Supplementary material: Conditional modeling of longitudinal data with terminal event

Shengchun Kong, Bin Nan, John D. Kalbfleisch, Rajiv Saran, and Richard Hirth

A General Theorems about M-estimators and Technical Lemmas

The following Lemma A.1 is a general M-estimation theory for a parametric model (see van der Geer (2000)) and it is a special case of Wellner and Zhang (2007). Thus its proof is omitted. Lemma A.2, which is used for the proof of Theorem 4.2, is similar to Theorem A.1 of Li and Nan (2011), but the former focuses on infinite-dimensional nuisance parameters, whereas the latter focuses on finite-dimensional nuisance parameters. Note that Lemma A.2 reduces to Lemma A.1 with the nuisance parameters fixed at their true values. We provide Lemma A.1 here for ease of reference in the proof for the complete case analysis, which is given in Theorem 4.1. Let $|\cdot|$ be the Euclidian norm and $\|\eta-\eta_0\| = \sup_{s,x}|\eta(s;x)-\eta_0(s;x)|$. We adopt the empirical process notation of van der Vaart and Wellner (1996).

Lemma A.1. *(Asymptotic normality for M-estimation) Given i.i.d. observations X_i, $i = 1, \cdots, n$, suppose that the estimates $\tilde{\psi}_n$ of unknown parameters ψ maximize the objective function $\mathbb{P}_n m(\psi; X)$. Let $\dot{m}(\psi; X) = \partial m(\psi; X)/\partial\psi$ and $\ddot{m}(\psi; X) = \partial^2 m(\psi; X)/(\partial\psi\partial\psi')$. Consider the following conditions:*

A1. $|\tilde{\psi}_n - \psi_0| = o_p(1)$.

A2. $A = -P\{\ddot{m}(\psi; X)\}$ *is non-singular.*

A3. $P\dot{m}(\psi_0; X) = 0$.

A4. The estimates $\tilde{\psi}_n$ satisfy $\mathbb{P}_n\dot{m}(\psi; X) = o_p(n^{-1/2})$.

A5. For any $\delta_n > 0$, let $\Psi_n = \{\psi : |\psi - \psi_0| \le \delta_n\}$, we have $\sup_{\psi\in\Psi_n}|\mathbb{G}_n\{\dot{m}(\psi; X) - \dot{m}(\psi_0; X)\}| = o_p(1)$.

A6. For $\psi \in \Psi_n$, $|P\{\dot{m}(\psi; X) - \dot{m}(\psi_0; X) - \ddot{m}(\psi_0; X)(\psi - \psi_0)\}| = o(|\psi - \psi_0|)$.

If Conditions A1-A6 hold, then we have

$$\sqrt{n}(\tilde{\psi}_n - \psi_0) = A^{-1}\sqrt{n}\mathbb{P}_n\dot{m}(\psi_0; X) + o_{p^*}(1).$$

Lemma A.2. *(Asymptotic normality for pseudo M-estimation) Given i.i.d. observations X_i, $i = 1, \cdots, n$, suppose that the estimates $\hat{\theta}_n$ of θ are chosen to maximize the objective* *function* $\mathbb{P}_n m(\theta, \tilde{\phi}_n, \tilde{\eta}_n; X)$, *where* $\tilde{\phi}_n$ *is an estimator of the true parameter* $\phi_0 \in \Phi \subset R^d$, *and* $\tilde{\eta}_n$ *is an estimator of the true parameter* $\eta_0 \in \mathcal{H}$, *which is an infinite dimensional Banach space. Suppose that* η_t *is a parametric submodel in* $\mathcal{F}$ *passing through* η, *that is,* $\eta_t \in \mathcal{F}$ *and* $\eta_{t=0} = \eta$. *Let* $H = \left\{h : h = \partial\eta_t/\partial t\big|_{t=0}\right\}$ *be the collection of all directions to approach* η. *Let* $\dot{m}_1(\theta, \phi, \eta; X) = \partial m(\theta, \phi, \eta; X)/\partial\theta$, $\dot{m}_2(\theta, \phi, \eta; X) = \partial m(\theta, \phi, \eta; X)/\partial\phi$, *and* $\dot{m}_3(\theta, \phi, \eta; X)[h] = \partial m(\theta, \phi, \eta; X)/\partial t$ *along the direction of* h. *Let* $\ddot{m}_{ij}$ *be the second order derivatives of* m *with respect to corresponding arguments defined in a similar way,* $i, j \in \{1, 2, 3\}$. *Consider the following conditions:*

