Web-based Supplementary Materials for "Semiparametric transformation models for semicompeting survival data" by Lin, Zhou, Li and Li

Huazhen Lin, Ling Zhou

Center of Statistical Research, School of Statistics

Southwestern University of Finance and Economics, Chengdu, Sichuan, China

Chunhong Li

Department of Mathematics

The Hong Kong University of Science and Technology, Hong Kong

Yi Li

Department of Biostatistics, University of Michigan, USA

Supplementary Material A: Notations

The following notations are used in Theorem 2. Let

$$f^{(k_1,k_2,\cdots)}(\mathbf{x}_1,\mathbf{x}_2,\cdots)=\frac{d^{(k_1+k_2+\cdots)}f(\mathbf{x}_1,\mathbf{x}_2,\cdots)}{dx_1^{k_1}dx_2^{k_2}\cdots},\ \ \gamma_{1i}(x,y)=-\frac{S_{\rho_0}^{(10)}(x-\mathbf{X}_i'\boldsymbol{\beta}_0,y-\mathbf{X}_i'\boldsymbol{\alpha}_0)}{S_{\rho_0}(x-\mathbf{X}_i'\boldsymbol{\beta}_0,y-\mathbf{X}_i'\boldsymbol{\alpha}_0)},$$

$$\gamma_{2i}(x,y)=-\frac{S_{\rho_0}^{(01)}(x-\mathbf{X}_i'\boldsymbol{\beta}_0,y-\mathbf{X}_i'\boldsymbol{\alpha}_0)}{S_{\rho_0}(x-\mathbf{X}_i'\boldsymbol{\beta}_0,y-\mathbf{X}_i'\boldsymbol{\alpha}_0)},\ \ \gamma_{3i}(x,y)=-\frac{\dot{S}_\rho(x-\mathbf{X}_i'\boldsymbol{\beta},y-\mathbf{X}_i'\boldsymbol{\alpha})}{S_\rho(x-\mathbf{X}_i'\boldsymbol{\beta},y-\mathbf{X}_i'\boldsymbol{\alpha})},$$

$$\dot{S}_\rho(x,y)=\frac{\partial S_\rho(x,y)}{\partial\rho},\ \ \gamma_1\{H_{10}(t)\}=\exp\left(\int_{t_0}^{t}\frac{E[Y_i(s)d\gamma_{1i}\{H_{10}(s),H_{20}(s)\}]}{E[Y_i(s)\gamma_{1i}\{H_{10}(s),H_{20}(s)\}]}\right),$$

$$\gamma_2(H_{20}(t)) = \exp\left(\int_{t_0}^{t} \frac{E[Y_i(s) d\gamma_{2i}\{H_{10}(s), H_{20}(s)\}]}{E[Y_i(s)\gamma_{2i}\{H_{10}(s), H_{20}(s)\}]}\right),$$

$$\lambda_{2i}(x) = \frac{\phi(x - \mathbf{X}_i'\boldsymbol{\alpha}_0)}{1 - \Phi(x - \mathbf{X}_i'\boldsymbol{\alpha}_0)}, \quad \lambda_2(H_{20}(t)) = \exp\left(\int_{t_0}^{t} \frac{E[Y_{2i}(s) d\lambda_{2i}\{H_{20}(s)\}]}{E[Y_{2i}(s)\lambda_{2i}\{H_{20}(s)\}]}\right),$$

$$\mu(t) = \int_{t_0}^{t} \frac{\lambda_2\{H_{20}(s)\} E\left[Y_{2i}(s) \frac{\partial \mathbf{X}_i'\boldsymbol{\alpha}_0}{\partial \boldsymbol{\Theta}} d\lambda_{2i}\{H_{20}(s)\}\right]}{E[Y_{2i}(s)\lambda_{2i}\{H_{20}(s)\}]}, \quad \mathcal{K}(s) = E[Y_{2i}(s)\lambda_{2i}\{H_{20}(s)\}],$$

$$B(s) = E[Y_i(s)\gamma_{1i}\{H_{10}(s), H_{20}(s)\}], \quad \mathcal{A}(s) = E[Y_i(s)\gamma_{2i}\{H_{10}(s), H_{20}(s)\}],$$

$$\mathcal{D}_1(s) = E\left[\frac{\partial^2 \log \mathcal{L}_i(\boldsymbol{\Theta}_0; H_{10}, H_{20})}{\partial \boldsymbol{\Theta} \partial H_1(U_{1i})} \frac{Y_i(s)}{\gamma_1\{H_{10}(U_{1i})\}}\right] \frac{\gamma_1\{H_{10}(s)\}}{B(s)},$$

