$ under assumption (C1). For condition (3.2), as m_{lc} does not depend here on G, we first note via the proof of Theorem 3 in Chen et al. that the constant s_j controlling for the regularity of the nuisance parameter and appearing in condition (3.2) may in fact be replaced by the constant s_{1j} already appearing in condition (3.1). Therefore, we will verify condition (3.2) with respect to s_{1j} instead of s_j as initially stated in Chen et al.. To that end, first it can be observed that for all positive values $\epsilon_n = o(1)$,

$$\sup_{\beta_\star:||\beta-\beta_\star||\leq\epsilon_n} |\mathbb{1}(Y > \beta_\star^\mathsf{T}\boldsymbol{X}) - \mathbb{1}(Y > \beta^\mathsf{T}\boldsymbol{X})| \leq \mathbb{1}(Y > \beta^\mathsf{T}\boldsymbol{X} - \epsilon_n||\boldsymbol{X}||) - \mathbb{1}(Y > \beta^\mathsf{T}\boldsymbol{X} + \epsilon_n||\boldsymbol{X}||).$$

Hence, we have that

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\beta_\star:||\beta-\beta_\star||\le\epsilon_n} \left|\left|\boldsymbol{X}\left(\mathbb{1}(Y>\beta_\star^\intercal\boldsymbol{X})-\mathbb{1}(Y>\beta^\intercal\boldsymbol{X})\right)\right|\right|^2\right] \\
&\le \mathbb{E}\left[||\boldsymbol{X}||^2\left\{\mathbb{1}(Y>\beta^\intercal\boldsymbol{X}-\epsilon_n||\boldsymbol{X}||)-\mathbb{1}(Y>\beta^\intercal\boldsymbol{X}+\epsilon_n||\boldsymbol{X}||)\right\}\right] \\
&= \mathbb{E}_{\boldsymbol{X}}\left[||\boldsymbol{X}||^2\left\{H\left(\beta^\intercal\boldsymbol{X}+\epsilon_n||\boldsymbol{X}||\,|\boldsymbol{X}\right)-H\left(\beta^\intercal\boldsymbol{X}-\epsilon_n||\boldsymbol{X}||\,|\boldsymbol{X}\right)\right\}\right] \\
&\le \mathbb{E}_{\boldsymbol{X}}\left[||\boldsymbol{X}||^3K_4\epsilon_n\right],
\end{aligned}$$

for some finite constant K_4 under assumptions (C2) and (C3). Hence, provided assumption (C1) is satisfied, we may observe that condition (3.2) holds for $s_{1j}=1/2$. For the last condition of Theorem 3 in Chen et al., for $\epsilon>0$ denote first by $N(\epsilon,\mathcal{G}_\delta,||\cdot||_{\mathcal{G}})$ the covering number (Van der Vaart and Wellner (1996, p. 83)) of the class $\mathcal{G}_\delta$ under the sup-norm metric we consider on the latter with a slight abuse of notation. Now, keeping in mind that $N(\epsilon,\mathcal{G}_\delta,||\cdot||_{\mathcal{G}})\le N_{[]}(\epsilon,\mathcal{G}_\delta,||\cdot||_{\mathcal{G}})$, and since all the functions in the class $\mathcal{G}_\delta$ have values between 0 and 1 by (C3), we first observe that only one ϵ-bracket suffices to cover $\mathcal{G}_\delta$ if $\epsilon>1$. Then, using Lemma 6.1 in Lopez (2011) for a bound on the bracketing number for the case $\epsilon\le1$, we have that

$$\begin{aligned}
\int_0^\infty\sqrt{\log N(\epsilon,\mathcal{G}_\delta,||\cdot||_{\mathcal{G}})}\,\mathrm{d}\epsilon &\le \int_0^1\sqrt{\log N_{[]}(\epsilon,\mathcal{G}_\delta,||\cdot||_{\mathcal{G}})}\,\mathrm{d}\epsilon \\
&\le K_5\int_0^1\epsilon^{-\frac{1}{1+\eta}}\,\mathrm{d}\epsilon \\
&<\infty,
\end{aligned}$$

for some finite constant K_5, hereby satisfying condition (3.3) in Chen et al. for $s_j=1$. It then follows from their Theorem 3 that condition (C.5) in Birke et al. holds in our context.

Lastly, for condition (C.6) we need to establish that $n^{1/2}\left[M_n(\beta_\tau,G_C)+\Gamma_2(\beta_\tau,G_C)[\widehat{G}_C-G_C]\right]$

converges to a normal distribution $\mathcal{N}(0,\Sigma)$ for some positive definite matrix Σ. Recalling that $M_n(\beta_\tau, G_C) = n^{-1}\sum_{i=1}^n m(Y_i, \boldsymbol{X}_i, \beta_\tau, G_C)$ is the average of independent random vectors with mean 0, this follows easily using the same arguments as for the verification of condition (C.3) for the particular case of $\beta = \beta_\tau$. Hence, we obtain

$$n^{1/2}\left[M_n(\beta_\tau, G_C) + \Gamma_2(\beta_\tau, G_C)[\widehat{G}_C - G_C]\right] \xrightarrow{\mathcal{L}} \mathcal{N}(0,\Sigma),$$

where $\Sigma = Cov(\Lambda_i)$ with

$$\Lambda_i = m(Y_i, \boldsymbol{X}_i, \beta_\tau, G_C) - (1-\tau)\,\boldsymbol{X}_i\,\xi(Y_i, \Delta_i, \beta_\tau^{\mathsf{T}}\boldsymbol{X}_i|\boldsymbol{X}_i). \tag{A.5}$$

Theorem 3.2 then follows directly from an application of Proposition 2 in Birke et al., hereby concluding the proof.

$\square$