B1. $|\tilde{\phi}_n - \phi_0| = o_p(1)$, $|\hat{\theta}_n - \theta_0| = o_p(1)$ *and* $\|\tilde{\eta}_n - \eta_0\| = O_p(n^{-\nu})$ *for some* $\nu > 0$ *and some norm* $\|\cdot\|$.

B2. $A = -P\{\ddot{m}_{11}(\theta_0, \phi_0, \eta_0; X)\}$ *is non-singular.*

B3. $P\dot{m}_1(\theta_0, \phi_0, \eta_0; X) = 0$.

B4. The estimator $\hat{\theta}_n$ *satisfy* $\mathbb{P}_n\dot{m}_1(\hat{\theta}_n, \tilde{\phi}_n, \tilde{\eta}_n; X) = o_p(n^{-1/2})$.

B5. For any $\delta_n \downarrow 0$ *and constant* $C > 0$, *let* $\Theta_n = \{(\theta, \phi, \eta) : |(\theta, \phi) - (\theta_0, \phi_0)| \leq \delta_n, \|\tilde{\eta}_n - \eta_0\|_2 \leq Cn^{-\nu}\}$, *we have* $\sup_{(\theta,\phi,\eta)\in\Theta_n} |\mathbb{G}_n\{\dot{m}_1(\theta, \phi, \eta; X) - \dot{m}_1(\theta_0, \phi_0, \eta_0; X)\}| = o_p(1)$.

B6. For some $\varsigma > 1$ *satisfying* $\varsigma\nu > 1/2$, *and for* $(\theta, \phi, \eta) \in \Theta_n$,

$$\begin{aligned}
&|P\{\dot{m}_1(\theta, \phi, \eta; X) - \dot{m}_1(\theta_0, \phi_0, \eta_0; X) \\
&\qquad -\ddot{m}_{11}(\theta_0, \phi_0, \eta_0; X)(\theta - \theta_0)\}| - \ddot{m}_{12}(\theta_0, \phi_0, \eta_0; X)(\phi - \phi_0) \\
&\qquad -\ddot{m}_{13}(\theta_0, \phi_0, \eta_0; X)[\eta - \eta_0]\}| \\
&\quad = o(|\theta - \theta_0|) + o(|\phi - \phi_0|) + O(\|\eta - \eta_0\|^\varsigma).
\end{aligned}$$

Suppose that Conditions B1-B6 hold, then we have

$$\begin{aligned}
&\sqrt{n}(\hat{\theta}_n - \theta_0) \\
&\quad = A^{-1}\sqrt{n}\mathbb{P}_n\dot{m}_1(\theta_0, \phi_0, \eta_0; X) + A^{-1}\sqrt{n}P\{\ddot{m}_{12}(\theta_0, \phi_0, \eta_0; X)\}(\tilde{\phi}_n - \phi_0) \\
&\qquad +A^{-1}\sqrt{n}P\{\ddot{m}_{13}(\theta_0, \phi_0, \eta_0; X)[\tilde{\eta}_n - \eta_0]\} + o_p(1).
\end{aligned}$$

