A general approach to goodness of fit for U-processes

University of Rochester and University of Colorado

Supplementary Material

Before proving main results, we make the following assumptions.

1. The parameter space $\mathbf{\Theta}$ is compact and the true parameter $\boldsymbol{\eta}_0$ is the interior point of $\mathbf{\Theta}$.

2. Let $||\cdot||$ be Euclidean norm. The functions $\mathbf{h}(\cdot,\cdot,\boldsymbol{\eta})$ and

$$\mathbf{u}(\cdot,\cdot,\boldsymbol{\eta},w)=\sup_{||\boldsymbol{\tau}-\boldsymbol{\eta}||\leq w}||\mathbf{h}(\cdot,\cdot,\boldsymbol{\tau})-\mathbf{h}(\cdot,\cdot,\boldsymbol{\eta})||$$

   are measurable functions of $\mathbf{V}_{i_1}$ and $\mathbf{V}_{i_2}$ for $1\leq i_1\neq i_2\leq n$ in some open neighborhood of $\mathbf{\Theta}$.

3. Let $\lambda(\boldsymbol{\eta})=E\{n^{-1/2}\mathbf{U}_n^{FR}(\boldsymbol{\eta})\}$. Then $\lambda(\boldsymbol{\eta}_0)=0$ and $\lambda(\boldsymbol{\eta})$ is differentiable at $\boldsymbol{\eta}_0$ with nonsingular derivative at $\boldsymbol{\eta}_0$.

4. For $1\leq i_1\neq i_2\leq n$, there exist positive constant a_0,b_0 and c_0 such that $E\{\mathbf{u}(\mathbf{V}_{i_1},\mathbf{V}_{i_2},\boldsymbol{\eta},r)\}\leq a_0r$ and $E\{\mathbf{u}(\mathbf{V}_{i_1},\mathbf{V}_{i_2},\boldsymbol{\eta},r)^2\}\leq b_0r$ for all $r\leq c_0$ and all $\boldsymbol{\eta}$ in an open neighborhood of $\boldsymbol{\eta}_0$.

5. There exists $K>0$ such that $E(||\mathbf{Z}||^2)\leq K$.

6. The error distribution has finite Fisher information.

7. The distribution of $\mathbf{Z}$ given $\Delta = 1$ is not concentrated on a proper hyperplane on $\mathbb{R}^p$.

8. The information bound (Bickel et al. (1993, Chapter 2, p23)) for estimating $\boldsymbol{\eta}_0$ is finite and invertible.

9. We assume that error terms ϵ_i of model (1.1) and (2.6) in main paper are i.i.d, so that $\mathbf{V}_i = (\epsilon_i, C_i, \mathbf{Z}_i)$ are i.i.d.

The first assumption represents a standard regularity assumption. Assumptions 2-4 are necessary for proving results regarding the tightness results regarding stochastic processes. These conditions also allow for an expansion analogous to a Taylor series expansion for nonsmooth estimating functions. The fifth assumption is a standard second moment condition for covariates needed to guarantee finite information for regression parameters. Assumptions 6 and 8 are conditions needed to guarantee the existence of the information bound for the estimator of $\boldsymbol{\eta}_0$. Assumption 7 guarantees that distribution of covariates is nondegenerate with respect to the censoring mechanism. Assumption 9 is one from model assumptions (1.1) and (2.6). Moreover, we assume regularity conditions C1-C4 in the Appendix of Peng and Fine (2006).

In this Appendix, we will prove the four theorems in the main paper. Before proving them, we introduce several definitions, including Euclidean class of functions, which are crucial to proving tightness. We will also use linear functional notations in the proofs.

Definition 1 (Nolan and Pollard (1987)). Let S be the set equipped with a pseudometric d. The covering number $N(\tau, d, S)$ is defined as the smallest value of N for which there exist N closed balls of diameter τ, and centers in S, whose union covers S.

Definition 2. Let $\mathcal{F}$ denote a class of functions defined from $\mathcal{X}$ to $\mathbb{R}$. The envelope function for $\mathcal{F}$, F, is given by

$$F(x) = \sup_{f \in \mathcal{F}} |f(x)|.$$

Definition 3. Let $\mathcal{F}$ be a class of functions and F be its envelope. Let Q be a measure on a space $\mathcal{X}$. For $q > 0$, we define $Q(F^q)$ as

$$Q(F^q) = \int F^q dQ.$$

Definition 4 (Nolan and Pollard (1987)). Let $\mathcal{F}$ be class of functions and F be envelope of $\mathcal{F}$. Define Q to be a measure on the space $\mathcal{X}$. If $0 < Q(F) < \infty$ and there exist constants A and B such that

$$N_1(\tau, Q, \mathcal{F}, F) \leq A\tau^{-B}, \text{ for } 0 < \tau \leq 1,$$

then $\mathcal{F}$ is called Euclidean class, and A and B are called Euclidean constants for F. If $\mathcal{F}$ is Euclidean, for each $p > 1$, if $0 < Q(F^p) < \infty$

$$N_p(\tau, Q, \mathcal{F}, F) \leq A2^{pV}\tau^{-pV}, \text{ for } 0 < \tau \leq 1,$$

