Manuscript submitted to *The Econometrics Journal*, pp. S1–S24.

Distribution Regression in Duration Analysis: an Application to Unemployment Spells Online supplement

Miguel A. Delgado[†], Andrés García-Suaza[‡] and Pedro H.C. Sant'Anna[††]

[†] *Universidad Carlos III de Madrid, Calle Madrid 126, 28903 Getafe, Madrid, Spain*
E-mail: delgado@est-econ.uc3m.es
[‡] *Universidad del Rosario, Calle 14 625, Bogotá, Colombia.*
E-mail: andres.garcia@urosario.edu.co
[‡] *Microsoft and Vanderbilt University, 415 Calhoun Hall, Nashville, TN 37240, USA.*
E-mail: pedro.h.santanna@vanderbilt.edu

Summary

In this Supplementary Appendix we provide details about the Monte Carlo simulations used to illustrated the finite sample properties of the proposed Kaplan-Meier Distribution Regression estimators and its functionals. In addition, we provide the mathematical proofs of all results stated in the main text.

S1. MONTE CARLO

In this section, we compare the finite sample performance of the proposed KMDR estimators with those based on the Cox (1972, 1975) proportional hazard (PH) model and on the Clayton (1976) and Bennett (1983) proportional odds (PO) model.

We consider the following three data generating processes (DGPs) :

1. $T = (-\ln(1-U))^{\frac{1}{2}} \exp(X)$,
2. $T = \exp\left(\frac{X}{4}\right)\left(\frac{1}{U}-1\right)^{\frac{1}{4}}$,
3. $T = \exp\left(\frac{\ln(-\ln(1-U)) - X}{1+2X}\right)$

where U and X are mutually independent and both follow a uniform distribution on $(0, 1)$.

Table 1 displays the conditional CDF, $F_{T|X}$, hazard, $\lambda_{T|X}$, and odds-ratio, $\Gamma_{T|X}$, associated with each DGP. In DGP 1, we have that $T|X$ follows a Weibull distribution with scale parameter $\exp(x)$, shape parameter 2, which satisfies the PH but not the PO assumption. In DGP 2, we have that $T|X$ follows a log-logistic distribution, which satisfies the PO but not the PH assumption. Finally, in DGP 3, we have $T|X$ follows a Weibull distribution with scale parameter $\exp(-x/(1+2x))$ and shape parameter $1+2x$. Furthermore, both PH and PO assumptions are violated in DGP 3, though it admits a DR specification with time-varying slope coefficient equal to $1 + \ln(t)$.

To allow for different levels of right-censoring, we generate censoring random variable C according to

$$C_{d,c} = a_d + b_{d,c}E,$$

Table 1. Conditional CDF, hazard, and odds-ratio associated with DGPs

DGP	$F_{T\|X}(t\|x)$	$\lambda_{T\|X}(t\|x)$	$\Gamma_{T\|X}(t\|x)$
1	$1-\exp\left(-\frac{t^2}{\exp(2x)}\right)$	$\frac{2t}{\exp(2x)}$	$\exp\left(\frac{t^2}{\exp(2x)}\right)-1$
2	$\frac{\exp(4\ln t - x)}{1+\exp(4\ln t - x)}$	$\frac{4t^3\exp(-x)}{1+t^4\exp(-x)}$	$t^4\exp(-x)$
3	$1-\exp(-\exp(\ln(t)+(1+\ln(t))x))$	$t^{2x}(1+2x)\exp(x)$	$\exp\left(t^{1+2x}\exp(x)\right)-1$

where $d=1,2,3$, $c=0,10,30$, E follows a standard exponential distribution, and a_d and $b_{d,c}$ are chosen such that the percentage of censoring is equal to 0, 10 or 30 percent. The observed data is $\{Y_i,\delta_i,X_i\}_{i=1}^n$, where $Y_i=\min(T_i,C_i)$ and $\delta_i=1_{\{T_i\le C_i\}}$.

The comparison between KMDR, PH and PO models is based on the conditional CDF, $F_{T|X}$, and the ADME as defined in (5.2). Both functionals have a clear economic interpretation.

We consider two alternative link functions for the KMDR model: the logistic link function $\Psi(u)=(1+\exp(-u))^{-1}$, and the complementary log-log link function $\Psi(u)=1-\exp(-\exp(u))$. These specifications lead to different estimators of the conditional CDF $F_{T|X}(t|X)$, namely

$$\hat{F}_n^{dr,\,\mathrm{l}}(t|X)=\left(1+\exp\left(-\left(\hat{\alpha}_{0,n}^{\mathrm{l}}(t)+\hat{\alpha}_{1,n}^{\mathrm{l}}(t)X\right)\right)\right)^{-1},$$
$$\hat{F}_n^{dr,\,\mathrm{cll}}(t|X)=1-\exp\left(-\exp\left(\hat{\alpha}_{0,n}^{\mathrm{cll}}(t)+\hat{\alpha}_{1,n}^{\mathrm{cll}}(t)X\right)\right),$$

where $\hat{\alpha}_{0,n}^{\mathrm{l}}(t)$, $\hat{\alpha}_{1,n}^{\mathrm{l}}(t)$ ($\hat{\alpha}_{0,n}^{\mathrm{cll}}(t)$, $\hat{\alpha}_{1,n}^{\mathrm{cll}}(t)$) are the logit (complementary log-log) KMDR estimators of the unknown parameters $\alpha_0(t)$ and $\alpha_1(t)$. We denote the KMDR specifications as DR_l and DR_{cll}, respectively. As discussed in Section 2, the PH and PO specifications lead to alternative estimators of $F_{T|X}(t|X)$, namely,

$$\hat{F}_n^{ph}(t|X)=1-\exp\left(-\exp\left(\ln\hat{\Lambda}_{0,n}(t)+\hat{\alpha}_n^{\mathrm{ph}}X\right)\right),$$
$$\hat{F}_n^{po}(t|X)=\frac{\hat{\Gamma}_{0,n}(t)\exp(-\hat{\alpha}_n^{\mathrm{po}}X)}{1+\hat{\Gamma}_{0,n}(t)\exp(-\hat{\alpha}_n^{\mathrm{po}}X)},$$

respectively, where $\hat{\alpha}_m^{ph}$ and $\hat{\Lambda}_{0,n}(t)$ are the Cox (1975) partial likelihood estimator of α^{ph} and the Breslow (1974) estimator of the cumulative baseline hazard $\Lambda_0(t)$, and $\hat{\Gamma}_{0,n}(t)$ and $\hat{\alpha}_n^{po}$ are the Hunter and Lange (2002) estimators for the baseline odds to death function $\Gamma_0(t)$, and α^{po}.

Tables 2, 3, and 4 list the performance of the estimators of $100\cdot F_{T|X}(t|0.5)$ and $100\cdot\mathrm{ADME}(t)$ in terms of average absolute bias and root mean square error (RMSE) over a hundred different values of thresholds t located at 100 equidistant points between the 0.10 and 0.90 marginal quantiles of T. In each table, we list the results with no censoring, 10% censoring, and 30% censoring, and with sample size $n=100,\ 400$, and 1600.

Table 2 shows that under DGP 1, for all sample sizes and censoring levels, both CDF and ADME estimators based on PH and DR_{cll} specifications have lower bias than those based on PO and DR_l specifications. These results are expected, as the PH and DR_{cll}

models are correctly specified, but the PO and DR_l models are misspecified. As the sample size increases from 100 to 400 and from 400 to 1600, i.e., when sample size increases from $\sqrt{n}$ to $2\sqrt{n}$, the average RMSE of our proposed KMDR estimators for both $F_{T|X}\,(t|\,0.5)$ and ADME(t) decreases by approximately 50%, as is expected from a $\sqrt{n}$-consistent estimator.

Table 2. Simulated finite-sample properties under DGP 1.

		No Censoring				10 % Censoring				30% Censoring			
	n	PH	PO	DR_{cll}	DR_l	PH	PO	DR_{cll}	DR_l	PH	PO	DR_{cll}	DR_l
Average	100	0.13	0.78	0.44	1.33	0.38	0.70	0.56	1.61	0.11	0.73	0.87	1.82
absolute bias	400	0.05	0.95	0.07	1.37	0.06	0.94	0.13	1.30	0.10	0.85	0.30	1.54
for $F\,(t\vert X=0.5)$	1600	0.07	0.97	0.07	1.42	0.06	0.94	0.07	1.42	0.01	0.89	0.06	1.40
Average	100	4.08	4.21	4.48	4.60	4.39	4.44	4.88	5.03	4.99	4.92	5.88	5.91
RMSE	400	2.09	2.41	2.26	2.69	2.16	2.47	2.33	2.71	2.44	2.60	2.76	3.15
for $F\,(t\vert X=0.5)$	1600	1.05	1.56	1.12	1.86	1.10	1.58	1.18	1.90	1.21	1.61	1.32	1.98
Average	100	0.60	5.76	0.71	2.55	0.87	6.53	0.60	1.93	0.62	7.15	1.10	1.71
absolute bias	400	0.10	5.45	0.18	2.28	0.07	5.65	0.17	2.15	0.32	6.53	0.58	2.01
for $AMDE\,(t)$	1600	0.07	5.46	0.04	2.08	0.06	5.65	0.07	2.11	0.07	6.25	0.07	2.04
Average	100	8.93	11.85	12.45	13.31	9.55	12.89	13.48	14.25	10.65	13.90	17.34	18.15
RMSE	400	4.64	7.87	6.20	7.07	4.40	7.79	6.29	7.14	5.20	8.99	7.73	8.43
for $AMDE\,(t)$	1600	2.20	6.21	3.03	4.05	2.32	6.40	3.20	4.21	2.56	7.05	3.81	4.70

Note: Simulations based on one thousand Monte Carlo experiments. "PH" stands for estimators based on the proportional hazard model. "PO" stands for estimators based on the proportional odds model. "DR_{cll}" and "DR_l" stand for estimators based on the proposed distribution regression mode with the *cloglog* and *logit* link functions, respectively.

Table 3 shows that under DGP 2, all considered estimators for the distribution function have little to no bias for all considered sample sizes and censoring levels. In terms of ADMEs, our proposed KMDR estimators have little to no bias, just like the PO model. On the other hand, ADME(t) estimators based on the PH specification are biased and the biases do not disappear when the sample size increases. Interestingly, even when our proposed KMDR model has a misspecified link function, they perform very similarly to the PO model in terms of both bias and RMSE, especially when the sample size is moderate.

Table 3. Simulated finite-sample properties under DGP 2.

		No Censoring				10 % Censoring				30% Censoring			
	n	PH	PO	DR_{cll}	DR_l	PH	PO	DR_{cll}	DR_l	PH	PO	DR_{cll}	DR_l
Average	100	0.27	0.52	0.26	0.32	0.35	0.69	0.34	0.48	0.15	0.74	0.85	1.12
absolute bias	400	0.13	0.14	0.17	0.10	0.14	0.15	0.14	0.09	0.13	0.25	0.22	0.42
for $F\,(t\vert X=0.5)$	1600	0.10	0.05	0.16	0.04	0.12	0.07	0.12	0.06	0.10	0.05	0.08	0.13
Average	100	4.11	4.14	4.24	4.26	4.45	4.48	4.66	4.70	4.87	4.91	5.41	5.45
RMSE	400	2.08	2.08	2.11	2.11	2.15	2.15	2.19	2.20	2.38	2.37	2.47	2.51
for $F\,(t\vert X=0.5)$	1600	1.03	1.02	1.04	1.04	1.10	1.09	1.11	1.11	1.23	1.22	1.26	1.27
Average	100	3.80	0.22	0.26	0.33	3.63	0.32	0.25	0.29	3.23	0.47	0.64	0.70
absolute bias	400	4.20	0.19	0.27	0.20	3.96	0.21	0.28	0.20	3.54	0.12	0.34	0.23
for $AMDE\,(t)$	1600	4.24	0.08	0.12	0.08	3.94	0.04	0.13	0.07	3.62	0.02	0.11	0.05
Average	100	11.14	10.78	14.35	14.44	11.37	10.84	15.29	15.44	12.65	11.87	19.47	19.59
RMSE	400	7.15	5.53	7.24	7.28	6.95	5.49	7.63	7.66	7.16	5.64	9.17	9.21
for $AMDE\,(t)$	1600	5.23	2.66	3.53	3.54	5.04	2.78	3.87	3.90	4.99	2.99	4.78	4.82

Note: Simulations based on one thousand Monte Carlo experiments. "PH" stands for estimators based on the proportional hazard model. "PO" stands for estimators based on the proportional odds model. "DR_{cll}" and "DR_l" stand for estimators based on the proposed distribution regression mode with the *cloglog* and *logit* link functions, respectively.