Proof. By B1, B3 and B5, we have

$$\mathbb{P}_n\dot{m}_1(\hat{\theta}_n, \tilde{\phi}_n, \tilde{\eta}_n; X) - P\dot{m}_1(\hat{\theta}_n, \tilde{\phi}_n, \tilde{\eta}_n; X) - \mathbb{P}_n\dot{m}_1(\theta_0, \phi_0, \eta_0; X) = o_p(n^{-1/2}).$$

In view of B4, this reduces to

$$P\dot{m}_1(\hat{\theta}_n, \tilde{\phi}_n, \tilde{\eta}_n; X) + \mathbb{P}_n\dot{m}_1(\theta_0, \phi_0, \eta_0; X) = o_p(n^{-1/2}).$$

Then by B6, it follows that

$$\begin{aligned}
&P\ddot{m}_{11}(\theta_0, \phi_0, \eta_0; X)(\hat{\theta}_n - \theta_0) + P\ddot{m}_{12}(\theta_0, \phi_0, \eta_0; X)(\tilde{\phi}_n - \phi_0) \\
&\quad +P\ddot{m}_{13}(\theta_0, \phi_0, \eta_0; X)[\tilde{\eta}_n - \eta_0] + \mathbb{P}_n\dot{m}_1(\theta_0, \phi_0, \eta_0; X) \\
&\quad +o(|\hat{\theta}_n - \theta_0|) + o(|\tilde{\phi}_n - \phi_0|) + O(\|\tilde{\eta}_n - \eta_0\|^\mu) = o_p(n^{-1/2}).
\end{aligned}$$

Thus,

$$\begin{aligned}
&-(A + o_p(1))(\hat{\theta}_n - \theta_0) \\
&\quad = -P\{\ddot{m}_{12}(\theta_0, \phi_0, \eta_0; X)(\tilde{\phi}_n - \phi_0) + \ddot{m}_{13}(\theta_0, \phi_0, \eta_0; X)[\tilde{\eta}_n - \eta_0]\} \\
&\qquad -\mathbb{P}_n\dot{m}_1(\theta_0, \phi_0, \eta_0; X) + o_p(n^{-1/2}).
\end{aligned}$$

$\square$

Now we provide technical preparations for the proofs of Theorem 4.1 and 4.2. In order to obtain the influence function of the conditional survival function given X in the Cox regression model, we introduce the following notation:

$$
\begin{aligned}
&W_i(s) = 1(V_i \geq s), && N_i(s) = 1(V_i \leq s, \Delta_i = 1),\\
&dA_i(s;\alpha) = W_i(s)\exp(\alpha' X_i)d\Lambda_0(s), && dM_i(s;\alpha) = dN_i(s) - dA_i(s;\alpha)\\
&\bar{M}(s) = \sum_{i=1}^n M_i(s), && J(s) = 1\left(\sum_{i=1}^n W_i(s) > 0\right),\\
&S^{(0)}(u;\alpha) = \mathbb{P}_n\{W(u)\exp(\alpha' X)\}, && s^{(0)}(u;\alpha) = P\{W(u)\exp(\alpha' X)\},\\
&S^{(1)}(u;\alpha) = \mathbb{P}_n\{XW(u)\exp(\alpha' X)\}, && s^{(1)}(u;\alpha) = P\{XW(u)\exp(\alpha' X)\},\\
&S^{(2)}(u;\alpha) = \mathbb{P}_n\{X^{\otimes 2}W(u)\exp(\alpha' X)\}, && s^{(2)}(u;\alpha) = P\{X^{\otimes 2}W(u)\exp(\alpha' X)\},\\
&\zeta(u;\alpha) = s^{(1)}(u;\alpha)/s^{(0)}(u;\alpha).
\end{aligned}
$$