Now we define a metric in the space and we focus on symmetric functions. Let $\mathcal{F}$ be class of symmetric functions and F be its envelope. Let the metric $d_{Q,p,F}$ which is defined on $\mathcal{F}$ be

$$d_{Q,p,F}(f,g) = \left[\frac{Q|f-g|^p}{Q(G^p)}\right]^{1/p} \quad f, g \in \mathcal{F}$$

where Q is a measure on $\mathcal{X} \otimes \mathcal{X}$ that satisfies $0 < Q(F^p) < \infty$. We define $N_p(\tau, Q, \mathcal{F}, F)$ to be covering number $N(\tau, d_{Q,p,F}, \mathcal{F})$. Let $y_1, \ldots y_{2n}$ be a sample from Q. Define T_n to be the measure which assigns mass one at each of the $4n(n-1)$ pairs of y_v, y_w in function g_{ij} for $u \in \mathcal{F}$, where

$$g_{ij} = u(y_{2i}, y_{2j}) - u(y_{2i}, y_{2j-1}) - u(y_{2i-1}, y_{2j}) + u(y_{2i-1}, y_{2j-1}),$$

This measure plays an important role in the construction of exponential inequalities and the proofs of convergence theorems in U-processes (Nolan and Pollard (1987); Nolan and Pollard (1988)). Finally, we define

$$J_n(s, Q, \mathcal{F}, F) = \int_0^s \log N_2(x, Q, \mathcal{F}, F)dx,$$

where $N_2(x, Q, \mathcal{F}, F)$ is the covering number $N(x, d_{Q,2,\mathcal{F}}, \mathcal{F})$ (Nolan and

Pollard (1987)). Then

$$\sup_n E\{J_n(s,Q,\mathcal{F},F)\} = \sup_n E\bigg\{\int_0^s \log N_2(x,Q,\mathcal{F},F)dx\bigg\}.$$

S1 Proof of tightness of $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$ and $\mathbf{U}_n^{P}(t;\hat{\boldsymbol{\gamma}})$

The first step of our proof is to show tightness. Let N_0 be an open neighborhood of $\boldsymbol{\eta}_0$. We will obtain an expansion similar to Taylor series expansion to prove tightness. However, since $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$ is nonsmooth, Taylor series expansion is not directly applicable. Assumptions 2-4 allow for the application of Lemma 2 from Honoré and Powell (1994), which yields

$$\sup_{\boldsymbol{\eta}\in N_0} \frac{||\mathbf{U}_n^{FR}(\boldsymbol{\eta}) - \mathbf{U}_n^{FR}(\boldsymbol{\eta}_0) - n^{1/2}\lambda(\boldsymbol{\eta})||}{1+n^{1/2}||\lambda(\boldsymbol{\eta})||} = o_p(1). \qquad \text{(S1.1)}$$

Then by a Taylor series expansion for $\lambda(\boldsymbol{\eta})$ and consistency of $\hat{\boldsymbol{\eta}}$ (Peng and Fine (2006)),

$$\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}}) = \mathbf{U}_n^{FR}(t;\boldsymbol{\eta}_0) + n^{1/2}\boldsymbol{\Gamma}_0(t)(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0) + o_p(1),$$

where $\boldsymbol{\Gamma}_0(t)$ is the expectation of slope matrix of $\mathbf{U}_n^{FR}(t;\boldsymbol{\eta}_0)$. Clearly, $n^{1/2}(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}_0)$ converges in distribution, so it is tight.

The next step is to show tightness of $\mathbf{U}_n^{FR}(t;\boldsymbol{\eta}_0)$. Note that $g(\mathbf{V}_i,\mathbf{V}_j,\boldsymbol{\eta}_0) = e_i(\boldsymbol{\eta}_0)\vee e_j(\boldsymbol{\eta}_0)$. For each t, a class of functions $g_t\{e_i(\boldsymbol{\eta}_0), e_j(\boldsymbol{\eta}_0)\} = e_i(\boldsymbol{\eta}_0)\vee e_j(\boldsymbol{\eta}_0) - t$ is a polynomial class (Note that for each t, $e_i(\boldsymbol{\eta}_0)\vee e_j(\boldsymbol{\eta}_0) - t$ is an element of a finite dimensional vector space of real functions).

Then by the arguments in Nolan and Pollard (1987), the class of functions $g_t\{e_i(\boldsymbol{\eta}_0), e_j(\boldsymbol{\eta}_0)\}$ is Euclidean with envelope 2. Let $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}_0, t) = \frac{1}{2}(\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}_0) > e_i(\boldsymbol{\eta}_0)\} - \Delta_j I\{e_i(\boldsymbol{\eta}_0) > e_j(\boldsymbol{\eta}_0)\}] I\{e_i(\boldsymbol{\eta}_0) \vee e_j(\boldsymbol{\eta}_0) \leq t\}$. Since $[\Delta_i I\{e_j(\boldsymbol{\eta}_0) > e_i(\boldsymbol{\eta}_0)\} - \Delta_j I\{e_i(\boldsymbol{\eta}_0) > e_j(\boldsymbol{\eta}_0)\}]$ are bounded and $(\mathbf{Z}_i - \mathbf{Z}_j)$ is a difference of two random variables, applying Lemma 22 in Nolan and Pollard (1987) with arguments about Euclidean class in Pollard (1986) implies that $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}_0, t)$ is also Euclidean with some envelope $G = G(\cdot, \cdot)$.