Table 4 shows that under DGP 3, when both proportional hazard and proportional odds assumptions are violated, the CDF estimators under PH and PO specifications are slightly biased, especially when censoring is heavier. Importantly, such bias does not disappear when sample size increases. Our proposed KMDR estimators for the CDF perform better than those based on the PH and PO specifications, especially when the censoring level is high and the sample size is moderate. Once the focus is shifted towards ADME(t), one can easily see that our KMDR estimators perform substantially better than those based on PH or PO specification. In fact, the average bias of the PH and PO estimators of ADME(t) is never lower than 15 percentage points, and such biases do not vanish as the sample size increases. Our proposed KMDR estimators of ADME(t), on the other hand, have little to no bias, and the RMSE decreases at the appropriate $\sqrt{n}$-rate, even when the link function is misspecified.

Table 4. Simulated finite-sample properties under DGP 3.

		No Censoring				10 % Censoring				30% Censoring			
	n	PH	$P0$	DR_{cll}	DR_l	PH	$P0$	DR_{cll}	DR_l	PH	$P0$	DR_{cll}	DR_l
Average	100	0.62	0.59	0.35	0.65	0.63	0.72	0.36	0.69	0.58	0.78	0.98	1.24
absolute bias	400	0.85	0.69	0.08	0.41	0.73	0.64	0.12	0.45	0.67	0.59	0.25	0.55
for $F(t\|X=0.5)$	1600	0.91	0.77	0.03	0.34	0.81	0.73	0.03	0.34	0.61	0.62	0.15	0.46
Average	100	4.29	4.42	4.29	4.32	4.42	4.58	4.55	4.57	4.86	4.97	5.36	5.35
RMSE	400	2.34	2.41	2.12	2.20	2.42	2.52	2.28	2.36	2.57	2.64	2.54	2.62
for $F(t\|X=0.5)$	1600	1.46	1.48	1.08	1.19	1.41	1.47	1.09	1.20	1.40	1.50	1.25	1.38
Average	100	17.13	18.09	0.29	0.70	16.50	18.31	0.56	0.74	16.16	19.15	3.14	2.85
absolute bias	400	17.44	18.42	0.14	0.67	16.72	18.68	0.50	0.58	16.15	19.42	1.40	1.07
for $AMDE(t)$	1600	17.44	18.53	0.07	0.51	16.78	18.72	0.06	0.53	16.19	19.56	0.76	0.67
Average	100	21.51	23.32	14.69	15.10	21.05	23.39	16.00	16.33	21.42	24.29	20.60	20.57
RMSE	400	18.84	20.02	7.09	7.37	18.21	20.31	7.46	7.75	17.86	21.06	9.62	9.87
for $AMDE(t)$	1600	17.90	19.04	3.56	3.79	17.26	19.23	3.77	4.01	16.71	20.10	4.73	4.89

Note: Simulations based on one thousand Monte Carlo experiments. "PH" stands for estimators based on the proportional hazard model. "PO" stands for estimators based on the proportional odds model. "DR_{cll}" and "DR_l" stand for estimators based on the proposed distribution regression mode with the *cloglog* and *logit* link functions, respectively.

Overall, the simulation evidence supports that the new KMDR procedure is a useful research tool when dealing with duration data.

S2. PROOFS OF MAIN RESULTS

Henceforth, $Z = (Y, X', \delta)'$ is a random vector defined in the sample space $(\mathcal{Z},\mathcal{A})$, $\mathcal{Z} = (\mathcal{Y} \times \mathcal{X} \times [0,1])$. We shall use the standard empirical process notation $Pf = \mathbb{E}f(Z) = \int f(Z)\, dP$. Furthermore, let $\mathbb{P}_n = n^{-1}\sum_{i=1}^n \delta_{Z_i}$ denote the empirical measure of $\{Z_i\}_{i=1}^n$, $Z_i = (Y_i, X_i', \delta_i)'$, i.e. ,$\mathbb{P}_n f = n^{-1}\sum_{i=1}^n f(Z_i)$. Analogously, let $\widehat{\mathbb{P}}_n$ denotes the empirical Kaplan-Meier measure, i.e. $\widehat{\mathbb{P}}_n f = \sum_{i=1}^n W_{ni} f\left(Y_{i:n}, X_{[i:n]}, \delta_{[i:n]}\right)$ for any function f, with W_{ni} as in (3.2), $Y_{i:n}$ the *i-th* order statistic of Y and $X_{[i:n]}$ its corresponding concomitant.

For any class of functions $\mathcal{G}$ and any probability function Q, we use the notation $\|a\| = \left(\sum_{j=1}^p a_j^2\right)^{1/2}$ for any vector $a = (a_1, ..., a_p)'$, $\|\varphi\|_{\mathcal{G}} = \sup_{g\in\mathcal{G}} \|\varphi(g)\|$, and $\|g\|_{r,Q} = (Qg^r)^{1/r}$, $r > 0$. Let e be a generic pseudo-norm on $\mathcal{G}$, and $N(\varepsilon, \mathcal{G}, e)$ denotes the covering number with respect to e, i.e.,

$$N(\varepsilon, \mathcal{G}, e) = \min\left\{s : \exists g_1, ..., g_s \text{ s.t. } \sup_{g\in\mathcal{G}} \min_{j\le s} e(g, g_j) \le \varepsilon\right\}.$$

Given two functions $l, u \in \mathcal{G}$, a bracket $[l, u]$ is the set of functions $f \in \mathcal{G}$ such that $l \le f \le u$. An ε-bracket with respect to $\|\cdot\|_{r,Q}$ is a bracket $[l, u]$ with $\|l - u\|_{r,Q} \le \varepsilon$, $\|l\|_{r,Q} < \infty$, and $\|u\|_{r,Q} < \infty$. The bracketing number $N_{[\,]}\left(\varepsilon, \mathcal{G}, \|\cdot\|_{r,Q}\right)$ is the minimal number of ε-brackets with respect to $\|\cdot\|_{r,Q}$ needed to cover $\mathcal{G}$, that is,

$$N_{[\ ]}\left(\varepsilon, \mathcal{G}, \|\cdot\|_{r,Q}\right) = \min\Big\{s : \exists g_1, ..., g_s \text{ and } \Delta_1, ..., \Delta_s \in L_r(Q) \ \ s.t. \ \|\Delta_j\|_{r,Q} < \varepsilon,$$
$$\text{and for all } g \in \mathcal{G} \text{ there exists } j \leq s \text{ with } |g_j - g| < \Delta_j\Big\}$$

We take by granted measurability where it is needed.

PROOF. (PROOF OF THEOREM 4.1) The proof of this theorem follows Example 5.40 in van der Vaart (1998) (VV, henceforth). Let

$$m_\theta(d, x) = \log\left(\frac{p_{\theta_0(t)} + p_\theta}{2}\right)(d, x),$$
$$M_t(\theta) = \int m_\theta(1_{\{T \leq t\}}, X) dP,$$
$$M_{nt}(\theta) = \sum_{i=1}^{n} W_{ni} \cdot m_\theta(1_{\{Y_{i:n} \leq t\}}, X_{[i:n]}).$$

Note that, by the concavity of the logarithm function and the characterization of $\hat{\theta}_n(t)$ as the maximizer of $\hat{Q}_n(\theta, t)$,

$$M_{nt}(\hat{\theta}(t)) \geq \frac{1}{2}\left(\hat{Q}_n(\hat{\theta}_n(t), t) + \hat{Q}_n(\theta_0(t), t)\right) \geq M_{nt}(\theta_0(t)).$$

Thus, in light of Theorem 5.7 of VV, it suffices to show that,

$$\sup_{\theta:\|\theta - \theta_0(t)\| > \epsilon} M_t(\theta) < M_t(\theta_0(t)) \text{ for all } \epsilon > 0, \tag{S.1}$$
$$\lim_{n\to\infty} \sup_{\theta\in\Theta} |M_{nt}(\theta) - M_t(\theta)| = 0 \ a.s.. \tag{S.2}$$

We focus on $M_t(\theta)$ instead of $Q(\theta, t)$ because $p_{\theta_0(t)}$ is bounded away from zero (and one), suggesting that the function m_0 is somewhat better behaved than $log p_\theta$. Then, using the fact that $\log x \leq 2(\sqrt{x} - 1)$ for all $x \geq 0$, and $p_\theta(0, X) + p_\theta(1, X) = 1$ a.s. for all $\theta \in \Theta$, and that for non-negative number a, b, $\sqrt{(a + b)/2} \geq \left(\sqrt{a} + \sqrt{b}\right)/2$, we have that

$$\begin{aligned}
M_t(\theta) - M_t(\theta_0(t)) &= E\left[\log\left(\frac{(p_\theta + p_{\theta_0(t)})}{2p_{\theta_0(t)}}(1_{\{T\leq t\}}, X)\right)\right] \\
&= \sum_{d=0}^{1} \mathbb{E}\left[\log\left(\frac{(p_\theta + p_{\theta_0(t)})}{2p_{\theta_0(t)}}(d, X)\right) \cdot p_{\theta_0(t)}(d, X)\right] \\
&\leq 2\sum_{d=0}^{1} \mathbb{E}\left[\left(\sqrt{\frac{(p_\theta + p_{\theta_0(t)})}{2p_{\theta_0(t)}}}(d, X) - 1\right) p_{\theta_0(t)}(d, X)\right] \\
&= 2\sum_{d=0}^{1} \mathbb{E}\left[\sqrt{\frac{(p_\theta + p_{\theta_0(t)})\, p_{\theta_0(t)}}{2}}(d, X)\right] - 2 \\
&= -\sum_{d=0}^{1} \mathbb{E}\left[\left(\sqrt{\frac{p_\theta + p_{\theta_0(t)}}{2}} - \sqrt{p_{\theta_0(t)}}\right)^2 (d, X)\right] \\
&\leq -\sum_{d=0}^{1} \mathbb{E}\left[\left(\frac{\sqrt{p_\theta} + \sqrt{p_{\theta_0(t)}}}{2} - \sqrt{p_{\theta_0(t)}}\right)^2 (d, X)\right]
\end{aligned}$$

$$
\begin{aligned}
&= -\sum_{d=0}^{1} \mathbb{E}\left[\left(\frac{\left(p_{\theta}-p_{\theta_0(t)}\right)}{2\left(\sqrt{p_{\theta}}+\sqrt{p_{\theta_0(t)}}\right)}\right)^2 (d,X)\right]^2 \\
&\leq -K \cdot \mathbb{E}\left[\left(\Lambda\left(\boldsymbol{X}'\theta\right)-\Lambda\left(\boldsymbol{X}'\theta_0(t)\right)\right)^2\right],
\end{aligned}
$$

where, henceforth, K is a generic constant. This is always non-positive and is zero only if $\theta = \theta_0(t)$. Thus, $\theta_0(t)$ is the unique maximum of $M_t(\theta)$. Furthermore, $M_t(\theta_k) \to M_t(\theta_0(t))$ implies that $X'\theta_k \to_p X'\theta_0(t)$. If the sequence θ_k is also bounded, then $\mathbb{E}\left((\theta_k - \theta_0(t))'\boldsymbol{X}\right)^2 \to 0$ and, hence, $\theta_k \to \theta_0(t)$ because $\mathbb{E}\left(\boldsymbol{X}\boldsymbol{X}'\right)$ is non-singular. On the other hand, $\|\theta_k\|$ cannot have a diverging subsequence because, in that case, $\theta_k'\boldsymbol{X}/\|\theta_k\| \to_p 0$ and hence, $\theta_k/\|\theta_k\| \to_p 0$ by the same argument. This proves (S.1).

In order to prove (S.2), applying arguments in VV Example 5.40, the functions $u \mapsto \Lambda(u)$ form a VC-class (see, e.g.,VVs Example 19.7), and the functions m_θ take the form $m_\theta(d,x) = \varphi\left(\Lambda(\theta'\boldsymbol{x}), d, \Lambda(\theta_0'\boldsymbol{x})\right)$, where the function $\varphi(\gamma, d, \eta)$ is Lipschitz in its first argument with Lipschitz constant bounded above by $1/\eta + 1/(1-\eta)$; this follows from Λ being Lipschitz (since it is continuously differentiable). Thus, it follows that the class $\mathcal{M}$ of functions $(x,d) \mapsto m_\theta(d,x)$ indexed by parameters $\theta \in \Theta$ has bracketing numbers bounded by the bracketing numbers of Θ (see VV's example 19.7) i.e., for any $\varepsilon \in (0, diam\ \Theta)$,

$$
N_{[\,]}\left(\varepsilon M, \mathcal{M}, \|\cdot\|_{r,P}\right) \leq K\left(\frac{diam\Theta}{\varepsilon}\right)^p < \infty,
$$

for some constant K independent of ε and p, where $M = \sup_{\theta\in\Theta}|\dot{m}_\theta(d,X)| < \infty$, and $\dot{m}_\theta$ is the derivative of m_θ with respect to θ. Hence, $\mathcal{M}$ is Glivenko-Cantelli, i.e.

$$
\sup_{\theta\in\Theta}\left|\left(\mathbb{P}_n - P\right)m_\theta\right| \to 0 \text{ a.s.-}\mathbb{P}
$$

applying Theorem 2.4.1 of van der Vaart and Wellner (1996).