Lemma A.3. *Under Conditions 5-7, we have*

$$
\begin{aligned}
&\sqrt{n}(\tilde{\eta}_n(t;\tilde{X}) - \eta_0(t;\tilde{X}))\\
&\quad = [1-\eta_0(t;\tilde{X})]\exp(\alpha_0'\tilde{X})\mathbb{G}_n\{A_1(\eta_0;t,\tilde{X};X,\Delta,V)\} + o_p(1),
\end{aligned}
$$

where

$$
\begin{aligned}
&A_1(\eta_0;t,\tilde{X};X,\Delta,V)\\
&\quad = \left[\tilde{X}' + h(t;\alpha_0)'\right] e(\alpha_0)^{-1}\Bigg[-\int_0^\tau \{X - \zeta(u;\alpha_0)\}\exp(\alpha_0' X)W(u)d\Lambda_0(u)\\
&\qquad + 1(V \leq t)\Delta/s^{(0)}(V;\alpha_0) + \{X - \zeta(V;\alpha_0)\}\Delta\\
&\qquad - \int_0^t \exp(\alpha_0' X)W(u)/s^{(0)}(u;\alpha_0)d\Lambda_0(u)\Bigg],
\end{aligned}
$$

with $h(t;\alpha_0) = -\int_0^t \zeta(u;\alpha_0)d\Lambda_0(u)$ *and* $e(\alpha) = P[\Delta\{s^{(2)}(V;\alpha)s^{(0)}(V;\alpha) - s^{(1)}(V;\alpha)^{\otimes 2}\}s^{(0)}(V;\alpha)^{-2}]$ *is the Fisher information matrix for the Cox regression.*

Proof. From Nan and Wellner (2013) or Theorem 8.3.2 of Fleming and Harrington (1991), we have

$$
\begin{aligned}
&\sqrt{n}(\tilde{\alpha}_n - \alpha_0) \qquad (1)\\
&= e(\alpha_0)^{-1}\mathbb{G}_n\left[\{X - \zeta(V;\alpha_0)\}\Delta - \int_0^\tau \{X - \zeta(u;\alpha_0)\}\exp(\alpha_0' X)W(u)d\Lambda_0(u)\right]\\
&\qquad + o_p(1).
\end{aligned}
$$

From the proof of Theorem 8.3.3 in Fleming and Harrington (1991), we have

$$
\sqrt{n}(\tilde{\Lambda}_n(t) - \Lambda_0(t)) = h(t;\alpha_0)'\sqrt{n}(\tilde{\alpha}_n - \alpha_0) + n^{-1/2}\int_0^t J(u)d\bar{M}(u)/S^{(0)}(u;\alpha_0) + o_p(1). \quad (2)
$$

The second term in the right hand side of equation (2) equals

$$
\begin{aligned}
&n^{-1/2}\int_0^t J(u)\left[\frac{1}{S^{(0)}(u;\alpha_0)}-\frac{1}{s^{(0)}(u;\alpha_0)}\right]d\bar{M}(u)\\
&\qquad +n^{-1/2}\int_0^t \{J(u)-1\}/s^{(0)}(u;\alpha_0)d\bar{M}(u)+n^{-1/2}\int_0^t d\bar{M}(u)/s^{(0)}(u;\alpha_0).
\end{aligned}
$$

Define martingales

$$
\begin{aligned}
A(t)&=n^{-1/2}\int_0^t J(u)\Big[1/S^{(0)}(u;\alpha_0)-1/s^{(0)}(u;\alpha_0)\Big]d\bar{M}(u),\\
B(t)&=n^{-1/2}\int_0^t \{J(u)-1\}/s^{(0)}(u;\alpha_0)d\bar{M}(u),
\end{aligned}
$$

with predictable variation processes:

$$
\begin{aligned}
&< A(t),A(t) >\\
&=\frac{1}{n}\sum_{i=1}^n\int_0^t\left\{J(u)\Big[1/S^{(0)}(u;\alpha_0)-1/s^{(0)}(u;\alpha_0)\Big]\right\}^2 W_i(u)\exp(\alpha_0'X)d\Lambda_0(u)\\
&=\int_0^t\left\{J(u)\Big[1/S^{(0)}(u;\alpha_0)-1/s^{(0)}(u;\alpha_0)\Big]\right\}^2 S^{(0)}(u;\alpha_0)d\Lambda_0(u)\xrightarrow{p}0,\\
&< B(t),B(t) >=\frac{1}{n}\sum_{i=1}^n\int_0^t[J(u)-1]^2 s^{(0)}(u;\alpha_0)^{-2}W_i(u)\exp(\alpha_0'X_i)d\Lambda_0(u)\\
&\qquad=\int_0^t[J(u)-1]^2 s^{(0)}(u;\alpha_0)^{-2}S^{(0)}(u;\alpha_0)d\Lambda_0(u)\xrightarrow{p}0.
\end{aligned}
$$

Hence, $A(t)\to 0$ and $B(t)\to 0$ for any t, and

$$
\sqrt{n}(\tilde{\Lambda}_n(t)-\Lambda_0(t))=h(t;\alpha_0)'\sqrt{n}(\tilde{\alpha}_n-\alpha_0)+\sum_{i=1}^n\int_0^t n^{-1/2}s^{(0)}(u;\alpha_0)^{-1}dM_i(u)+o_p(1).
$$

From the Taylor expansion,

$$
\begin{aligned}
&\sqrt{n}(\tilde{\eta}_n(t;\tilde{X})-\eta_0(t;\tilde{X}))\\
&=-\sqrt{n}\Big[\exp\{-\tilde{\Lambda}_n(t)\exp(\tilde{\alpha}_n'\tilde{X})\}-\exp\{-\Lambda_0(t)\exp(\alpha_0'\tilde{X})\}\Big]\\
&=-\sqrt{n}\Big[-\exp\{-\Lambda_0(t)\exp(\alpha_0'\tilde{X})\}\exp(\alpha_0'\tilde{X})(\tilde{\alpha}_n-\alpha_0)'\tilde{X}\\
&\qquad-\exp\{-\Lambda_0(t)\exp(\alpha_0'\tilde{X})\}\exp(\alpha_0'\tilde{X})[\tilde{\Lambda}_n(t)-\Lambda_0(t)]\\
&\qquad+o(|\tilde{\alpha}_n-\alpha_0|)+o(\|\tilde{\Lambda}_n-\Lambda_0\|)\Big].
\end{aligned}
$$

$\square$

Lemma A.4. *Under Conditions 1, 3-5, and 8, the class of functions defined in (8), $\{l(\theta, \phi, \eta(\alpha, \Lambda); Y, X, \Delta, V) : \theta \in \Theta, \phi \in \Phi, \eta \in \mathcal{F}\}$, belongs to Donsker class.*

Proof. We first want to shown that $\{\log f_{\theta,\phi}(Y|u, X) : \theta \in \Theta, \phi \in \Phi\}$ is Donsker. Under Conditions 1 and 3-5, we have $g(S - t, \xi)$ is Lipschitz function for ξ, $\log|\Sigma(\phi)|$ and all the elements in $\Sigma(\phi)$ are Lipschitz functions for ϕ, thus all belong to Donsker by Theorem 2.10.6 of van der Vaart and Wellner (1996). By Condition 3 and Theorem 2.10.6 of van der Vaart and Wellner (1996), we have $\{f_{\theta,\phi}(Y|u, X)\}$ is Donsker. By integration by parts,

$$\begin{aligned}
&\int_C^\tau f_{\theta,\phi}(Y|u, X)d[1 - \exp\{-\Lambda(u)\exp(\alpha' X)\}] \\
&= f_{\theta,\phi}(Y|\tau, X)[1 - \exp\{-\Lambda(\tau)\exp(\alpha' X)\}] - f_{\theta,\phi}(Y|C, X)[1 - \exp\{-\Lambda(C)\exp(\alpha' X)\}] \\
&\quad + \int_C^\tau f_{\theta,\phi}(Y|u, X)\,[\partial \mathbf{g}(u1 - t, \xi)/\partial u]'\,\Sigma(\phi)^{-1}r(\theta)[1 - \exp\{-\Lambda(u)\exp(\alpha' X)\}]du.
\end{aligned}$$