Let $\mathcal{G}$ be function space for $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}_0, t)$. Note that $\mathcal{G}$ is a class of functions in $\mathcal{L}^2$. The metric for $\mathcal{G}$ given measure Q is

$$d_{Q,2,G}(f^*, g^*) = \left[\frac{Q|f^* - g^*|^2}{Q(G^2)}\right]^{1/2} \quad f^*, g^* \in \mathcal{G}.$$

Let P be the distribution of $\mathbf{V}_1, \ldots, \mathbf{V}_n$. $P\mathcal{G}$ be a class of functions of $P\{g(x, \cdot)\}$, where $g \in \mathcal{G}$. Clearly, $P\mathcal{G}$ is the class of functions of $E\{h(\mathbf{v}, \mathbf{V}, \boldsymbol{\eta}_0, t)\}$. Since $E(\mathbf{Z}^2)$ is bounded, $E\{\mathbf{h}(\mathbf{v}, \mathbf{V}, \boldsymbol{\eta}_0, t)\}$ and $E\{\mathbf{h}^2(\mathbf{v}, \mathbf{V}, \boldsymbol{\eta}_0, t)\}$ are also bounded for all $\mathbf{v}$. By Corollary 21 in Nolan and Pollard (1987), $P\mathcal{G}$ is also Euclidean with envelope $PG = P[G(x, \cdot)]$. Since $T_n(G^2) = \int G^2 dT_n$ takes value between 0 and ∞, by argument in Nolan and Pollard (1987) about Euclidean class, there exists positive constant A_1 and B_1 such that

$N_2(x, T_n, \mathcal{G}, G) \leq A_1 4^{B_1} x^{-2B_1}$ for $0 < x \leq 1$. Then

$$\int_0^1 \log N_2(x, T_n, \mathcal{G}, G)dx \leq \int_0^1 (\log A_1 + B_1 \log 4 - 2B_1 \log x)dx$$

$$= \log A_1 + B_1 \log 4 - 2B_1(x \log x - x)|_0^1 = \log A_1 + B_1 \log 4 + 2B_1 < \infty. \tag{S1.2}$$

Let μ be a measure defined on $\mathcal{X} \otimes \mathcal{X}$. Define

$$J(t, \mu, \mathcal{G}, G) = \int_0^t \log N_2(x, \mu, \mathcal{G}, G)dx$$

Then by (S1.2),

$$\sup_n E\{J(1, T_n, \mathcal{G}, G)\}^2 < \infty. \tag{S1.3}$$

Let P_n be the empirical measure on sample $\mathbf{V}_1, \ldots, \mathbf{V}_n$. Since $E\{\mathbf{h}^2(\mathbf{v}, \mathbf{V}, \boldsymbol{\eta}_0, t)\}$ is bounded, $0 < P_n(P(G^2)) = \int(\int G^2 dP)dP_n < \infty$. Since $P\mathcal{G}$ is also Euclidean, by using similar arguments as for (S1.2),

$$\sup_n E\{J(1, P_n, P\mathcal{G}, PG)\}^2 < \infty. \tag{S1.4}$$

Since $(P \otimes P)(G^2) = \int\int G^2 d(P \otimes P)$ is also positive and finite, we have,

$$J(1, P \otimes P, \mathcal{G}, G) < \infty. \tag{S1.5}$$

Thus it is enough to show that for every $\zeta > 0$ and $\delta > 0$, we can find $\nu > 0$ such that

$$\limsup_{n\to\infty} E\{J(\nu, P_n, P\mathcal{G}, PG) > \zeta\} < \delta. \tag{S1.6}$$

Since $P\mathcal{G}$ is also Euclidean and $0 < P_n(P(G^2)) < \infty$ is satisfied, by similar calculation of (S1.2), there exists a constant w such that

$$J(t, P_n, P\mathcal{G}, PG) = \int_0^t \log N_2(x, P_n, P\mathcal{G}, PG)dx = w.$$

For $\zeta > 0$, by taking t to be the solution of

$$\int_0^t \log N_2(x, P_n, P\mathcal{G}, PG)dx = \zeta.$$

Thus (S1.6) holds. Hence by Theorem 5 of Nolan and Pollard (1988), $\mathbf{U}_n^{FR}(t; \boldsymbol{\eta}_0)$ is tight. Hence $\mathbf{U}_n^{FR}(t; \hat{\boldsymbol{\eta}})$ is also tight. For the dependent censoring case, combining the above argument with those in Lin, Wei, and Ying (1993), the process $\mathbf{W}_n(t, s; \hat{\boldsymbol{\gamma}})$ is tight.

S2 Proof of Theorem 1 and Theorem 2

S2.1 Proof of Theorem 1

Let $\mathbf{h}(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\eta}_0) = \dfrac{1}{2}(\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j(\boldsymbol{\eta}_0) > e_i(\boldsymbol{\eta}_0)\} - \Delta_j I\{e_i(\boldsymbol{\eta}_0) > e_j(\boldsymbol{\eta}_0)\}]$. Define

$$2\mathbf{h}_1(\mathbf{v}, \boldsymbol{\eta}_0, t) = 2E\{\mathbf{h}(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\eta}_0, t)\}$$

where $\mathbf{h}(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\eta}_0, t) = \mathbf{h}(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\eta}_0)I\{g(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\eta}_0) \leq t\}$ and $2\mathbf{h}_1(\mathbf{v}, \boldsymbol{\eta}_0) = 2E\{\mathbf{h}(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\eta}_0)\}$. Let

$$\mathbf{H}_i(t) = 2\mathbf{h}_1(\mathbf{V}_i, \boldsymbol{\eta}_0, t) \quad \mathbf{H}_i = 2\mathbf{h}_1(\mathbf{V}_i, \boldsymbol{\eta}_0).$$