To conclude the proof, we combine Stute (1993) Theorem and Theorem 2.4.1 in van der Vaart and Wellner (1996) to show that

$$
\sup_{\theta\in\Theta}\left|\left(\widehat{\mathbb{P}}_n - P\right)m_\theta\right| \to 0 \text{ a.s.-}\mathbb{P}.
$$

Towards this end, fix $\varepsilon > 0$. Choose finitely many ε-brackets $[\ell_j, u_j]$ whose union contains $\mathcal{M}$ and such that $P(u_j - \ell_j) < \varepsilon$ for every j. Then for every $\theta \in \Theta$, there is a bracket such that

$$
\left(\widehat{\mathbb{P}}_n - P\right)m_\theta \leq \left(\widehat{\mathbb{P}}_n - P\right)u_j + P(u_j - m_\theta) \leq \left(\widehat{\mathbb{P}}_n - P\right)u_j + \varepsilon.
$$

Consequently,

$$
\sup_{\theta\in\Theta}\left(\widehat{\mathbb{P}}_n - P\right)m_\theta \leq \max_j\left(\widehat{\mathbb{P}}_n - P\right)u_j + \varepsilon.
$$

The right side converges almost surely to ε by Stute (1993)'s Theorem. Combination with a similar argument for $\inf_{\theta\in\Theta}\left(\widehat{\mathbb{P}}_n - P\right)m_\theta$ yields that $\limsup\sup_{\theta\in\Theta}\left|\left(\widehat{\mathbb{P}}_n - P\right)m_\theta\right| \leq \varepsilon$ *a.s.*, for every $\varepsilon > 0$. Take a sequency $\varepsilon_m \downarrow 0$ to see that the $\limsup$ must actually be zero almost surely.□

In order to prove Theorems 4.2 and 4.3, we provide a uniform version of the central limit theorem's in Stute (1995, 1996). The Kaplan-Meier integral

$$S_n(\varphi) = \widehat{\mathbb{P}}_n \varphi = \sum_{i=1}^{n} W_{ni} \varphi\left(Y_{i:n}, X_{[i:n]}\right) = \int_{\mathbb{R}^{1+k}} \varphi(t,x) \hat{F}_n(dt, dx),$$

can be viewed as a random process indexed by a vector of functions $\varphi : \mathbb{R}^{1+k} \mapsto \mathbb{R}^q$ in a class $\mathcal{F}$. The proof follows the following steps.

First, we show that $\|S_n(\varphi) - V_n(\varphi)\|_{\mathcal{F}} = o_P\left(n^{-1/2}\right)$, where

$$\begin{aligned}
V_n(\kappa_\varphi) &= \frac{1}{n^3} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{m=1}^{n} \kappa_\varphi(Z_i, Z_j, Z_m) \\
&= -\int \varphi(t_3, x_3) \gamma^{(0)}(t_3) \int \frac{1_{\{t_1 < t_3\}}}{1 - F_Y(t_1)} F_Y^0(dt_1) \tilde{F}_n^1(dt_3, dx_3) \\
&\quad + \int \varphi(t_3, x_3) \gamma^{(0)}(t_3) \tilde{F}_n^1(dt_3, dx_3) \\
&\quad + 2 \int \varphi(t_3, x_3) \gamma^{(0)}(t_3) \int \frac{1_{\{t_1 < t_3\}}}{1 - F_Y(t_1)} \tilde{F}_{Yn}^0(dt_1) \tilde{F}_n^1(dt_3, dx_3) \\
&\quad - \int \varphi(t_3, x_3) \gamma^{(0)}(t_3) \int \int \frac{1_{\{t_1 < t_2, t_1 < t_3\}}}{[1 - F_Y(t_1)]^2} \tilde{F}_{Yn}(dt_2) \tilde{F}_{Yn}^0(dt_1) \tilde{F}_n^1(dt_3, dx_3) \\
&= (\mathbb{P}_n \otimes \mathbb{P}_n \otimes \mathbb{P}_n) \kappa_\varphi
\end{aligned}$$

is a V-process of order 3 with

$$\begin{aligned}
\kappa_\varphi(z_1, z_2, z_3) &= \kappa_{1\varphi}(z_3) + \kappa_{2\varphi}(z_1, z_3) + \kappa_{3\varphi}(z_1, z_2, z_3), \\
\kappa_{1\varphi}(z_3) &= \varphi(t_3, x_3) \gamma^{(0)}(t_3) \delta_3 \left[1 - \int \frac{1_{\{t_1 < t_3\}}}{1 - F_Y(t_1)} F_Y^0(dt_1)\right], \\
\kappa_{2\varphi}(z_1, z_3) &= 2\varphi(t_3, x_3) \gamma^{(0)}(t_3) \delta_3 \frac{1_{\{t_1 < t_3\}}(1 - \delta_1)}{1 - F_Y(t_1)}, \\
\kappa_{3\varphi}(z_1, z_2, z_3) &= -\varphi(t_3, x_3) \gamma^{(0)}(t_3) \delta_3 \frac{1_{\{t_1 < t_2, t_1 < t_3\}}(1 - \delta_1)}{[1 - F_Y(t_1)]^2},
\end{aligned} \tag{S.3}$$

for $z_j = \left(y_j, x_j', d_j\right)'$, $j = 1, 2, 3$, and the empirical distributions $\tilde{F}_{Yn}(t) = n^{-1} \sum_{i=1}^{n} 1_{\{Y_i \le t\}}$, $\tilde{F}_{Yn}^0(t) = n^{-1} \sum_{i=1}^{n} (1 - \delta_i) 1_{\{Y_i \le t\}}$ and $\tilde{F}_n^1(t, x) = n^{-1} \sum_{i=1}^{n} \delta_i 1_{\{Y_i \le t, X_i \le x\}}$ are the sample analogues of $F_Y(t) = \mathbb{P}(Y \le t)$, $F_Y^0(t) = \mathbb{P}(Y \le t, \delta = 0)$ and $F_{YX}^1(t, x) = \mathbb{P}(Y \le t, X \le x, \delta = 1)$, respectively.

Second, we show that $\|V_n(\varphi) - U_n(\varphi)\|_{\mathcal{F}} = o_P\left(n^{-1/2}\right)$ under suitable restrictions on the $\mathcal{F}$'s bracketing numbers, where

$$U_n(\varphi) = U_n^{(1)}(\kappa_{1\varphi}) + U_n^{(2)}(\kappa_{2\varphi}) + U_n^{(3)}(\kappa_{3\varphi}),$$

and for $(z_1, z_2, z_3) \mapsto g(z_1, z_2, z_3)$,

$$U_n^{(m)}(g) = \frac{(n-m)!}{n!} \sum_{\pi} g(Z_{i_1}, ..., Z_{i_m}),$$

is the U-process of order m, with π running on all 3-permutations $\pi = (i_1, i_2, i_3)$ of integers $0 \le i_j \le n$ with $i_1 \ne i_2 \ne i_3$. We then obtain the Hájek projection

$$\hat{U}_n(\varphi) = \mathbb{E}[U_n(\kappa_\varphi) | Z_i] - (n-1)\mathbb{E}U_n(\kappa_\varphi) = \mathbb{P}_n \zeta_\varphi,$$

where

$$\zeta_\varphi(z) = \varphi(t,x)\gamma^{(0)}(t)d + \gamma_\varphi^{(1)}(y)(1-d) - \gamma_\varphi^{(2)}(y),$$

with

$$\gamma^{(0)}(t) = \exp\left\{\int_0^{t-} \frac{F_Y^0(dt)}{1-F_Y(t)}\right\},$$
$$\gamma_\varphi^{(1)}(t) = \frac{1}{1-F_Y(t)}\int 1_{\{t<t_2\}}\varphi(t_2,x)\gamma^{(0)}(t_2)F^1(dt_2,dx),$$
$$\gamma_\varphi^{(2)}(t) = \int\int \frac{1_{\{t_1<t,t_1<t_2\}}\varphi(t_2,x)\gamma^{(0)}(t_2)}{[1-F_Y(t_1)]^2}F_Y^0(dt_1)F^1(dt_2,dx).$$

Third, we show that $\|U_n(\varphi) - \mathbb{P}_n\zeta_\varphi\|_{\mathcal{F}} = o_P\left(n^{-1/2}\right)$, where $\mathbb{P}_n\zeta_\varphi$ is the Hájek projection of $U_n(\varphi)$. It then follows that $P\zeta_\varphi = P\varphi$, and that $P\zeta_\varphi^2 < \infty$, by (S.4) below. Thus, provided that $\{\zeta_\varphi : \varphi \in \mathcal{F}\}$ satisfies some entropy conditions that guarantee it forms a Donsker class of functions,

$$\left\{\sqrt{n}\left(\widehat{\mathbb{P}}_n - P\right)\varphi : \varphi \in \mathcal{F}\right\} \to_d \{\mathbb{G}_P\zeta_\varphi : \varphi \in \mathcal{F}\} \text{ in the space } \ell^\infty(\mathcal{F}),$$

where $\mathbb{G}_P$ is the *Brownian bridge* associated to P, i.e. a centered tight Gaussian process indexed by functions squared integrable with respect to P with covariance,

$$\mathbb{E}\mathbb{G}_P\varphi_1\mathbb{G}_P\varphi_2 = P(\varphi_1 - P\varphi_1)(\varphi_2 - P\varphi_2) = P\varphi_1\varphi_2 - P\varphi_1P\varphi_2,\ P\varphi_j^2 < \infty,$$

for given φ_j such that $P\|\varphi_j\|^2 < \infty$, $j = 1,2$.

LEMMA S1. *Let $\mathcal{F}$ be a class of p-valued vectors of functions, $(t,x) \to \varphi(t,x)$ that admits an envelope function $\Phi \in \mathcal{F}$ such that $P\Phi^2 < \infty$. Let Assumptions 3.1, 3.2, 3.3 hold, and assume*

$$\mathbb{E}\left[\Phi^2(Y,X)\gamma^{(0)^2}(X)\delta\right] < \infty, \tag{S.4}$$

$$\int \|\Phi(t,x)\| S^{1/2}(t) F(dt,dx) < \infty, \tag{S.5}$$

with

$$S(t) = \int_0^{t-} \frac{F_C(d\bar{t})}{[1-F_Y(\bar{t})][1-F_C(\bar{t})]},$$

If for $\varepsilon \in (0,1)$,

$$\sup_Q N_{[\,]}\left(\varepsilon\|\Phi\|_{2,Q}, \mathcal{F}, \|\cdot\|_{2,Q}\right) \le K\left(\frac{1}{\varepsilon}\right)^v, \tag{S.6}$$

with K and v independent of ε, then,

$$\sup_{\varphi\in\mathcal{F}} \|S_n(\varphi) - V_n(\varphi)\|_{\mathcal{F}} = o_P\left(n^{-1/2}\right), \tag{S.7}$$

$$\sup_{\varphi\in\mathcal{F}} \|V_n(\varphi) - \mathbb{P}_n\zeta_\varphi\| = o_P\left(n^{-1/2}\right), \tag{S.8}$$

and

$$\left\{\sqrt{n}\left(\widehat{\mathbb{P}}_n - P\right)\varphi : \varphi \in \mathcal{F}\right\} \to_d \{\mathbb{G}_P\zeta_\varphi : \varphi \in \mathcal{F}\} \text{ in } \ell^\infty(\mathcal{F})^q. \tag{S.9}$$

PROOF. Stute (1996)'s Lemma 5.1 establishes that

$$\begin{aligned}\widehat{\mathbb{P}}_n\varphi &= \int_{\mathbb{R}^{1+k}} \varphi(t,x)\exp\left\{n\int_0^{t-}\ln\left[1+\frac{1}{n\left(1-\tilde{F}_{Yn}(\bar{t})\right)}\right]\tilde{F}^0_{Yn}(d\bar{t})\right\}\tilde{F}^1_{YXn}(dt,dx)\\ &= \frac{1}{n}\sum_{i=1}^{n}\varphi(Y_i,X_i)\,\delta_i\exp\left\{n\int_0^{Y_i-}\ln\left[1+\frac{1}{n\left(1-\tilde{F}_{Yn}(\bar{t})\right)}\right]\tilde{F}^0_{Yn}(d\bar{t})\right\}\end{aligned}$$

Applying the mean value theorem to the exponential term,

$$\begin{aligned}\widehat{\mathbb{P}}_n\varphi &= \frac{1}{n}\sum_{i=1}^{n}\varphi(Y_i,X_i)\,\delta_i\left(\exp\left\{\int_0^{Y_i-}\frac{F^0_Y(d\bar{t})}{1-F_Y(\bar{t})}\right\}\right.\\ &\quad\times\left[1+n\int_0^{Y_i-}\ln\left(1+\frac{1}{n\left(1-\tilde{F}_{Yn}(\bar{t})\right)}\right)\tilde{F}^0_{Yn}(d\bar{t})-\int_0^{Y_i-}\frac{F^0_Y(d\bar{t})}{1-F_Y(\bar{t})}\right]\\ &\quad\left.+\frac{1}{2}e^{\Delta_i}\left\{n\int_0^{Y_i-}\ln\left[1+\frac{1}{n\left(1-\tilde{F}_{Yn}(\bar{t})\right)}\right]\tilde{F}^0_{Yn}(d\bar{t})-\int_0^{Y_i-}\frac{F^0_Y(d\bar{t})}{1-F_Y(\bar{t})}\right\}^2\right)\\ &= \frac{1}{n}\sum_{i=1}^{n}\varphi(Y_i,X_i)\,\gamma^{(0)}(Y_i)\,\delta_i\left[1+B_{in}+C_{in}\right]+\frac{1}{2n}\sum_{i=1}^{n}\varphi(Y_i,X_i)\,\delta_i e^{\Delta_i}\left[B_{in}+C_{in}\right]^2,\end{aligned}$$

where Δ_i is between the two terms in squared brackets, and

$$\begin{aligned}B_{in} &= n\int_0^{Y_i-}\ln\left[1+\frac{1}{n\left(1-\tilde{F}_{Yn}(\bar{t})\right)}\right]\tilde{F}^0_{Yn}(d\bar{t})-\int_0^{Y_i-}\frac{\tilde{F}^0_{Yn}(d\bar{t})}{1-\tilde{F}_{Yn}(\bar{t})},\\ C_{in} &= \int_0^{Y_i-}\frac{\tilde{F}^0_{Yn}(d\bar{t})}{1-\tilde{F}_{Yn}(\bar{t})}-\int_0^{Y_i-}\frac{F^0_Y(d\bar{t})}{1-F_Y(\bar{t})}.\end{aligned}$$

Reasoning as Stute (1995, 1996), we temporarily focus on functions such that

$$\varphi(t,x)=0 \text{ for all } t_0<t \text{ and some } t_0<\tau_Y. \tag{S.10}$$

Remember that $\tau_\xi=\inf(t:\mathbb{P}(\xi\le t)=1)$. So we ensure that all the denominators are bounded away from zero.