In the above, $\exp\{-\Lambda(u)\exp(\alpha' X)\}$ is Lipschitz for Λ and α from Condition 5, and $\partial \mathbf{g}(u1 - t, \xi)/\partial u$ is Lipschitz function for ξ from Condition 1, thus belong to Donsker classes by Theorem 2.10.6 of van der Vaart and Wellner (1996). By Theorem 2.10.3 of van der Vaart and Wellner (1996), the permanence of the Donsker property for the closure of the convex hull, we have $\{\int_C^\tau f_{\theta,\phi}(Y|u, X)d[1 - \exp\{-\Lambda(u)\exp(\alpha' X)\}] : \theta \in \Theta, \phi \in \Phi, \eta \in \mathcal{F}\}$ is Donsker. By Condition 8, $\delta_1 \le \int_C^\tau f_{\theta,\phi}(Y|u, X)d[1 - \exp\{-\Lambda(u)\exp(\alpha' X)\}] \le \sup_{Y,u,X} f_{\theta,\phi}(Y|u, X)$, which is bounded from Condition 3. Hence, $\{\log \int_C^\tau f_{\theta,\phi}(Y|u, X)d[1 - \exp\{-\Lambda(u)\exp(\alpha' X)\}] : \theta \in \Theta, \phi \in \Phi, \eta \in \mathcal{F}\}$ is Donsker from Theorem 2.10.6 of van der Vaart and Wellner (1996). □

Lemma A.5. *Under Conditions 1, 2(b), 3-4 and 8, we have*
$$\sup_{\theta\in\Theta} |Pl(\theta, \tilde{\phi}_n, \tilde{\eta}_n; Y, X, \Delta, V) - Pl(\theta, \phi_0, \eta_0; Y, X, \Delta, V)| \le O(|\tilde{\phi}_n - \phi_0|) + O(\|\tilde{\eta}_n - \eta_0\|).$$

Proof. From triangular inequality,

$$\begin{aligned}
&\sup_{\theta\in\Theta} |Pl(\theta, \tilde{\phi}_n, \tilde{\eta}_n; Y, X, \Delta, V) - Pl(\theta, \phi_0, \eta_0; Y, X, \Delta, V)| \\
&\le 0.5|P\{\log|\Sigma(\tilde{\phi}_n)| - \log|\Sigma(\phi_0)|\}| \\
&\quad + 0.5\left|P\left\{r(\theta_0; V, Y, X)'\left[\Sigma(\tilde{\phi}_n)^{-1} - \Sigma(\phi_0)^{-1}\right]r(\theta_0; V, Y, X)\right\}\right| \\
&\quad + \sup_{\theta\in\Theta} 0.5\left|P\left\{d(\theta; V, X)'\left[\Sigma(\tilde{\phi}_n)^{-1} - \Sigma(\phi_0)^{-1}\right]d(\theta; V, X)\right\}\right| \\
&\quad + \sup_{\theta\in\Theta}\left|P\left\{\log\int_C^\tau f_{\theta,\tilde{\phi}_n}(Y|u, X)d\tilde{\eta}_n(u; X) - \log\int_C^\tau f_{\theta,\phi_0}(Y|u, X)d\tilde{\eta}_n(u; X)\right\}\right| \\
&\quad + \sup_{\theta\in\Theta}\left|P\left\{\log\int_C^\tau f_{\theta,\phi_0}(Y|u, X)d\tilde{\eta}_n(u; X) - \log\int_C^\tau f_{\theta,\phi_0}(Y|u, X)d\eta_0(u; X)\right\}\right| \\
&\le O(|\tilde{\phi}_n - \phi_0|) \\
&\quad + \sup_{\theta\in\Theta}\left|P\left\{\log\int_C^\tau f_{\theta,\phi_0}(Y|u, X)d\tilde{\eta}_n(u; X) - \log\int_C^\tau f_{\theta,\phi_0}(Y|u, X)d\eta_0(u; X)\right\}\right|,
\end{aligned}$$