By the arguments in the appendices of Lin, Robins, and Wei (1996) and Peng and Fine (2006),

$$\begin{aligned}\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}}) &= \mathbf{U}_n^{FR}(t;\boldsymbol{\eta}_0) - \boldsymbol{\Gamma}_0(t)\boldsymbol{\Gamma}_0^{-1}\mathbf{U}_n^{FR}(\boldsymbol{\eta}_0) + o_p(1) \\ &= n^{-1/2}\sum_{i=1}^{n}\mathbf{H}_i(t) - \boldsymbol{\Gamma}_0(t)\boldsymbol{\Gamma}_0^{-1}n^{-1/2}\sum_{i=1}^{n}\mathbf{H}_i + o_p(1), \end{aligned} \tag{S2.1}$$

When (1.1) is true, by the multivariate central limit theorem, $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$ converges in finite dimensional distribution to a mean-zero Gaussian process with covariance matrix

$$E[\{\mathbf{H}_1(t) - \boldsymbol{\Gamma}_0(t)\boldsymbol{\Gamma}_0^{-1}\mathbf{H}_1\}\{\mathbf{H}_1(t) - \boldsymbol{\Gamma}_0(t)\boldsymbol{\Gamma}_0^{-1}\mathbf{H}_1\}^T]. \tag{S2.2}$$

By uniform strong law of large numbers by Pollard (1990, Section 8, p. 41) and strong consistency of $\hat{\boldsymbol{\eta}}$, covariance matrix of $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$ converged to (S2.2) uniformly t almost surely. By the tightness of $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$ and the arguments in the Appendix of Lin et al. (2000), $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$ converges weakly to a Gaussian process.

For the dependent censoring case, we assume regularity conditions C1-C4 in the Appendix of Peng and Fine (2006). Let

$$\bar{Z}^{(1)}(u;\boldsymbol{\eta}) = \frac{\sum_{j=1}^{n} I\{\tilde{D}_i^*(\eta) \geq u\}\mathbf{Z}_j}{\sum_{j=1}^{n} I\{\tilde{D}_i^*(\eta) \geq u\}}$$

$$M_{1i}(t;\boldsymbol{\eta}_0) = N_{1i}(t;\boldsymbol{\eta}_0) - \int_{-\infty}^{t} I\{\tilde{D}_i^*(\boldsymbol{\eta}_0) \geq u\}\lambda_{10}(u)du$$

$$\mathbf{h}^*(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\gamma}) = (\mathbf{Z}_i - \mathbf{Z}_j)\psi_{ij}(\boldsymbol{\gamma})$$

$$\mathbf{h}^*(\mathbf{V}_i, \mathbf{V}_j, \boldsymbol{\gamma}, t) = (\mathbf{Z}_i - \mathbf{Z}_j)\psi_{ij}(\boldsymbol{\gamma})I\{\tilde{X}_{i(j)}^*(\boldsymbol{\gamma}) \vee \tilde{X}_{j(i)}^*(\boldsymbol{\gamma}) \leq t\}.$$

where $\bar{z}^{(1)}(u) = \lim_{n\to\infty} \bar{Z}_1(u;\boldsymbol{\eta}_0)$ and λ_{10} is true common hazard function for $\{\tilde{D}_i^*(\boldsymbol{\eta}_0)\}_{i=1}^n$ (Lin, Robins, and Wei (1996)). Let $\mathbf{h}_1^*(\mathbf{v}, \boldsymbol{\gamma}_0, t) = E\{\mathbf{h}^*(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\gamma}_0, t)\}$ and $\mathbf{h}_1^*(\mathbf{v}, \boldsymbol{\gamma}_0) = E\{\mathbf{h}^*(\mathbf{v}, \mathbf{V}_2, \boldsymbol{\gamma}_0)\}$.

$$\mathbf{H}_i^*(t,s) = \begin{pmatrix} \int_{-\infty}^{t} \{Z_i - \bar{z}^{(1)}(u)\}dM_{1i}(u;\boldsymbol{\eta}_0) \\ 2\mathbf{h}_1^*(\mathbf{V}_i, \boldsymbol{\gamma}_0, s) \end{pmatrix} \quad \mathbf{H}_i^* = \begin{pmatrix} \int_{-\infty}^{\infty} \{Z_i - \bar{z}^{(1)}(u)\}dM_{1i}(u;\boldsymbol{\eta}_0) \\ 2\mathbf{h}_1^*(\mathbf{V}_i, \boldsymbol{\gamma}_0) \end{pmatrix}$$

By replacing $\mathbf{H}_i(t)$ and $\mathbf{H}_i$ to $\mathbf{H}_i^*(t,s)$ and $\mathbf{H}_i^*$, respectively with substituting $\boldsymbol{\Gamma}_0(t)$ and $\boldsymbol{\Gamma}_0$ with $\boldsymbol{\Upsilon}_0(t,s)$ and $\boldsymbol{\Upsilon}_0$, respectively, where $\boldsymbol{\Upsilon}_0(t,s)$ is the expectation of the slope matrix of $\mathbf{W}_n(t,s;\boldsymbol{\gamma}_0)$, (S2.1) still holds. Using arguments from $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$ and Appendix A.2 of Lin et al. (2000), we can show that when (2.6) is true, $\mathbf{W}_n(t,s;\hat{\boldsymbol{\gamma}})$ converges in finite dimensional distribution to Gaussian process with zero mean and covariance matrix

$$E[\{\mathbf{H}_1^*(t,s) - \boldsymbol{\Upsilon}_0(t,s)\boldsymbol{\Upsilon}_0^{-1}\mathbf{H}_1^*\}\{\mathbf{H}_1^*(t,s) - \boldsymbol{\Upsilon}_0(t,s)\boldsymbol{\Upsilon}_0^{-1}\mathbf{H}_1^*\}^T]. \qquad \text{(S2.3)}$$

Similar to proof of convergence for covariance matrix of $\mathbf{U}_n(t;\hat{\boldsymbol{\eta}})$, the covariance matrix of $\mathbf{W}_n(t,s;\hat{\boldsymbol{\gamma}})$ converges almost surely uniformly (t,s) to (S2.3).