Since $a-(a^2/2)\le\ln(1+a)\le a$ for $a\ge 0$,

$$-\frac{1}{2n}\int_0^{Y_i-}\frac{\tilde{F}^0_{Yn}(d\bar{t})}{\left[1-\tilde{F}_{Yn}(\bar{t})\right]^2}\le B_{in}\le 0. \tag{S.11}$$

Thus,

$$\left|\frac{1}{n}\sum_{i=1}^{n}\varphi(Y_i,X_i)\,\gamma(X_i)\,\delta_iB_{in}\right|\le\sup_{t:t<T}\int_0^{t-}\frac{\tilde{F}^0_{Yn}(d\bar{t})}{\left[1-\tilde{F}_{Yn}(\bar{t})\right]^2}\frac{1}{n^2}\sum_{i=1}^{n}\left|\varphi(Y_i,X_i)\right|\left|\gamma(Y_i)\right|$$

and applying a Glivenko-Cantelli theorem for cumulative hazards (e.g., Stute, 1994),

$$\sup_{t:t<t_0} \int_0^{t-} \frac{\tilde{F}^0_{Yn}(d\bar{t})}{\left[1-\tilde{F}_{Yn}(\bar{t})\right]^2} = O(1) \text{ with probability 1,}$$

and, since $\mathcal{F}$ is Glivenko-Cantelli under (S.6), see, e.g., van der Vaart and Wellner (1996)'s Theorem 2.4.1,

$$\sup_{\varphi\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^{n} |\varphi(Y_i, X_i)| \, |\gamma(Y_i)| = O(1) \text{ with probability 1.}$$

Therefore,

$$\sup_{\varphi\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^{n} \varphi(Y_i, X_i)\,\gamma(X_i)\,\delta_i B_{in} \right| = O\left(\frac{1}{n}\right) \text{ with probability 1.}$$

Likewise, under (S.10), (S.11), using the fact that $\mathcal{F}$ is Glivenko-Cantelli, and applying the LIL for cumulative hazard functions on compacta (e.g., Stute, 1994),

$$\sup_{\varphi\in\mathcal{F}} \left| \frac{1}{n}\sum_{i=1}^{n} \varphi(Y_i, X_i)\,\delta_i e^{\Delta_i}\left[B_{in}+C_{in}\right]^2 \right| = O\left(\frac{\ln n}{n}\right) \text{ with probability 1.}$$

Hence,

$$\left\| \widehat{\mathbb{P}}_n\varphi - \frac{1}{n}\sum_{i=1}^{n} \varphi(Y_i, X_i)\,\gamma(X_i)\,\delta_i\left[1+C_{in}\right] \right\|_{\mathcal{F}} = O\left(\frac{\ln n}{n}\right) \text{with probability 1.} \quad \text{(S.12)}$$

We now study the term involving C_{in} in greater detail. Towards this end, note that for $t < Y_{n:n}$,

$$\frac{1}{1-\tilde{F}_{Yn}(t)} = -\frac{1-\tilde{F}_{Yn}(t)}{\left[1-F_Y(t)\right]^2} + \frac{2}{1-F_Y(t)} + \frac{\left[\tilde{F}_{Yn}(t)-F_Y(t)\right]^2}{\left[1-F_Y(t)\right]^2\left[1-\tilde{F}_{Yn}(t)\right]}.$$

Thus,

$$\begin{aligned} C_{in} = &-\int_0^{Y_i-} \frac{1-\tilde{F}_{Yn}(t)}{\left[1-F_Y(t)\right]^2}\tilde{F}^0_{Yn}(dt) + \int_0^{Y_i-} \frac{2}{1-F_Y(t)}\tilde{F}^0_{Yn}(dt) \\ &-\int_0^{Y_i-} \frac{1}{1-F_Y(t)}\tilde{F}^0_{Y}(dt) + \int_0^{Y_i-} \frac{\left[\tilde{F}_{Yn}(t)-F_Y(t)\right]^2}{\left[1-F_Y(t)\right]^2\left[1-\tilde{F}_{Yn}(t)\right]}\tilde{F}^0_{Yn}(dt), \end{aligned} \quad \text{(S.13)}$$

and

$$\frac{1}{n}\sum_{i=1}^{n}\varphi\left(Y_i,X_i\right)\gamma^{(0)}\left(Y_i\right)\delta_i\left(1+C_{in}\right) \tag{S.14}$$

$$= -\int\int\int \varphi\left(t_3,x_3\right)\gamma^{(0)}\left(t_3\right)\frac{1_{\{t_1<t_2,t_1<t_3\}}}{\left[1-F_Y\left(t_1\right)\right]^2}\tilde{F}_{Yn}\left(dt_2\right)\tilde{F}^0_{Yn}\left(dt_1\right)\tilde{F}^1_n\left(dt_3,dx_3\right)$$
$$+2\int\int\frac{1_{\{t_1<t_3\}}\varphi\left(t_3,x_3\right)\gamma^{(0)}\left(t_3\right)}{1-F_Y\left(t_1\right)}\tilde{F}^0_{Yn}\left(dt_1\right)\tilde{F}^1_n\left(dt_3,dx_3\right)$$
$$+\int\varphi\left(t_3,x_3\right)\gamma^{(0)}\left(t_3\right)\left[1-\int\frac{1_{\{t_1<t_3\}}}{1-F_Y\left(t_1\right)}F^0_Y\left(dt_1\right)\right]\tilde{F}^1_n\left(dt_3,dx_3\right)+R^{\varphi}_n$$

where

$$R^{\varphi}_n=\int\int 1_{\{t_1<t_3\}}\varphi\left(t_3,x_3\right)\gamma^{(0)}\left(t_3\right)\frac{\left[\tilde{F}_{Yn}\left(t_1\right)-F_Y\left(t_1\right)\right]^2}{\left[1-F_Y\left(t_1\right)\right]^2}\frac{\tilde{F}^0_{Yn}\left(dt_1\right)}{\left[1-\tilde{F}_{Yn}\left(t_1\right)\right]}\tilde{F}^1_n\left(dt_3,dx_3\right).$$

Since, applying the Glivenko-Cantelli Theorem for hazard functions (Stute, 1994 1994), the LIL for empirical distributions (Finkelstein, 1971), and the uniform LLN for U-processes in Arcones and Giné (1993)'s Corollary 3.5, under (S.6),

$$\sup_{\varphi\in\mathcal{F}}\left|R^{\varphi}_n\right|\leq\sup_{t:t<T}\int_0^t\frac{\tilde{F}_{Yn}\left(dt_1\right)}{\left[1-\tilde{F}_{Yn}\left(t_1\right)\right]}\sup_{t\in\mathcal{T}}\left[\tilde{F}_{Yn}\left(t\right)-F_Y\left(t\right)\right]^2 \tag{S.15}$$
$$\times\left\|\frac{1}{n}\sum_{i=1}^{n}\left|\varphi\left(Y_i,X_i\right)\right|\right\|_{\mathcal{F}}$$
$$=O(1)\cdot O\left(\frac{\ln n}{n}\right)\cdot O(1)=O\left(\frac{\ln n}{n}\right)\text{ with probability 1,}$$

This proves (S.7) under (S.10).

In order to proof (S.8) under (S.10), let

$$\kappa_{1\varphi}\left(z_3\right)=\varphi\left(t_3,x_3\right)\gamma^{(0)}\left(t_3\right)\delta_3\left[1-\int\frac{1_{\{t_1<t_3\}}}{1-F_Y\left(t_1\right)}F^0_Y\left(dt_1\right)\right],$$
$$\kappa_{2\varphi}\left(z_1,z_3\right)=2\varphi\left(t_3,x_3\right)\gamma^{(0)}\left(t_3\right)\delta_3\frac{1_{\{t_1<t_3\}}\left(1-\delta_1\right)}{1-F_Y\left(t_1\right)},$$
$$\kappa_{3\varphi}\left(z_1,z_2,z_3\right)=-\varphi\left(t_3,x_3\right)\gamma^{(0)}\left(t_3\right)\delta_3\frac{1_{\{t_1<t_2,t_1<t_3\}}(1-\delta_1)}{\left[1-F_Y\left(t_1\right)\right]^2}.$$

Recall that

$$U_n\left(\varphi\right)=U^{(1)}_n\left(\kappa_{1\varphi}\right)+U^{(2)}_n\left(\kappa_{2\varphi}\right)+U^{(3)}_n\left(\kappa_{3\varphi}\right),$$

where, for $\left(z_1,z_2,z_3\right)\mapsto g\left(z_1,z_2,z_3\right)$,

$$U^{(m)}_n\left(g\right)=\frac{(n-m)!}{n!}\sum_{\pi}g\left(Z_{i_1},...,Z_{i_m}\right),$$

is the U-process of order m, with π running on all 3-permutations $\pi=\left(i_1,i_2,i_3\right)$ of integers $0\leq i_j\leq n$ with $i_1\neq i_2\neq i_3$. Then, by noticing that $\kappa_{3\varphi}\left(Z_i,Z_i,Z_j\right)=$

$\kappa_{3\varphi}(Z_i, Z_j, Z_i) = \kappa_{3\varphi}(Z_i, Z_i, Z_i) = k_{2\varphi}(Z_i, Z_i) = 0$ a.s., it follows that

$$V_n(\varphi) - U_n(\varphi) = \left[\frac{n!}{n^3(n-3)!} - 1\right] U_n(\kappa_{3\varphi}) + \frac{1}{n^3}\sum\sum_{i \neq j} \kappa_{3\varphi}(Z_j, Z_i, Z_i).$$

Notice that the class $\mathcal{K}_3$ of functions

$$(z_1, z_2, z_3) \mapsto \kappa_{3\varphi}(z_1, z_2, z_3) = -\varphi(t_3, x_3) \cdot g_3(d_1, t_1, t_2, d_3, t_3),$$

with

$$g_3(d_1, t_1, t_2, d_3, t_3) = \gamma^{(0)}(t_3)\, \delta_3 \frac{1_{\{t_1 < t_2, t_1 < t_3\}}(1-\delta_1)}{[1 - F_Y(t_1)]^2}$$

is completely bounded under (S.10). Therefore, $\mathcal{K}_3 = \mathcal{F} \cdot g_3$ has identical bracketing numbers than $\mathcal{F}$, whose bound is given by (S.6). Using an identical argument, the class $\mathcal{K}_2$ of functions $(z_1, z_2) \mapsto \kappa_2(z_1, z_2) = \varphi(t_2, x_2) \cdot g_2(\delta_1, t_1, t_2)$ and the class $\mathcal{K}_1$ of functions $z \mapsto \kappa_2(z) = \varphi(t, x) \cdot g_1(\delta, t)$ have also the same bracketing numbers than $\mathcal{F}$.