where $d(\theta; V, X) = \mathbf{X}(\beta - \beta_0) + [\mathbf{g}(V1 - t, \xi) - \mathbf{g}(V1 - t, \xi_0)]$, which is uniformly bounded for any (V, X) and $\theta \in \Theta$ from Condition 1. The last inequality is obtained from the mean value theorem and Conditions 3-4 and 8. Again, from the mean value theorem and Condition 8,

$$\sup_{\theta\in\Theta}\left|P\left\{\log\int_C^\tau f_{\theta,\phi_0}(Y|u,X)d\tilde{\eta}_n(u;X) - \log\int_C^\tau f_{\theta,\phi_0}(Y|u,X)d\eta_0(u;X)\right|\right\}$$
$$\leq \delta_1 \sup_{\theta\in\Theta} P\left|\left\{\int_C^\tau f_{\theta,\phi_0}(Y|u,X)d[\tilde{\eta}_n(u;X) - \eta_0(u;X)]\right\}\right|.$$

By integration by parts,

$$\sup_{\theta\in\Theta} P\left|\int_C^\tau f_{\theta,\phi_0}(Y|u,X)d\tilde{\eta}_n(u;X) - \int_C^\tau f_{\theta,\phi_0}(Y|u,X)d\eta_0(u;X)\right|$$
$$\leq \sup_{u\in[0,\tau]}\left(2 + \tau P\left\{\left|\partial\mathbf{g}(u1-t,\xi)/\partial u\right|' \Sigma(\phi_0)^{-1}\left|r(\theta;u,Y,X)\right|\right\}\right)$$
$$\times \sup_{\theta\in\Theta, Y\in\mathcal{Y}, X\in\mathcal{X}, u\in[0,\tau]} f_{\theta,\phi_0}(Y|u,X) \times \|\tilde{\eta}_n - \eta_0\| = O(\|\tilde{\eta}_n - \eta_0\|).$$

The last equality holds because all the elements in $\int_{-\infty}^{\infty}|y|f_{\theta_0,\phi_0}(y|u,x)dy$ are bounded uniformly for all $u \in [0, \tau]$ and $x \in \mathcal{X}$ from Kamat (1953). Hence,

$$\sup_{\theta\in\Theta}\left|Pl(\theta, \tilde{\phi}_n, \tilde{\eta}_n; Y, X, \Delta, V) - Pl(\theta, \phi_0, \eta_0; Y, X, \Delta, V)\right|$$
$$\leq O(|\tilde{\phi}_n - \phi_0|) + O(\|\tilde{\eta}_n - \eta_0\|).$$

□

B Additional Simulation

To investigate the impact of model misspecification to the proposed method, we consider the same simulation setup in the main text with equally spaced visit times except that we set the terminal event time $S_i = 1 + S_{0i}$ with S_{0i} generated from an exponential distribution with conditional hazard function

$$\exp\{\theta_0 + \theta_1 X_{1i} + \theta_2 X_{2i}\},$$

where $\theta_1 = -1$, θ_2 varies between 0 and 10, and θ_0 are chosen to yield about 40% censoring rate. In the analysis, however, we only use X_1 as covariate in the Cox model. The results are presented in Figure 1, which show that the misspecification of the Cox model can yield biased results, and the bias increases as the absolute value of θ_2 increases.