This and the tightness of $\mathbf{W}_n(t,s;\hat{\boldsymbol{\gamma}})$ shows that the process $\mathbf{W}_n(t,s;\hat{\boldsymbol{\gamma}})$ converges weakly to a Gaussian process. Thus Theorem 1 is proved.

S2.2 Proof of Theorem 2

By Appendix of Lin, Robins, and Wei (1996) ,

$$\mathbf{U}_n^{FR}(\boldsymbol{\eta}^*) = \mathbf{U}_n^{FR}(\hat{\boldsymbol{\eta}}) + n^{1/2}\boldsymbol{\Gamma}_0(\boldsymbol{\eta}^* - \hat{\boldsymbol{\eta}}) + o_p(1) = n^{1/2}\boldsymbol{\Gamma}_0(\boldsymbol{\eta}^* - \hat{\boldsymbol{\eta}}) + o_p(1) \tag{S2.4}$$

Combining (2.5) and (S2.4) provides

$$\begin{aligned}
\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*) &= \frac{n^{1/2}}{n(n-1)}\sum_{i\neq j}\mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})I\{g(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})\leq t\}Q_i \\
&- \boldsymbol{\Gamma}_0(t)n^{1/2}(\boldsymbol{\eta}^* - \hat{\boldsymbol{\eta}}) + o_p(1) \\
&= \frac{n^{1/2}}{n(n-1)}\sum_{i\neq j}\mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})I\{g(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})\leq t\}Q_i \\
&- \boldsymbol{\Gamma}_0(t)\boldsymbol{\Gamma}_0^{-1}\frac{n^{1/2}}{n(n-1)}\sum_{i\neq j}\mathbf{h}(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\eta}})Q_i + o_p(1).
\end{aligned}$$

We need to show that when (2.1) is true, conditional on the observed data $(Y_i,\Delta_i,\mathbf{Z}_i), i=1,\ldots,n$, the limiting distribution of $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*)$ converges weakly to a Gaussian process and that the limiting covariance matrix of $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*)$ is same as that of $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$. From now, our statements for $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*)$ are always ones which condition on the observed data. By arguments in Appendix of Peng and Fine (2006) and the strong consistency of $\hat{\boldsymbol{\eta}}$, we have expansion

$$\hat{\mathbf{U}}_n(t;\boldsymbol{\eta}^*) = n^{-1/2}\sum_{i=1}^{n} 2\mathbf{h}_1(\mathbf{V}_i,\boldsymbol{\eta}_0,t)Q_i - \boldsymbol{\Gamma}_0^{-1}\boldsymbol{\Gamma}(t)n^{-1/2}\sum_{i=1}^{n} 2\mathbf{h}_1(\mathbf{V}_i,\boldsymbol{\eta}_0)Q_i + o_p(1) \tag{S2.5}$$

Given the observed data, the only random term is Q_is, so by the multivariate central limit theorem, the first term of (S2.5) and the second term of (S2.5) converge to Gaussian process in finite dimension. Since $2\mathbf{h}_1(\mathbf{V}_i,\boldsymbol{\eta}_0,t)$ is increasing function of t, $\mathbf{V}_i$ and Q_i are i.i.d, the first term of expansion (S2.5) is manageable and satisfies the conditions for functional central limit theorem (Pollard (1990, Section 7, p38 and Section 9, p53); Lin et al. (2000)). Hence the first term of (S2.5) is tight and the second term is clearly tight. Thus $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*)$ is tight. Using the finite-distributional convergence results, we can show that $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*)$ converges weakly to Gaussian process and its asymptotic covariance function is $E(\mathbf{L}_1\mathbf{L}_1^T)$, where

$$\mathbf{L} = 2\mathbf{h}_1(\mathbf{V}_1,\boldsymbol{\eta}_0,t) - \boldsymbol{\Gamma}_0(t)\boldsymbol{\Gamma}_0^{-1}2\mathbf{h}_1(\mathbf{V}_1,\boldsymbol{\eta}_0). \tag{S2.6}$$

Moreover, conditional on the observed data covariance function of $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*)$ converges almost surely uniformly t to (S2.6) by uniform strong law of large numbers (Pollard (1990, Section 8, p41)). In addition, the components of $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}$ represent manageable processes in the sense of Pollard (1990, Section 7, page 38). As can be seen, the limiting covariance matrix of $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*)$, conditional on the observed data is the same as that of $\mathbf{U}_n^{FR}(t;\hat{\boldsymbol{\eta}})$. For the dependent censoring case, let

$$\hat{\mathbf{H}}_i^{\ *}(t,s) = \begin{pmatrix} \int_{-\infty}^{t} \{Z_i - \bar{Z}^{(1)}(u)\}\hat{M}_{1i}(u;\hat{\boldsymbol{\eta}}) \\ \frac{2}{n-1}\sum_{j=1}^{n} \mathbf{h}^*(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\gamma}},t) \end{pmatrix} \hat{\mathbf{H}}_i^{\ *} = \begin{pmatrix} \int_{-\infty}^{\infty} \{Z_i - \bar{Z}^{(1)}\}(u;\hat{\boldsymbol{\eta}})\hat{M}_{1i}(u;\hat{\boldsymbol{\eta}}) \\ \frac{2}{n-1}\sum_{j=1}^{n} \mathbf{h}^*(\mathbf{V}_i,\mathbf{V}_j,\hat{\boldsymbol{\gamma}}) \end{pmatrix}$$