Applying Corollary 3.5 of Arcones and Giné (1993), it follows that

$$\sup_{\varphi \in \mathcal{F}} \left\| \frac{1}{n^3}\sum\sum_{i \neq j} \kappa_{3\varphi}(Z_j, Z_i, Z_i) \right\| = O\left(\frac{1}{n}\right) \text{ with probability 1,}$$

$$\sup_{\varphi \in \mathcal{F}} \left\| \left[\frac{n!}{n^3(n-3)!} - 1\right] U_n(\kappa_{3\varphi}) \right\| = O\left(\frac{1}{n}\right) \text{ with probability 1,}$$

and, hence,

$$\|V_n - U_n\|_{\mathcal{F}} = O\left(\frac{1}{n}\right) \text{ with probability 1.} \qquad \text{(S.16)}$$

In order to obtain the Hájek projection of $U_n(\varphi)$, $\hat{U}_n(\varphi)$, define

$$\hat{U}_n^{(1)}(\varphi) = \frac{1}{n}\sum_{i=1}^{n}\left[\kappa_{1\varphi}^{(1)}(Z_i) + \kappa_{1\varphi}^{(2)}(Z_i)\right]$$

$$\mathbb{E}\left[\hat{U}_n^{(2)}(\varphi)\middle| Z_i\right] = \frac{1}{n}\sum_{i=1}^{n}\left[\kappa_{2\varphi}^{(1)}(Z_i) + \kappa_{2\varphi}^{(2)}(Z_i)\right] + \frac{(n-2)}{n}\kappa_{2\varphi}^{(0)},$$

$$\mathbb{E}\left[\hat{U}_n^{(3)}(\varphi)\middle| Z_i\right] = \frac{1}{n}\sum_{i=1}^{n}\left[\kappa_{3\varphi}^{(1)}(Z_i) + \kappa_{3\varphi}^{(2)}(Z_i) + \kappa_{3\varphi}^{(3)}(Z_i)\right] + \frac{n-3}{n}\kappa_{1\varphi}^{(0)},$$

where, for $z = (t, x', d)'$,

$$\kappa_{1\varphi}^{(1)}(z) = \varphi(t, x)\,\gamma(t)\, d$$

$$\kappa_{1\varphi}^{(2)}(z) = -\varphi(t, x)\,\gamma(t)\, d \int \frac{1_{\{t_1 < t\}}}{1 - F_Y(t_1)} F_Y^0(dt_1)$$

$$\kappa_{2\varphi}^{(1)}(z) = \mathbb{E}\kappa_{2\varphi}(z, Z_1) = 2\frac{(1-d)}{1 - F_Y(t)} \int 1_{\{t < t_3\}}\varphi(t_3, x_3)\,\gamma(t_3)\, F^1(dt_3, dx_3),$$

$$\kappa_{2\varphi}^{(2)}(z) = \mathbb{E}\kappa_{2\varphi}(Z_1, z) = 2\varphi(t, x)\,\gamma(t)\, d \int 1_{\{t_1 < t\}} \frac{F_Y^0(t_1)}{1 - F_Y(t_1)}$$

$$\kappa_{2\varphi}^{(0)} = \mathbb{E}\kappa_{2\varphi}(Z_1, Z_2) = 2\int\int \frac{1_{\{t_1 < t_3\}}\varphi(t_3, x_3)\,\gamma(t_3)}{1 - F_Y(t_1)} F_Y^0(t_1)\, F^1(dt_3, dx_3).$$

$$\begin{aligned}
\kappa_{3\varphi}^{(1)}(z) &= \mathbb{E}\kappa_{3\varphi}(z, Z_1, Z_2) \\
&= -\frac{(1-d)}{[1-F_Y(t)]^2} \int\int 1_{\{t<t_2, t<t_3\}} \varphi(t_3, x_3) \gamma(t_3) F_Y(dt_2) F^1(dt_3, dx_3) \\
\kappa_{3\varphi}^{(2)}(z) &= \mathbb{E}\kappa_{3\varphi}(Z_1, z, Z_2) \\
&= -\int\int \frac{1_{\{t_1<t, t_1<t_3\}} \varphi(t_3, x_3) \gamma(t_3)}{[1-F_Y(t_1)]^2} F_Y^0(dt_1) F^1(dt_3, dx_3), \\
\kappa_{3\varphi}^{(3)}(z) &= \mathbb{E}\kappa_{3\varphi}(Z_1, Z_2, z) \\
&= -\varphi(t, x) \gamma(t) d \int\int \frac{1_{\{t_1<t_2, t_1<t\}}}{[1-F_Y(t_1)]^2} F_Y^0(dt_1) F_Y(dt_2) \\
\kappa_{3\varphi}^{(0)} &= \mathbb{E}\kappa_{3\varphi}(Z_1, Z_2, Z_3) \\
&= -\int\int\int \frac{1_{\{t_1<t_2, t_1<t_3\}} \varphi(t_3, x_3) \gamma(t_3)}{[1-F_Y(t_1)]^2} F_Y^0(dt_1) F_Y(dt_2) F^1(dt_3, dx_3)
\end{aligned}$$

Thus, the $U_n^{(j)}(\varphi)'s$ Hájek projections are

$$\hat{U}_n^{(j)}(\varphi) = \sum_{i=1}^n \mathbb{E}\left[U_n^{(j)}(\varphi) \middle| Z_i \right] - (n-1) \mathbb{E} U_n^{(j)}(\varphi), \ j = 1, 2, 3$$

with

$$\mathbb{E}\left[U_n^{(j)}(\varphi) \middle| Z_i \right] = \frac{(n-j)}{n} \kappa_j^{(0)} + \frac{1}{n} \sum_{\ell=1}^j \kappa_j^{(\ell)}(Z_i), \ j = 1, 2, 3.$$

Therefore,

$$\hat{U}_n^{(j)}(\varphi) = (1-j) \kappa_j^{(0)} + \frac{1}{n} \sum_{i=1}^n \sum_{\ell=1}^j \kappa_j^{(\ell)}(Z_i), \ j = 1, 2, 3.$$

Taking into account that

$$\begin{aligned}
\kappa_{2\varphi}^{(2)}(z) &= -2\kappa_{3\varphi}^{(3)}(z) = -2\kappa_{1\varphi}^{(2)}(z) \\
\frac{\kappa_{2\varphi}^{(1)}(z)}{2} &= -\kappa_{3\varphi}^{(1)}(z) = -\frac{(1-d)}{1-F_Y(t)} \gamma_\varphi^{(1)}(t) \\
\kappa_{3\varphi}^{(2)}(z) &= -\gamma_\varphi^{(2)}(t) \\
\kappa_{2\varphi}^{(0)} &= -2\kappa_{3\varphi}^{(0)} = 2\mathbb{E}\hat{U}_n(\varphi) = 2P\varphi,
\end{aligned}$$

it follows that the $U_n(\varphi)'s$ Hájek projection is

$$\begin{aligned}\hat{U}_n(\varphi) &= \sum_{j=1}^{3} \hat{U}_n^{(j)}(\varphi)\\ &= \frac{1}{n}\sum_{i=1}^{n}\left\{\kappa_{1\varphi}^{(1)}(Z_i) - \kappa_{3\varphi}^{(1)}(Z_i) + \kappa_{3\varphi}^{(2)}(Z_i)\right\}\\ &= \frac{1}{n}\sum_{i=1}^{n}\left\{\varphi(Y_i, X_i)\gamma(Y_i)\delta_i + \gamma_{\varphi}^{(1)}(Y_i)(1-\delta_i) - \gamma_{\varphi}^{(2)}(Y_i)\right\}.\\ &= \frac{1}{n}\sum_{i=1}^{n}\zeta_{\varphi}(Y_i, X_i, \delta_i).\end{aligned}$$

Given that the class of functions $\mathcal{K}_1$, $\mathcal{K}_2$ and $\mathcal{K}_3$ are of the form $\mathcal{F}{\cdot}g$ as they have bounded envelope and identical bracketing numbers bounds as $\mathcal{F}$ under (S.10), it follows that, $\mathcal{F}{\cdot}g$ is Donsker under (S.6). The classes $\mathcal{G}_1$ and $\mathcal{G}_2$ of functions $(t,d) \mapsto \gamma_{\varphi}^{(1)}(t)(1-d)$ and $t \mapsto \gamma_{\varphi}^{(2)}(t)$, respectively, are of bounded variation, i.e. they are absolutely bounded and can be expressed as the difference of two increasing functions, it follows from van der Vaart and Wellner (1996) Theorem 2.7.5 that

$$\log N_{[\,]}\left(\varepsilon, \mathcal{G}_j, \|\cdot\|_{2,Q}\right) \leq K\left(\frac{1}{\varepsilon}\right),\ j = 1, 2,\ 0 < \varepsilon < 1. \tag{S.17}$$

Using Kosorok (2008) Lemma 9.18, for any probability measure Q, such that $\|\Phi\|_{Q,r} < \infty$,

$$N\left(\varepsilon, \mathcal{F}, \|\cdot\|_{r,Q}\right) \leq N_{[\,]}\left(\varepsilon, \mathcal{F}, \|\cdot\|_{r,Q}\right),$$

and, under (S.6),

$$\ln\sup_Q N\left(\varepsilon\|\Phi\|_{2,Q}, \mathcal{F}, \|\cdot\|_{2,Q}\right) \leq \ln\sup_Q N_{[\,]}\left(\varepsilon\|\Phi\|_{2,Q}, \mathcal{F}, \|\cdot\|_{2,Q}\right) \lesssim \ln(1/\varepsilon), \tag{S.18}$$

where

$$\int_0^1 \log(1/\varepsilon)\, d\varepsilon = \int_0^\infty u^{1/2}e^{-u}du = \frac{\sqrt{\pi}}{2} < \infty.$$

Taking into account that $\mathcal{K}_2$ and $\mathcal{K}_2$ have uniform bracketing entropies bounded as (S.18), condition (1.3) in Arcones and Giné (1995) is satisfied, and applying their law of iterated LIL for canonical U-processes (see also de la Peña and Giné, 1999 Theorem 5.4.1 and the discussion on page 256)

$$\begin{aligned}\mathbb{E}\left[\sup_{n\in\mathbb{N}}\left\|U_n^{(m)}(\kappa_{m\varphi}) - \hat{U}_n^{(m)}(\kappa_{m\varphi})\right\|_{\mathcal{K}_m}\right] &\leq [2n\ln\ln(n)]^{m/2}\frac{(n-m)!}{n!}\|\Phi\|_{2,P}\\ &= O\left(\left(\frac{\ln\ln n}{n}\right)^{m/2}\right),\ m = 2, 3.\end{aligned}$$

Therefore, under (S.10),

$$\left\|U_n(\varphi) - \hat{U}_n(\varphi)\right\|_{\mathcal{F}} = O_p\left(\frac{\ln\ln n}{n}\right), \tag{S.19}$$

which establishes (S.8).

In order to prove (S.7) and (S.8) for general $\varphi \in \mathcal{F}$, which possibly do not satisfy (S.10), assume first that F_Y if continuous. For given $\varepsilon > 0$ choose an $\tilde{\varphi} \in \mathcal{F}$ satisfying (S.10) and such that $\varphi_1 = (\varphi - \tilde{\varphi}) \in \mathcal{F}$ and

$$K_1 \varepsilon \leq \sup_{\varphi_1 \in \mathcal{F}} \mathbb{E}\left[\varphi_1^2 (Y, X) \gamma^{(0)^2} (X) \delta\right] \leq K_2 \varepsilon, \tag{S.20}$$

$$\sup_{\varphi_1 \in \mathcal{F}} \int \|\varphi_1 (y, x)\| S^{1/2} (y) F(dy, dx) \leq \varepsilon, \tag{S.21}$$

for finite positive constants K_1, K_2, which is possible under (S.4) and (S.5). Then, we prove that

$$\sup_{\varphi_1 \in \mathcal{F}} \left\| \frac{1}{\sqrt{n}} \int \varphi_1 (t, x) \left(\hat{F}_n - F\right) (dt, dx) \right\| = O_P \left(\varepsilon^{1/2} \sqrt{\log\left(1/\varepsilon^{1/2}\right)}\right) \tag{S.22}$$

for some $\delta > 0$. In view that (S.15), (S.16) and (S.19), a Cramér-Slutzky type argument will then complete the proof of the lemma.