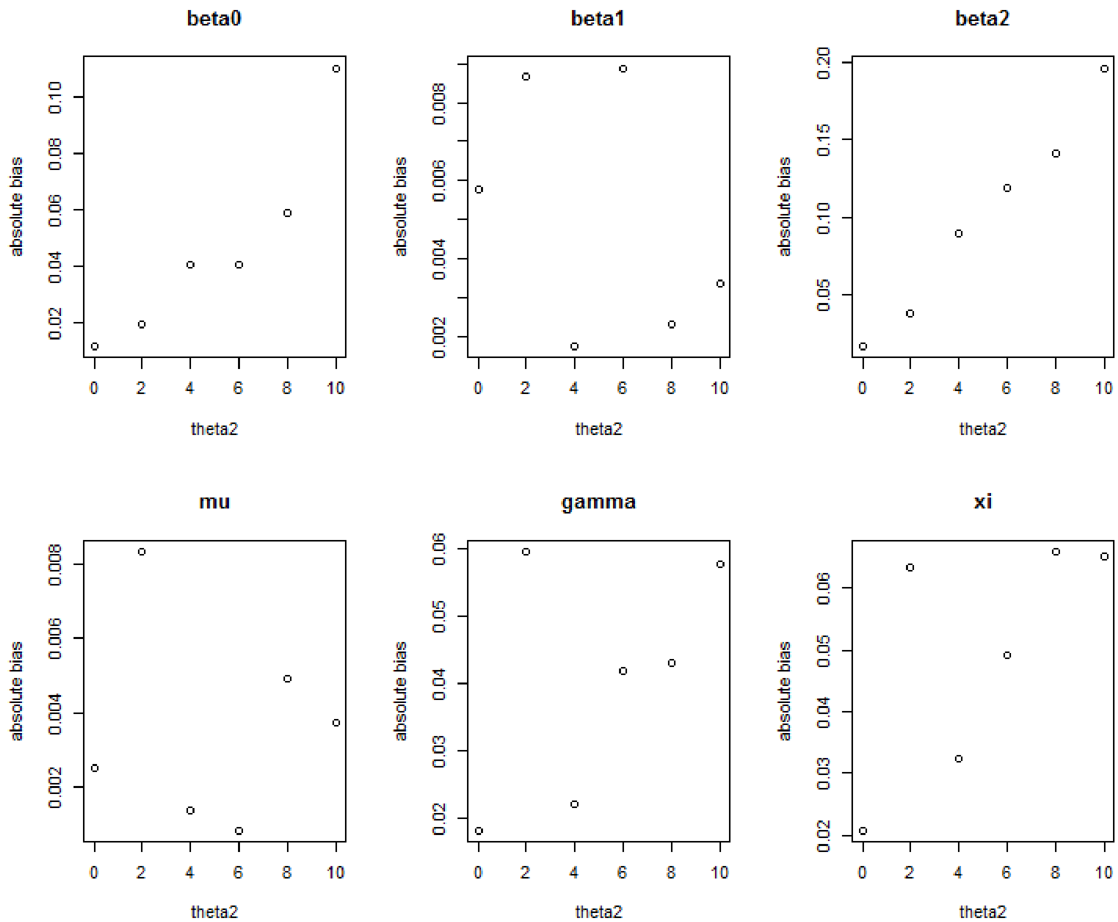

Figure 1: Severity of misspecification of the Cox regression model.

References

Fleming, T. R. and D. P. Harrington (1991). *Counting Processes and Survival Analysis*. John Wiley & Sons, Inc.

Kamat, A. R. (1953). Incomplete and absolute moments of the multivariate normal distribution with some applications. *Biometrika 44*(1-2), 20–34.

Li, Z. and B. Nan (2011). Relative risk regression for current status data in case-cohort studies. *The Canadian Journal of Statistics 39*(4), 557–577.

Nan, B. and J. A. Wellner (2013). A general semiparametric z-estimation approach for case-cohort studies. *Statistica Sinica 23*, 1155–1180.

van der Geer, S. A. (2000). *Empirical Processes in M-Esimation*. Cambridge: Cambridge University Press.

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

Wellner, J. A. and Y. Zhang (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Annals of Statistics 35*(5), 2106–2142.