By strong consistency of $\hat{\boldsymbol{\gamma}}$, $\hat{\mathbf{W}}_n(t,s)$ has expansion

$$\begin{aligned} \hat{\mathbf{W}}_n(t,s) &= n^{-1/2}\sum_{i=1}^{n}\hat{\mathbf{H}}_i^*(t,s)Q_i - \boldsymbol{\Upsilon}_0^{-1}\boldsymbol{\Upsilon}_0(t,s)n^{-1/2}\sum_{i=1}^{n}\hat{\mathbf{H}}_i^*Q_i + o_p(1) \\ &= n^{-1/2}\sum_{i=1}^{n}\mathbf{H}_i^*(t,s)Q_i - \boldsymbol{\Upsilon}_0^{-1}\boldsymbol{\Upsilon}_0(t,s)n^{-1/2}\sum_{i=1}^{n}\mathbf{H}_i^*Q_i + o_p(1) \qquad \text{(S2.7)} \end{aligned}$$

We can argue similarly for $\hat{\mathbf{W}}_n(t,s;\hat{\boldsymbol{\gamma}})$ as we did for $\hat{\mathbf{W}}_n(t,s;\hat{\boldsymbol{\gamma}})$. By the multivariate central limit theorem, $\hat{\mathbf{W}}_n(t,s)$ converges in finite-dimensional distribution to a mean zero Gaussian process with covariance function $E(\mathbf{L}_2\mathbf{L}_2^T)$, where

$$L_2 = \begin{pmatrix} \int_{-\infty}^{t} \{Z_1 - \bar{z}^{(1)}(u)\}dM_{11}(u;\boldsymbol{\eta}_0) \\ 2\mathbf{h}_1^*(\mathbf{V}_1,\boldsymbol{\gamma}_0,s) \end{pmatrix} - \boldsymbol{\Upsilon}_0^{-1}\boldsymbol{\Upsilon}_0(t,s) \begin{pmatrix} \int_{-\infty}^{\infty} \{Z_1 - \bar{z}^{(1)}(u)\}dM_{11}(u;\boldsymbol{\eta}_0) \\ 2\mathbf{h}_1^*(\mathbf{V}_1,\boldsymbol{\gamma}_0), \end{pmatrix} \qquad \text{(S2.8)}$$

which is equivalent to (S2.3). Applying arguments similar to $\hat{\mathbf{U}}_n^{FR}(t;\boldsymbol{\eta}^*)$ for $\hat{\mathbf{U}}_n^{P}(s;\boldsymbol{\gamma}^*)$ with arguments of Lin, Robins, and Wei (1996) and Lin et al. (2000) for $\hat{\mathbf{S}}(t;\boldsymbol{\eta}^*)$ leads that the first term in (S2.7) is manageable and hence they are tight by the functional central limit theorem. Moreover, the second term of (S2.7) converges to normal distribution. Hence $\hat{\mathbf{W}}_n(t,s;\boldsymbol{\gamma}^*)$

is tight conditional on the observed data and when (2.6) is true, it converges weakly to a Gaussian process as $\mathbf{W}_n(t, s; \hat{\boldsymbol{\gamma}}^*)$ converges unconditionally. By using arguments similar to Lin, Wei, and Ying (1993) and convergence results of the covariance function for $\hat{\mathbf{U}}_n^{FR}(t; \boldsymbol{\eta}^*)$, covariance function of $\hat{\mathbf{W}}_n(t, s; \boldsymbol{\gamma}^*)$ conditional on the observed data converges almost surely to (S2.8) uniformly in (t, s). This concludes the proof.

S3 Proof of Theorem 3

In this section, we will prove consistency of the estimator under model misspecification. Let the full covariate vector be $\mathbf{W} = \{\mathbf{Z}^T, (\mathbf{Z}^*)^T\}^T$. Let $\boldsymbol{\eta}_0$ and $\boldsymbol{\theta}_0$ be true parameter values corresponding to $\mathbf{Z}$ and $\mathbf{Z}^*$, respectively. Assume that the true model is

$$T = \mathbf{W}^T \boldsymbol{\tau}_0 + \epsilon$$

where $\boldsymbol{\tau}_0 = (\boldsymbol{\eta}_0^T, \boldsymbol{\theta}_0^T)^T$. Let $e_i^*(\boldsymbol{\eta}) = Y_i - \mathbf{Z}_i^T \boldsymbol{\eta}$. Then the estimating equation is

$$\mathbf{U}_n^{FRmis}(\boldsymbol{\eta}) = \frac{n^{1/2}}{2n(n-1)} \sum_{i \neq j} (\mathbf{Z}_i - \mathbf{Z}_j)[\Delta_i I\{e_j^*(\boldsymbol{\eta}) > e_i^*(\boldsymbol{\eta})\} - \Delta_j I\{e_i^*(\boldsymbol{\eta}) > e_j^*(\boldsymbol{\eta})\}] = 0. \tag{S3.1}$$