Since,

$$\hat{S}_n (\varphi_1) = \frac{1}{n} \sum_{i=1}^{n} \varphi_1 (Y_i, X_i) \delta_i \gamma^{(0)} (Y_i) \exp (B_{in} + C_{in}),$$

we have that,

$$\begin{aligned} n^{1/2} \int \varphi_1 (t, x) \left(\hat{F}_n - F\right) (dt, dx) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\varphi_1 (Y_i, X_i) \delta_i \gamma^{(0)} (Y_i) - \mathbb{E} \varphi_1 (T, X)\right] \\ &\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \varphi_1 (Y_i, X_i) \delta_i \gamma^{(0)} (Y_i) \left[\exp (B_{in} + C_{in}) - 1\right] \\ &= A_{n1} (\varphi_1) + A_{n2} (\varphi_1). \end{aligned} \tag{S.23}$$

In order to show that $\|A_{n1} (\varphi_1)\|_{\mathcal{F}} = O_P \left(\varepsilon^{1/2} \sqrt{\log\left(1/\varepsilon^{1/2}\right)}\right)$, we apply Theorem 3.1 of Giné and Koltchinskii (2006) (GK). Notice that, by (S.6), for all probability measure Q

$$\log N_{[\,]} \left(\varepsilon, \mathcal{F}, \|\cdot\|_{2,Q}\right) \lesssim \log \left(\frac{K \|\Phi\|_{2,Q}}{\varepsilon}\right),$$

for $0 < \varepsilon \leq \|F\|_{2,Q}$. Also, note that $u \mapsto H(u) = v \log (Vu)$, is a slowly varying function, where $V = K \|F\|_{2,Q}$. Therefore, by GK's Theorem 3.1, for n large,

$$\mathbb{E} \sup_{\varphi \in \mathcal{F}} \|A_{n1} (\varphi_1)\| \leq \sigma \sqrt{v \log (V/\sigma)}$$

where $\sigma^2 = \sup_{\varphi_1 \in \mathcal{F}} \mathbb{E}\left[\varphi_1^2 (Y, X) \gamma^{(0)^2} (X) \delta\right]$. Hence, by (S.20), $\sup_{\varphi_1 \in \mathcal{F}} \mathbb{E} \|A_{n1} (\varphi_1)\| \leq \varepsilon^{1/2} \sqrt{v \log (V/\varepsilon)}$ which proves that

$$\|A_{n1} (\varphi_1)\|_{\mathcal{F}} = O_P \left(\varepsilon^{1/2} \sqrt{\log\left(1/\varepsilon^{1/2}\right)}\right) \tag{S.24}$$

Now,

$$\|A_{n2}(\varphi_1)\|_{\mathcal{F}} \le \sup_{\varphi_1 \in \mathcal{F}} \left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |\varphi_1(Y_i, X_i)| \, \delta_i \gamma^{(0)}(Y_i) \, |B_{ni} + C_{ni}| \exp(|B_{ni}| + |C_{ni}|) \right\|. \tag{S.25}$$

By (S.11),

$$|B_{ni}| \le \frac{1}{n(1 - H_n(Y_i-))} \le 1. \tag{S.26}$$

Stute (1995), page 436, shows that $\sup_{1 \le i \le n} |C_{ni}| = O_P(1)$. Thus, the exponent in (S.25) is an $O_P(1)$ uniformly in $i = 1, ..., n$. Therefore, it remains to bound

$$A_{n3}(\varphi_1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |\varphi_1(Y_i, X_i)| \, \delta_i \gamma^{(0)}(Y_i) \, |B_{ni}|,$$
$$A_{n4}(\varphi_1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} |\varphi_1(Y_i, X_i)| \, \delta_i \gamma^{(0)}(Y_i) \, |C_{ni}|,$$

By (S.11) and (S.26),

$$\begin{aligned}
\|A_{n3}(\varphi_1)\|_{\mathcal{F}} &\le \left\| \frac{1}{n} \sum_{i=1}^{n} |\varphi_1(Y_i, X_i)| \, \delta_i \gamma^{(0)}(Y_i) \left[\int_0^{Y_i-} \frac{\tilde{F}_Y^0(dt)}{\left[1 - \tilde{F}_Y(t)\right]^2} \right]^{1/2} \right\|_{\mathcal{F}} \\
&\lesssim \frac{1}{n} \sum_{i=1}^{n} \Phi(Y_i, X_i) \, \delta_i \gamma^{(0)}(Y_i) \left[\int_0^{Y_i-} \frac{\tilde{F}_Y^0(dt)}{\left[1 - F_Y(t)\right]^2} \right]^{1/2}
\end{aligned}$$

since $\sup_{t < Y_{n:n}} (1 - F_Y) / \left(1 - \tilde{F}_Y\right)(z) = O_P(1)$ (see, e.g., Shorack and Wellner, 1986, page 415). Thus, by a symmetry argument and applying Hölder inequality

$$\begin{aligned}
\mathbb{E}\|A_{n3}(\varphi_1)\|_{\mathcal{F}} &\le \mathbb{E}\left\{ \Phi(Y_1, X_1) \, \delta_1 \gamma^{(0)}(Y_1) \cdot \mathbb{E}\left(\left[\int_0^{Y_1-} \frac{\tilde{F}_Y^0(dt)}{\left[1 - F_Y(t)\right]^2} \right]^{1/2} \middle| Y_1 \right) \right\} \\
&\le \mathbb{E}\left\{ \Phi(Y_1, X_1) \, \delta_1 \gamma^{(0)}(Y_1) \cdot \mathbb{E}^{1/2}\left(\int_0^{Y_1-} \frac{\tilde{F}_Y^0(dt)}{\left[1 - F_Y(t)\right]^2} \middle| Y_1 \right) \right\} \\
&= \sqrt{\frac{n-1}{n}} \mathbb{E}\left\{ \Phi(Y_1, X_1) \, \delta_1 \gamma^{(0)}(Y_1) \cdot \mathbb{E}^{1/2}\left(\frac{(1-\delta) 1_{\{Y_2 < Y_1\}}}{\left[1 - F_Y(Y_2)\right]^2} \right) \right\} \\
&= \sqrt{\frac{n-1}{n}} \mathbb{E}\left\{ \Phi(Y_1, X_1) \, \delta_1 \gamma^{(0)}(Y_1) \cdot \left(\int_0^{Y_1-} \frac{F_Y^0(dt)}{\left[1 - F_Y(t)\right]^2} \right)^{1/2} \right\} \\
&= \sqrt{\frac{n-1}{n}} \mathbb{E}\left\{ \Phi(Y_1, X_1) \, \delta_1 \gamma^{(0)}(Y_1) \cdot S^{1/2}(Y_1) \right\} \\
&\le \int \Phi(t, x) \cdot S^{1/2}(t) \, F(dt, dx) \\
&\le \varepsilon.
\end{aligned}$$

by (S.21). Likewise,

$$\begin{aligned}\|A_{n3}(\varphi_1)\|_{\mathcal{F}} &\le \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Phi(Y_i,X_i)\,\delta_i\gamma(Y_i)\,|C_{ni}| \\ &\le \int \Phi(t,x)\cdot S^{1/2}(t)\,F(dt,dx) \\ &\le \varepsilon,\end{aligned}$$

using the bounds to (2.11), (2.12) and (2.13) in Stute (1995). Also use the arguments in Stute (1995) page 437 and 438 to show that the lemma also follows for discontinuous $F_{Y,X}$. This completes the proof of (S.7) and (S.8).

Let $\mathcal{E}$ be the class of functions $(t,x,d)\mapsto\zeta_{\varphi}(t,x,d)$. That is, $\mathcal{E}=\mathcal{F}\cdot g+\mathcal{G}_1+\mathcal{G}_2$. By lemma 9.25 (i) in Kosorok (2008), for any probability measure Q and $1\le r\le\infty$,

$$\begin{aligned}N_{[\,]}\left(4\varepsilon,\mathcal{E},\|\cdot\|_{r,Q}\right) &\le N_{[\,]}\left(2\varepsilon,\mathcal{F}\cdot g,\|\cdot\|_{r,Q}\right)\cdot N_{[\,]}\left(2\varepsilon,\mathcal{G}_1+\mathcal{G}_2,\|\cdot\|_{r,Q}\right) \\ &\le N_{[\,]}\left(2\varepsilon,\mathcal{F}\cdot g,\|\cdot\|_{r,Q}\right)\cdot N_{[\,]}(\varepsilon,\mathcal{G}_1,Q)\cdot N_{[\,]}\left(\varepsilon,\mathcal{G}_2,\|\cdot\|_{r,Q}\right) \\ &\lesssim K\left(\frac{1}{\varepsilon}\right)^{v+2},\end{aligned}$$

with K a constant independent of ε and v. Therefore,

$$\int_0^1\sqrt{\sup_Q\ N_{[\,]}\left(\varepsilon\|\Phi\|_{2,Q},\mathcal{F},\|\cdot\|_{2,Q}\right)}d\varepsilon\lesssim\int_0^1\sqrt{\log(1/\varepsilon)}d\varepsilon<\infty,$$

and $\mathcal{E}$ is Donsker, which proves (S.9).□

PROOF. (PROOF OF THEOREM 4.2) This proof extends Theorem 5.21 in van der Vaart (1998) to the current circumstances, using the Kaplan-Meier empirical measure $\hat{\mathbb{P}}_n$ rather than standard empirical measure $\mathbb{P}_n$.

The score function is

$$\frac{\partial}{\partial\theta}\hat{Q}_{n,t}(\theta)=\widehat{\mathbb{P}}_n\psi_{\theta,t}$$

with

$$\psi_{\theta,t}(y,x)=\frac{1_{\{y\le t\}}-\Lambda(\mathbf{x}'\theta)}{\Lambda(\mathbf{x}'\theta)\left[1-\Lambda(\mathbf{x}'\theta)\right]}\lambda(\mathbf{x}'\theta)\mathbf{x}, \tag{S.27}$$

Since Λ and λ are Lipschitz continuous, and $\Lambda\in(0,1)$, the function $(y,x)\mapsto\psi_{\theta,t}(y,x)$ satisfies that, for θ_1 and θ_2 in a neighborhood of θ_0,

$$\|\psi_{\theta_1,t}(t,x)-\psi_{\theta_2,t}(t,x)\|\le K\dot{\psi}(x)\|\theta_1-\theta_2\|, \tag{S.28}$$

where $\dot{\psi}(x)=\|x\|$ and K is an universal constant. By assumption $P\|X\|^2<\infty$, Theorem 4.1, and by the dominated convergence theorem,

$$\begin{aligned}\left(\frac{\partial}{\partial\theta'}P\psi_{\theta,t}\right)_{\theta=\theta_0(t)} &= \left(P\frac{\partial}{\partial\theta'}\psi_{\theta,t}\right)_{\theta=\theta_0(t)} \\ &= \mathbb{E}\left\{h\left(X'\theta_0\left(t\right)\right)\left[\Lambda\left(X'\theta_0\left(t\right)\right)-1_{\{Y\leq t\}}\right]\right. \\ &\quad \left.-H\left(X'\theta_0\left(t\right)\right)\lambda\left(X'\theta_0\left(t\right)\right)XX'\right\} \\ &= -\mathcal{I}_0\left(t\right), \end{aligned} \tag{S.29}$$

with $H=\lambda/\Lambda\left(1-\Lambda\right)$ and $h\left(u\right)=dH\left(u\right)/du$.

For a fixed function f, we abbreviate $\sqrt{n}\left(\widehat{\mathbb{P}}_n-P\right)f$ and $\sqrt{n}\left(\mathbb{P}_n-P\right)f$ to $\widehat{\mathbb{G}}_nf$ and $\mathbb{G}_nf$, respectively. Next, we show that

$$\widehat{\mathbb{G}}_n\psi_{\hat{\theta}_n(t),t}-\widehat{\mathbb{G}}_n\psi_{\theta_0(t),t}=o_P\left(1\right). \tag{S.30}$$

Since by Lemma S1,

$$\widehat{\mathbb{G}}_n\psi_{\hat{\theta}_n(t),t}-\widehat{\mathbb{G}}_n\psi_{\theta_0(t),t}=\mathbb{G}_n\left(\zeta_{\psi_{\hat{\theta}_n(t),t}}-\zeta_{\psi_{\theta_0(t),t}}\right)+o_P\left(1\right),$$

it suffices to show that

$$\mathbb{G}_n\left(\zeta_{\psi_{\hat{\theta}_n(t),t}}-\zeta_{\psi_{\theta_0(t),t}}\right)=o_P\left(1\right). \tag{S.31}$$

First, we check that the map $z\mapsto\zeta_{\psi_{\theta,t}}\left(z\right)$, $z=\left(y,x',d\right)'$ satisfies a Lipschitz condition. From (S.28),

$$\left\|\zeta_{\psi_{\theta_1,t}}\left(z\right)-\zeta_{\psi_{\theta_2,t}}\left(z\right)\right\|\leq K\cdot\zeta_{\dot{\psi}}\left(z\right)\cdot\left\|\theta_1-\theta_2\right\|, \tag{S.32}$$

and $P\zeta_{\dot{\psi}}^2<\infty$ by Assumptions 4.2 and 4.3.

For a non-random sequence $\hat{\theta}_n\left(t\right)$, (S.31) is immediate since the mean of the random variable on the left hand side of (S.31) has mean zero and variance bounded by $P\left\|\zeta_{\psi_{\hat{\theta}_n(t),t}}-\zeta_{\psi_{\theta_0(t),t}}\right\|^2\leq K\cdot\left\|\hat{\theta}_n\left(t\right)-\theta_0\left(t\right)\right\|P\zeta_{\dot{\psi}}^2$ and, hence, converge to zero, since $\hat{\theta}_n\left(t\right)$ is consistent. For random $\hat{\theta}_n\left(t\right)$, we first check that the bracketing numbers of the class of functions $z\mapsto\zeta_{\psi_{\theta,t}}\left(z\right)$ satisfy (S.6) and, hence, is Donsker. This is the case, since the class $\mathcal{E}$ of functions $\left(y,x,d\right)\mapsto\zeta_{\psi_{\theta,t}}\left(y,x,d\right)$ satisfies (S.32) and, by VV's Example 19.7,

$$N_{[\,]}\left(\varepsilon\left\|\zeta_{\dot{\psi}}\right\|_{2,Q},\mathcal{E},\left\|\cdot\right\|_{2,Q}\right)\leq K\left(\frac{diam\Theta}{\varepsilon}\right)^p<\infty \tag{S.33}$$

for all Q such that $Q\zeta_{\dot{\psi}}^2<\infty$. And taking into account that, by (S.32) and the dominated convergence theorem,

$$P\left\|\zeta_{\psi_{\hat{\theta}_n(t),t}}-\zeta_{\psi_{\theta_0(t),t}}\right\|^2\rightarrow_p 0,$$

we can apply VV's Lemma 19.24 to prove (S.31), and hence (S.30).