By Theorem 2.1(i) in Fygenson and Ritov (1994), the solution of equation (S3.1) exists. Denote this solution by $\hat{\boldsymbol{\eta}}^{mis}$. By the strong law of large numbers,

$$n^{-1/2}\mathbf{U}_n^{FRmis}(\boldsymbol{\eta}) = \lambda^*(\boldsymbol{\eta}) + o(1).$$

Assume that $\lambda^*(\boldsymbol{\eta})$ has a unique solution $\boldsymbol{\eta}^{mis}$. Let us consider the antiderivative of $\mathbf{U}_n^{FRmis}(\boldsymbol{\eta})$, say $\mathbf{G}_n^{FRmis}(\boldsymbol{\eta})$,

$$\mathbf{G}_n^{FRmis}(\boldsymbol{\eta}) = \frac{n^{1/2}}{2n(n-1)}\sum_{i\neq j}\{e_i^*(\boldsymbol{\eta}) - e_j^*(\boldsymbol{\eta})\}[\Delta_i I\{e_j^*(\boldsymbol{\eta}) > e_i^*(\boldsymbol{\eta})\} - \Delta_j I\{e_i^*(\boldsymbol{\eta}) > e_j^*(\boldsymbol{\eta})\}]$$

Since $\mathbf{G}_n^{FRmis}(\boldsymbol{\eta})$ is convex (Jin et al. (2003)), then $\hat{\boldsymbol{\eta}}^{mis}$ is minimizer of $\mathbf{G}_n^{FRmis}(\boldsymbol{\eta})$. The kernel function of $\mathbf{G}_n^{FRmis}(\boldsymbol{\eta})$ is Euclidean and hence we have

$$\sup_{\eta\in\Theta} ||\mathbf{G}_n^{FRmis}(\boldsymbol{\eta}) - E[\mathbf{G}_n^{FRmis}(\boldsymbol{\eta})]|| \to 0$$

almost surely. Moreover, $\mathbf{G}_n^{FRmis}(\boldsymbol{\eta})$ satisfies conditions in Proposition A1 of Jin, Ying, and Wei (2001). Since $\boldsymbol{\eta}^{mis}$ is unique solution of $\mathbf{G}_n^{FRmis}(\boldsymbol{\eta})$, so $\hat{\boldsymbol{\eta}}^{mis}$ converges almost surely to $\boldsymbol{\eta}^{mis}$ (Jin, Ying, and Wei (2001)). In the presence of dependent censoring, one can show the consistency for the least-false parameter for time to dependent censoring using arguments from Struthers and Kalbfleisch (1986) and Lin and Wei (1989). For the event of interest, we can apply the argument in Appendix of Peng and Fine (2006) for misspecified model.

S4 Proof of Theorem 4

In this section, we will prove consistency of the proposed test. Suppose that the alternative hypothesis is that $\boldsymbol{\eta}$ in the AFT model depends on time, i.e.,

$$T = \mathbf{Z}^T \boldsymbol{\eta}(s) + \epsilon. \tag{S4.1}$$

Let $\hat{\boldsymbol{\eta}}^{mt}$ be estimator of $\boldsymbol{\eta}$ assuming AFT model that has time independent parameters while in the true model parameters actually depend on time. Then by applying similar arguments for the misspecified AFT model, $\hat{\boldsymbol{\eta}}^{mt}$ converges almost surely to constant vector, say $\boldsymbol{\eta}^{mt}$. To show consistency of test, it suffices to show that $n^{-1/2}\mathbf{U}_n^{FR}(t; \hat{\boldsymbol{\eta}}^{mt})$ converges to nonzero limit (Lin, Wei, and Ying (1993); Arbogast and Lin (2004)) against the alternative hypothesis. Under the alternative hypothesis, by the strong law of large number of U-statistics, $n^{-1/2}\mathbf{U}_n^{FR}(t; \hat{\boldsymbol{\eta}}^{mt})$ converges almost surely to

$$\begin{aligned}&\tfrac{1}{2}E[(\mathbf{Z}_1 - \mathbf{Z}_2)\times\\ &E[I\{e_1(\boldsymbol{\eta}^{mt}) \vee e_2(\boldsymbol{\eta}^{mt}) \leq t\}(\Delta_1 I\{e_2(\boldsymbol{\eta}^{mt}) > e_1(\boldsymbol{\eta}^{mt})\} - \Delta_2 I\{e_1(\boldsymbol{\eta}^{mt}) > e_2(\boldsymbol{\eta}^{mt})\})|\mathbf{Z}_1, \mathbf{Z}_2].\end{aligned} \tag{S4.2}$$

Then given $e_1(\boldsymbol{\eta}^{mt}) \vee e_2(\boldsymbol{\eta}^{mt}) \leq t$, the inner expectation of (5.12) is $P[\Delta_1 I\{e_2(\boldsymbol{\eta}^{mt}) > e_1(\boldsymbol{\eta}^{mt})\} - \Delta_2 I\{e_1(\boldsymbol{\eta}^{mt}) > e_2(\boldsymbol{\eta}^{mt})\}]$. Then,