The rest of the proof follows in the lines of VV's Theorem 5.21. We provide the details for completeness.

By definition of $\hat{\theta}_n(t)$ and $\theta_0(t)$,

$$\begin{aligned}\widehat{\mathbb{G}}_n\psi_{\hat{\theta}_n(t),t} &= -\sqrt{n}P\psi_{\hat{\theta}_n(t),t} + o_P(1) \\ &= \sqrt{n}P\left(\psi_{\theta_0(t),t} - \psi_{\hat{\theta}_n(t),t}\right) + o_P(1) \\ &= \sqrt{n}P\left(\psi_{\theta_0(t),t} - \psi_{\theta_0(t)+(\hat{\theta}_n(t)-\theta_0(t)),t}\right) + o_P(1) \\ &= -\sqrt{n}\mathcal{I}_0(t)\left(\theta_0(t) - \hat{\theta}_n(t)\right) + \sqrt{n}\cdot o_P\left(\left\|\hat{\theta}_n(t) - \theta_0(t)\right\|\right),\end{aligned} \tag{S.34}$$

where the last equality follows from (S.29) and using the differentiability of the map $\theta \to P\psi_{\theta,t}$ at $\theta_0(t)$, i.e.,

$$P\left(\psi_{\theta_0(t)+h,t} - \psi_{\theta_0(t),t}\right) = \left(\frac{\partial}{\partial\theta'}P\psi_{\theta,t}\right)_{\theta=\theta_0(t)} \cdot h + o(\|h\|) \text{ as } h \to 0.$$

Then, by (S.30), (S.31), and (S.34),

$$\begin{aligned}-\sqrt{n}\mathcal{I}_0(t)\left(\theta_0(t) - \hat{\theta}_n(t)\right) + \sqrt{n}\cdot o_P\left(\left\|\theta_0(t) - \hat{\theta}_n(t)\right\|\right) \\ = \widehat{\mathbb{G}}_n\psi_{\theta_0(t),t} + o_P(1) \\ = \mathbb{G}_n\zeta_{\psi_{\theta_0(t),t}} + o_P(1)\end{aligned} \tag{S.35}$$

by Lemma S1.

Because of the invertibility of $\mathcal{I}_0(t)$

$$\begin{aligned}\sqrt{n}\left\|\hat{\theta}_n(t) - \theta_0(t)\right\| &\le \left\|\mathcal{I}_0^{-1}(t)\right\|\sqrt{n}\left\|\mathcal{I}_0(t)\left(\hat{\theta}_n(t) - \theta_0(t)\right)\right\| \\ &= O_P(1) + o_P\left(\sqrt{n}\left\|\theta_0(t) - \hat{\theta}_n(t)\right\|\right).\end{aligned}$$

This implies that the left side is bounded in probability and, hence, $\hat{\theta}_n(t)$ is $\sqrt{n}$-consistent. Inserting this into (S.35), we obtain that $\sqrt{n}\mathcal{I}_0(t)\left(\hat{\theta}_n(t) - \theta_0(t)\right) = \mathbb{G}_n\zeta_{\psi_{\theta_0(t),t}} + o_P(1)$. The theorem follows by multiplying by the inverse $\mathcal{I}_0^{-1}(t)$ on both left and right hand sides. Because matrix multiplication is a continuous map, the remainder term still converges to zero.

Therefore, applying the CLT and a Cramér-Wold device, given fixed times $t_1, ..., t_m$,

$$\left\{\sqrt{n}\left(\hat{\theta}_n(t_j) - \theta_0(t_j)\right)\right\}_{j=1}^{m} \to_d \left\{\mathcal{I}_0^{-1}(t)\cdot Z(t_j)\right\}_{j=1}^{m},$$

where $\{Z(t_j)\}_{j=1}^m$ are $p\times 1$ Gaussian random vectors with zero mean and covariance matrix

$$\mathbb{E}Z(t_j)Z(t_\ell) = \Omega(j,\ell),$$

where $\Omega(j,\ell) = P\zeta_{\psi_{\theta_0(t_j),t_j}}\zeta'_{\psi_{\theta_0(t_\ell),t-\ell}}$. □

Proof. (Proof of Corollary 4.1) First notice that

$$\hat{\Omega}_n(j,m) = \mathbb{P}_n\zeta_{\psi_{\theta_0(t_j),t_j}}\zeta'_{\psi_{\theta_0(t_m),t_m}} \text{ estimates } \Omega(j,m) = P\zeta_{\psi_{\theta_0(t_j),t_j}}\zeta'_{\psi_{\theta_0(t_m),t_m}}$$

and

$$\widehat{\mathcal{I}}_n(t) = \widehat{\mathbb{P}}_n\psi_{\theta,t}\psi'_{\theta,t} \text{ estimates } \mathcal{I}_0(t) = P\psi_{\theta,t}\psi'_{\theta,t}.$$

Let $\mathcal{F}$ and $\mathcal{E}$ be the classes of functions $(y,x) \mapsto \psi_{\theta_0(t),t}(y,x)$ and $(y,x,d) \mapsto \zeta_{\psi_{\theta,t}}(y,x,d)$, respectively. Then, by Theorem 9.23 and Lemma 9.25 (ii) in Kosorok (2008) and VV's Example 19.7, for any probability measure Q and $1 \leq r \leq \infty$, for any $\varepsilon \in (0, diam\Theta)$,

$$N_{[\,]}\left(2\varepsilon, \mathcal{F}\cdot\mathcal{F}, \|\cdot\|_{1,Q}\right) \leq \left[N_{[\,]}\left(\varepsilon, \mathcal{F}, \|\cdot\|_{1,Q}\right)\right]^2 \leq \left[N\left(\varepsilon, \Theta, \|\cdot\|\right)\right]^2 \leq K\left(\frac{diam\Theta}{\varepsilon}\right)^{2p} < \infty,$$

and

$$N_{[\,]}\left(2\varepsilon, \mathcal{E}\cdot\mathcal{E}, \|\cdot\|_{1,Q}\right) \leq \left[N_{[\,]}\left(\varepsilon, \mathcal{E}, \|\cdot\|_{1,Q}\right)\right]^2 \leq \left[N\left(\varepsilon, \Theta, \|\cdot\|\right)\right]^2 \leq K\left(\frac{diam\Theta}{\varepsilon}\right)^{2p} < \infty.$$

Therefore, $\mathcal{F}\cdot\mathcal{F}$ and $\mathcal{E}\cdot\mathcal{E}$ satisfy (S.6) in Lemma S1, and

$$\sup_{\theta\in\Theta}\left\|\left(\mathbb{P}_n - P\right)\zeta_{\psi_{\theta,t_j}}\zeta'_{\psi_{\theta,t_m}}\right\| \to 0 \text{ with probability 1,}$$
$$\sup_{\theta\in\Theta}\left\|\left(\widehat{\mathbb{P}}_n - P\right)\psi_{\theta,t_j}\psi'_{\theta,t_m}\right\| \to 0 \text{ with probability 1,}$$

which implies that

$$\left(\mathbb{P}_n - P\right)\zeta_{\psi_{\hat{\theta}_n(t_j),t_j}}\zeta'_{\psi_{\hat{\theta}_n(t_m),t_m}} \to 0 \text{ with probability 1,}$$
$$\left(\widehat{\mathbb{P}}_n - P\right)\psi_{\hat{\theta}_n(t_j),t_j}\psi'_{\hat{\theta}_n(t_m),t_m} \to 0 \text{ with probability 1.}$$

By (S.28) and (S.30), Theorem 4.1 and applying the dominated convergence theorem,

$$\left\|P\left(\zeta_{\psi_{\hat{\theta}_n(t_m),t_m}}\zeta'_{\psi_{\hat{\theta}_n(t_j),t_j}} - \zeta_{\psi_{\theta_0(t_m),t_m}}\zeta'_{\psi_{\theta_0(t_j),t_j}}\right)\right\| \to 0 \text{ with probability 1,}$$

$$\left\|P\left(\psi_{\hat{\theta}_n(t_j),t_j}\psi'_{\hat{\theta}_n(t_m),t_m} - P\psi_{\theta_0(t_j),t_j}\psi'_{\theta_0(t_m),t_m}\right)\right\| \to 0 \text{ with probability 1,}$$

concluding the proof.□

Proof. (Proof of Theorem 4.3) First notice that

$$\begin{aligned}\sqrt{n}\left(\hat{\theta}^*_n(t) - \hat{\theta}_n(t)\right) &= \widehat{\mathcal{I}}_n^{-1}(t)\sqrt{n}\mathbb{P}^*_n\zeta_{\psi_{\hat{\theta}_n(t),t}} \\ &= \left[\mathcal{I}_0^{-1}(t) + o(1)\right]\sqrt{n}\mathbb{P}^*_n\zeta_{\psi_{\hat{\theta}_n(t),t}} \text{ with probability 1}\end{aligned}$$

by Corollary 4.1, where $\mathbb{P}^*_n = n^{-1}\sum_{i=1}^n V_i\delta_{Z_i}$, and δ_z is the Dirac measure. Then, using the same arguments as in the proof of Stute et al. (1998)'s Theorem (under H_0), and applying a Cramér-Wold device, with probability 1, $\left\{\sqrt{n}\mathbb{P}^*_n\zeta_{\psi_{\hat{\theta}_n(t_j),t_j}}\right\}_{j=1}^m$ converges in distribution to $\{Z(t_j)\}_{j=1}^m$ under the bootstrap law. □

Proof. (Proof of Theorem 4.4) We show that the bracketing numbers of the class

$\mathcal{S}$ of score functions $(y,x,d) \mapsto \psi_{\theta t}(y,x,d)$ in (S.27) satisfy condition (S.6) in Lemma S1, and hence

$$\sqrt{n}\left(\widehat{\mathbb{P}}_n - P\right)\psi_{\theta t} \text{ converges in distribution to } \mathbb{G}_P\psi_{\theta t} \text{ in } \ell^{\infty}\left(\Theta \times \mathcal{T}_0\right)^{k+1}. \tag{S.36}$$

Write the class as $\mathcal{S} = \mathcal{S}_1 \cdot (\mathcal{S}_2 - \mathcal{S}_3) \cdot b$, where $\mathcal{S}_1$ is the class of functions $(y,x) \mapsto H(x'\theta)$ with $H = \lambda / [\Lambda(1-\Lambda)]$, $\mathcal{S}_2$ is the class of functions $x \mapsto \Lambda(x'\theta)$, $\mathcal{B}_3$ is the class of indicator functions $y \mapsto 1_{\{y \leq t\}}$, and b is the fixed function $b(x) = x$.

Since H is bounded and continuously differentiable, and Θ is compact, we have that $\|H(x'\theta_1) - H(x'\theta_2)\| \leq K\|x\| \cdot \|\theta_1 - \theta_2\|$ for all $\theta_1, \theta_2 \in \Theta$, where K is a generic constant. Therefore, by VV's Example 19.7, for any $\varepsilon \in (0, diam\Theta)$, any probability measure Q, such that $Q\|x\|^2 < \infty$,

$$N_{[\,]}\left(\|b\|_{2,Q}\,\varepsilon, \mathcal{S}_1, \|\cdot\|_{2,Q}\right) \leq K\left(diam\Theta/\varepsilon\right)^p,$$

where K is independent of ε and p . Likewise, the $\mathcal{S}_2's$ bracketing numbers are of the same order. The indicator functions' class $\mathcal{S}_3$ satisfies, $N_{[\,]}\left(\varepsilon, \mathcal{S}_3, \|\cdot\|_{2,Q}\right) \leq K(1/\varepsilon)^2$ for any Q. See VV's Example 19.6.