$$
\begin{aligned}
& P[\Delta_1 I\{e_2(\boldsymbol{\eta}^{mt}) > e_1(\boldsymbol{\eta}^{mt})\} - \Delta_2 I\{e_1(\boldsymbol{\eta}^{mt}) > e_2(\boldsymbol{\eta}^{mt})\}] \\
& = P\{(T_1 - \mathbf{Z}_1^T\boldsymbol{\eta}^{mt}) \leq (T_2 - \mathbf{Z}_2^T\boldsymbol{\eta}^{mt}) \wedge (C_1 - \mathbf{Z}_1^T\boldsymbol{\eta}^{mt}) \wedge (C_2 - \mathbf{Z}_2^T\boldsymbol{\eta}^{mt})\} \\
& - P\{(T_2 - \mathbf{Z}_2^T\boldsymbol{\eta}^{mt}) \leq (T_1 - \mathbf{Z}_1^T\boldsymbol{\eta}^{mt}) \wedge (C_1 - \mathbf{Z}_1^T\boldsymbol{\eta}^{mt}) \wedge (C_2 - \mathbf{Z}_2^T\boldsymbol{\eta}^{mt})\} \\
& = P[\{\epsilon_1 + \mathbf{Z}_1^T(\boldsymbol{\eta}(s_1) - \boldsymbol{\eta}^{mt})\} \leq [\epsilon_2 + \mathbf{Z}_2^T(\boldsymbol{\eta}(s_2) - \boldsymbol{\eta}^{mt})] \wedge (C_1 - \mathbf{Z}_1^T\boldsymbol{\eta}^{mt}) \wedge (C_2 - \mathbf{Z}_2^T\boldsymbol{\eta}^{mt})] \\
& - P[\{\epsilon_2 + \mathbf{Z}_2^T(\boldsymbol{\eta}(s_2) - \boldsymbol{\eta}^{mt})\} \leq \{\epsilon_1 + \mathbf{Z}_1^T(\boldsymbol{\eta}(s_1) - \boldsymbol{\eta}^{mt})\} \wedge (C_1 - \mathbf{Z}_1^T\boldsymbol{\eta}^{mt}) \wedge (C_2 - \mathbf{Z}_2^T\boldsymbol{\eta}^{mt})].
\end{aligned}
\tag{S4.3}
$$

Since $\boldsymbol{\eta}(\cdot)$ depends on time and there are covariates, $\epsilon_1+\mathbf{Z}_1^T(\boldsymbol{\eta}(s_1)-\boldsymbol{\eta}^{mt})$ and $\epsilon_2+\mathbf{Z}_2^T(\boldsymbol{\eta}(s_2)-\boldsymbol{\eta}^{mt})$ do not have the same distribution, thus the probability in expression (S4.3) is not 0. For the expression in (S4.3) to be 0, the distribution of $\epsilon_1 + \mathbf{Z}_1^T(\boldsymbol{\eta}(s_1) - \boldsymbol{\eta}^{mt})$ and $\epsilon_2 + \mathbf{Z}_2^T(\boldsymbol{\eta}(s_2) - \boldsymbol{\eta}^{mt})$ should be same. Thus for the function in (S4.3) to be 0, $\boldsymbol{\eta}(s_1) = \boldsymbol{\eta}(s_2) = \boldsymbol{\eta}^{mt}$. Under the presence of the dependent censoring, one can apply the argument of Lin, Wei, and Ying (1993), and for the event of interest, we can apply the arguments similar to those in this section.

References

Arbogast, P. G. and Lin, D. Y. (2004). Goodness-of-fit methods for matched case-control studies. *Can. J. Stat.* **32**, 373-386.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993).

Efficient and Adaptive Estimation for Semiparametric Models. Johns Hopkins University Press, Baltimore.

Fygenson, M. and Ritov, Y. (1994). Monotone estimating equations for censored data, *Ann. Stat.* **22**, 732-746.

Honoré, B. E. and Powell, J. L. (1994). Pairwise difference estimators of censored and truncated regression models. *J. Econom.* **64**, 241-278.

Lin, D. Y. and Wei, L. J. (1989) The robust inference for the Cox proportional hazards model. *J. Am. Stat. Assoc.* **84**,1074-1078.

Lin, D. Y., Wei, L. J. and Ying, Z. (1993). Checking the Cox model with cumulative sums of martingale-based residuals, *Biometrika* **80**, 557-572.

Lin, D. Y., Robins, J. M. and Wei, L. J. (1996). Comparing two failure time distributions in the presence of dependent censoring. *Biometrika* **83**, 381-393.

Lin, D. Y., Wei, L. J., Yang, I. and Ying, Z. (2000). Semiparametric regression for the mean and rate functions of recurrent events. *J Roy Stat Soc B*, 62, 711-730.

Jin, Z., Ying, Z. and Wei, L. J. (2001), A simple resampling method by perturbing the minimand, *Biometrika* **88**, 381-390.

Jin, Z., Lin, D. Y., Wei, L. J. and Ying, Z. (2003). Rank-based inference for the accelerated failure time model. *Biometrika*, **90**, 341-353.

Nolan, D. and Pollard, D. (1987). U-processes : rate of convergence, *Ann. Stat.* **15**, 780-799.

Nolan, D. and Pollard, D. (1988). Functional limit theorems for U-processes, *Ann. Probab.* **16**, 1291-1298.

Peng, L. and Fine, J. P. (2006). Rank estimation of accelerated lifetime models with dependent censoring. *J. Am. Stat. Assoc.* **101**, 1085-1093.

Pollard, D. (1986). Rates of uniform almost-sure convergence for empirical processes indexed by unbounded classes of functions. *Unpublished manuscript.*

Pollard, D. (1990). *Empirical processes: theory and applications.* In NSF-CBMS regional conference series in probability and statistics (pp. i-86). Institute of Mathematical Statistics and the American Statistical Association.

Struthers, C. A. and Kalbfleisch, J. D. (1986). Misspecified proportional hazard models. *Biometrika* **73**, 363-369.