Therefore, applying Kosorok (2008)'s Lemma 9.25, for any measure Q,

$$\begin{aligned} N_{[\,]}\left(4\|b\|_{2,Q}\,\varepsilon, \mathcal{S}, \|\cdot\|_{2,Q}\right) &\leq N_{[\,]}\left(2\|b\|_{2,Q}\,\varepsilon, \mathcal{S}_1, \|\cdot\|_{2,Q}\right) \cdot N_{[\,]}\left(2\|b\|_{2,Q}\,\varepsilon, (\mathcal{S}_2 - \mathcal{S}_3), \|\cdot\|_{2,Q}\right) \\ &\leq N_{[\,]}\left(2\|b\|_{2,Q}\,\varepsilon, \mathcal{S}_1, \|\cdot\|_{2,Q}\right) \cdot N_{[\,]}\left(\|b\|_{2,Q}\,\varepsilon, \mathcal{S}_2, \|\cdot\|_{2,Q}\right) \\ &\quad \cdot N_{[\,]}\left(\|b\|_{2,Q}\,\varepsilon, \mathcal{S}_1, \|\cdot\|_{2,Q}\right) \cdot N_{[\,]}\left(\|b\|_{2,Q}\,\varepsilon, \mathcal{S}_3, \|\cdot\|_{2,Q}\right) \\ &\leq K\left(\frac{1}{\varepsilon}\right)^{2p+2}. \end{aligned}$$

Hence, by Lemma S1,

$$\sup_{\theta \in \Theta, t \in \mathcal{T}_0}\left\|\widehat{\mathbb{P}}_n\psi_{\theta t} - \mathbb{P}_n\zeta_{\psi_{\theta t}}\right\| = o_P\left(\frac{1}{\sqrt{n}}\right),$$

and $(S.36)$ holds from Chernozhukov et al. (2013)'s Theorem 5.2, since Assumptions (4.2)-(4.4) are sufficient for Chernozhukov et al. (2013)'s Assumption DR. □

Proof. (Proof of Theorem 4.5) Define the weighted bootstrap empirical measure $\mathbb{P}_n^* = n^{-1}\sum_{i=1}^n \delta_{X_i} V_i$. Since

$$\hat{\theta}_n^*(t) = \hat{\theta}_n(t) + \widehat{\mathcal{I}}_n^{-1}(t)\,\mathbb{P}_n^*\zeta_{\psi_{\hat{\theta}_n(t)t}},$$

we must first show that

$$\sup_{t \in \mathcal{T}_0}\left\|\widehat{\mathcal{I}}_n(t) - \mathcal{I}_0(t)\right\| \to 0. \tag{S.37}$$

This follows using the fact that the class of functions $(y,x) \longmapsto \psi_{\theta(t),t}(y,x)\,\psi'_{\theta(t),t}(y,x)$ has bounded bracketing numbers, which is proved mimicking the arguments in Theorem 4.4 for the class of functions $\mathcal{S}$. Therefore, this class of functions is Glivenko-Cantelli. Then, (S.37) follows applying Theorem 2.4.1 in van der Vaart and Wellner (1996), as in the proof of (S.2) in the proof of Theorem 4.1 above.

With (S.37) on hands, we must show that $\left\{\sqrt{n}\mathbb{P}_n^*\zeta_{\psi_{\hat{\theta}_n}(t)t}\right\}_{t\in\mathcal{T}_0}$ converges in distribution (under the bootstrap law) to $\left\{\mathbb{B}_P\zeta_{\psi_{\theta_0}(t)t}\right\}_{t\in\mathcal{T}_0}$ in the space $\ell^\infty\left(\mathcal{T}_0\right)$ for almost all sample $\{Y_i, X_i, \delta_i\}_{i=1}^n$ where $\mathbb{B}_P$ is the P-Brownian Motion, i.e., a centered Gaussian process indexed by squared integrable (with respect to P) functions and with covariance function $\mathbb{E}\mathbb{B}_{\varphi_1}\mathbb{B}_{\varphi_2} = P\varphi_1\varphi_2$, for $\varphi_1, \varphi_2 \in \ell^\infty\left(\mathcal{T}_0\right)$.

From Lemma S1 and the proof of Theorems 4.2 and 4.4, we have established that the class $\mathcal{B}_0$ of functions $(y,x,d) \mapsto \zeta_{\psi_{\theta_0}(t)t}\left(y,x,d\right)$ is Donsker; see also Stute et al. (2000) for related results. Thus, from van der Vaart and Wellner (1996)'s Theorem 3.6.13, we have that

$$\sup_{h\in BL_1}\left\|\mathbb{E}\left[h\left(\sqrt{n}\mathbb{P}_n^*\zeta_{\psi_{\hat{\theta}_n(t)t}}\right)\right] - \mathbb{E}\left[h\left(\mathbb{B}_P\zeta_{\psi_{\theta_0(t)t}}\right)\right]\right\| \to 0 \text{ with probability 1,}$$

where BL_1 is the class of functions $h : \ell^\infty\left(\mathcal{B}_0\right) \mapsto \mathbb{R}^{k+1}$ such that $|h(z_1) - h(z_2)| \leq ||z_1 - z_2||_{\mathcal{B}_0}$ for every z_1, z_2.The above display is necessary and sufficient for convergence in distribution under the bootstrap law. □

PROOF. (PROOF OF THEOREM 5.1) Note that, by adding and subtracting different terms, we have that

$$\begin{aligned}
&\sqrt{n}\left(\hat{\eta}_n - \eta\right)(t) \\
&= \frac{1}{n}\sum_{i=1}^n\left[\lambda(\mathbf{X}_i'\hat{\theta}_n(t))\right]\sqrt{n}\left(\hat{\beta}_n\left(t\right) - \beta_0\left(t\right)\right) \\
&+\beta_0\left(t\right)\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\lambda(\mathbf{X}_i'\hat{\theta}_n(t)) - \lambda\left(\mathbf{X}_i'\theta_0(t)\right)\right] \qquad \text{(S.38)}\\
&+\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^n\left[\beta_0\left(t\right)\lambda(\mathbf{X}_i'\theta_0(t))\right] - \mathbb{E}\left[\beta_0\left(t\right)\lambda(\mathbf{X}'\theta_0(t))\right]\right).
\end{aligned}$$

We analyze each of these terms separately.

From Theorem 4.4 and a standard Glivenko-Cantelli argument, we have that, uniformly in $t \in \mathcal{T}_0$,

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^n\left[\lambda(\mathbf{X}_i'\hat{\theta}_n(t))\right]\sqrt{n}\left(\hat{\beta}_n\left(t\right) - \beta_0\left(t\right)\right) \\
&= \mathbb{E}\left[\lambda(\mathbf{X}'\theta_0(t))\right]\cdot H\cdot\mathcal{I}_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^n\zeta_i(t) + o_p\left(1\right). \qquad \text{(S.39)}
\end{aligned}$$

where $\zeta_i(t)$ is as defined in (4.2), with $\varphi = \psi_{\theta t}$, $\mathcal{I}_0$ as defined in (4.1) and $H = [0_{k\times1}, I_k]$ is the $k \times (k+1)$ selection matrix for the slope coefficients.

Next, from the mean value theorem, we have that

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^n\left[\lambda(\mathbf{X}_i'\hat{\theta}_n(t)) - \lambda\left(\mathbf{X}_i'\theta_0(t)\right)\right] \\
&= \frac{1}{n}\sum_{i=1}^n\left[\dot{\lambda}(\mathbf{X}_i'\bar{\theta}(t))\mathbf{X}_i'\right]\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)(t), \qquad \text{(S.40)}
\end{aligned}$$

with $||\bar{\theta} - \theta_0|| \leq ||\hat{\theta}_n - \theta_0||$. Given that $\dot{\lambda}$ is a continuous function by Assumption 4.2, $\hat{\theta}_n(\cdot)$ is a uniformly consistent estimator for $\theta_0(\cdot)$, and the function $\dot{\lambda}$ is bounded on a neighborhood of θ_0, from a Glivenko-Cantelli argument, we have that, uniformly in $t \in \mathcal{T}_0$,

$$\frac{1}{n}\sum_{i=1}^{n}\left[\dot{\lambda}(\mathbf{X}_i'\hat{\theta}_n(t))\mathbf{X}_i'\right] = \mathbb{E}\left[\dot{\lambda}(\mathbf{X}'\theta_0(t))\mathbf{X}'\right] + o_p(1). \tag{S.41}$$

Therefore, combining (S.40) and (S.41) with the uniform linear representation of $\sqrt{n}(\hat{\theta} - \theta_0)(t)$ derived in the proof of Theorem 4.4 (see, also, Theorem 4.2), we have

$$\begin{aligned} &\beta_0(t)\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[\lambda(\mathbf{X}_i'\hat{\theta}_n(t)) - \lambda(\mathbf{X}_i'\theta_0(t))\right] \\ &= \beta_0(t)\,\mathbb{E}\left[\dot{\lambda}(\mathbf{X}'\theta_0(t))\mathbf{X}'\right]\,\mathcal{I}_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\zeta_i(t) + o_p(1) \end{aligned} \tag{S.42}$$

uniformly in $t \in \mathcal{T}_0$.

The third term of (S.38) can be rewritten as

$$\begin{aligned} &\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}[\beta_0(t)\,\lambda(\mathbf{X}_i'\theta_0(t))] - \mathbb{E}[\beta_0(t)\,\lambda(\mathbf{X}'\theta_0(t))]\right) \\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\beta_0(t)\left(\lambda(\mathbf{X}_i'\theta_0(t)) - \mathbb{E}[\lambda(\mathbf{X}'\theta_0(t))]\right). \end{aligned} \tag{S.43}$$

Thus, from (S.38), (S.39), (S.42) and (S.43), we have that, uniformly in $t \in \mathcal{T}_0$,

$$\sqrt{n}(\hat{\eta}_n - \eta)(t) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\zeta_{i,ADME}(t) + o_P(1), \tag{S.44}$$

where

$$\begin{aligned} \zeta_{i,ADME}(t) = {} & \beta_0(t)\cdot\left(\lambda(\mathbf{X}_i'\theta_0(t)) - \mathbb{E}[\lambda(\mathbf{X}'\theta_0(t))]\right) \\ & + \mathbb{E}[\lambda(\mathbf{X}'\theta_0(t))]\cdot H\cdot\mathcal{I}_0^{-1}\cdot\zeta_i(t) \\ & + \beta_0(t)\cdot\mathbb{E}\left[\dot{\lambda}(\mathbf{X}'\theta_0(t))\,\mathbf{X}'\right]\cdot\mathcal{I}_0^{-1}\cdot\zeta_i(t). \end{aligned}$$

From Example 19.7 of van der Vaart (1998), Theorem 4.4, and Lemma 9.25 of Kosorok (2008), see Lemma S1's proof (page S17), it then follows that $(y, x, d) \mapsto \zeta_{ADME;t}(y, x, d)$ forms a Donsker class of functions, completing the weak convergence proof. The validity of the multiplier bootstrap now follows from van der Vaart and Wellner (1996)'s Theorem 3.6.13, completing the proof. □

REFERENCES

Arcones, M. A. and E. Giné (1993, jul). Limit Theorems for \$U\$-Processes. *The Annals of Probability 21*(3), 347–370.

Arcones, M. A. and E. Giné (1995). On the law of the iterated logarithm for canonical U-statistics and processes. *Journal of Theoretical Probability 58*, 217–245.

Bennett, S. (1983). Analysis of survival data by the proportional odds model. *Statistics in Medicine 2*, 273–277.

Breslow, N. (1974). Covariance analysis of censored survival data. *Biometrics 30*, 89–99.

Chernozhukov, V., I. Fernández-Val, and B. Melly (2013). Inference on counterfactual distributions. *Econometrica 81*, 2205–2268.

Clayton, A. D. G. (1976). An Odds Ratio Comparison for Ordered Categorical Data with Censored Observations. *Biometrika 63*, 405–408.

Cox, D. (1975). Partial Likelihood. *Biometrika 62*, 269–276.

Cox, D. R. (1972). Regression models and life-tables (with discussion). *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 34*, 187–220.

de la Peña, V. H. and E. Giné (1999). *Decoupling : from dependence to independence.* New York, NY: Springer.

Finkelstein, H. (1971). The Law of the Iterated Logarithm for Empirical Distribution. *The Annals of Mathematical Statistics 42*(2), 607–615.

Giné, E. and V. Koltchinskii (2006). Concentration inequalities and asymptotic results for ratio type empirical processes. *Annals of Probability 34*(3), 1143–1216.

Hunter, D. R. and K. Lange (2002). Computing estimates in the proportional odds model. *Annals of the Institute of Statistical Mathematics 54*, 155–168.

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

Shorack, G. R. and J. A. Wellner (1986). *Empirical processes with applications to statistics.* New York: Wiley.

Stute, W. (1993). Consistent estimation under random censorship when covariables are present. *Journal of Multivariate Analysis 45*, 89 – 103.

Stute, W. (1994). Strong and weak representations of cumulative hazard function and Kaplan-Meier estimators on increasing sets. *Journal of Statistical Planning and Inference 42*(3), 315–329.

Stute, W. (1995). The central limit theorem under random censorship. *Annals of Statistics 23*, 422–439.

Stute, W. (1996). Distributional convergence under random censorship when covariables are present. *Scandinavian Journal of Statistics 23*, 461–471.

Stute, W., W. González-Manteiga, and M. P. Quindimil (1998). Bootstrap Approximations in Model Checks for Regression. *Journal of the American Statistical Association 93*(441), 141–149.

Stute, W., W. González-Manteiga, and C. . Sánchez Sellero (2000). Nonparametric model checks in censored regression. *Communications in Statistics - Theory and Methods 29*, 1611 – 1629.

van der Vaart, A. W. (1998). *Asymptotic Statistics.* Cambridge: Cambridge University Press.

van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes.* New York: Springer.