$$\mathcal{D}_2(s) = E\left\{\frac{\partial^2 \log \mathcal{L}_i(\boldsymbol{\Theta}_0; H_{10}, H_{20})}{\partial \boldsymbol{\Theta} \partial H_2(U_{2i})} \frac{Y_{2i}(s)}{\lambda_2\{H_{20}(U_{2i})\}}\right\} \frac{\lambda_2\{H_{20}(s)\}}{\mathcal{K}(s)} - \frac{\mathcal{D}_1(s)\mathcal{A}(s)}{\mathcal{K}(s)}$$
$$- \int_s^{\tau} \frac{\mathcal{D}_1(v)\mathcal{A}(v)}{\mathcal{K}(s)} \frac{\lambda_2\{H_{20}(s)\}}{\gamma_2\{H_{20}(v)\}} d\frac{\gamma_2\{H_{20}(v)\}}{\lambda_2\{H_{20}(v)\}},$$

$$C(t) = \int_{t_0}^{t} E\left[Y_i(s) \frac{\partial \mathbf{X}_i'\boldsymbol{\beta}_0}{\partial \boldsymbol{\Theta}} d\gamma_{1i}\{H_{10}(s), H_{20}(s)\}\right] + E\left[Y_i(s) \frac{\partial \mathbf{X}_i'\boldsymbol{\alpha}_0}{\partial \boldsymbol{\Theta}} d\gamma_{2i}\{H_{10}(s), H_{20}(s)\}\right]$$
$$- \frac{\partial \rho_0}{\partial \boldsymbol{\Theta}} E\left[Y_i(s) d\gamma_{3i}\{H_{10}(s), H_{20}(s)\}\right],$$

$$\Delta = E\left\{\frac{\partial \log \mathcal{L}_i(\boldsymbol{\Theta}_0; H_{10}, H_{20})}{\partial \boldsymbol{\Theta}} + \int_{t_0}^{\tau} \mathcal{D}_1(s) dM_i(s) + \int_{t_0}^{\tau} \mathcal{D}_2(s) dM_{2i}(s)\right\}^2,$$

$$\boldsymbol{\Sigma} = E\left\{\frac{\partial^2 \log \mathcal{L}_i(\boldsymbol{\Theta}_0; H_{10}, H_{20})}{\partial \boldsymbol{\Theta} \partial \boldsymbol{\Theta}'}\right\} + \int_{t_0}^{\tau} \mathcal{D}_1(s) dC'(s)$$
$$- \int_{t_0}^{\tau} \mathcal{D}_1(s)\mathcal{A}(s) \frac{d\mu'(s)}{\lambda_2\{H_{20}(s)\}} - \int_{t_0}^{\tau} \int_s^{\tau} \frac{\mathcal{D}_1(v)\mathcal{A}(v)}{\gamma_2\{H_{20}(v)\}} d\left[\frac{\gamma_2\{H_{20}(v)\}}{\lambda_2\{H_{20}(v)\}}\right] d\mu'(s)$$
$$+ \int_{t_0}^{\tau} E\left[\frac{\partial^2 \log \mathcal{L}_i(\boldsymbol{\Theta}_0; H_{10}, H_{20})}{\partial \boldsymbol{\Theta} \partial H_2(U_{2i})} \frac{Y_{2i}(s)}{\lambda_2\{H_{20}(U_{2i})\}}\right] d\mu'(s).$$

The following notations are used in Theorem 3. Denote

$$\zeta(t) = \frac{1}{\gamma_1\{H_{10}(t)\}} \int_{t_0}^{t} \frac{\gamma_1\{H_{10}(s)\} dC(s)}{B(s)} - \frac{1}{\gamma_1\{H_{10}(t)\}} \int_{t_0}^{t} \left[\frac{\mathcal{A}(s)\gamma_1\{H_{10}(s)\}}{B(s)\lambda_2\{H_{20}(s)\}}\right.$$
$$\left. + \int_s^{t} \frac{\gamma_1\{H_{10}(v)\}\mathcal{A}(v)}{\gamma_2\{H_{20}(v)\}B(v)} d\left(\frac{\gamma_2\{H_{20}(v)\}}{\lambda_2\{H_{20}(v)\}}\right)\right] d\mu(s).$$

$$\omega_1(s,t) = I(s \le t)\frac{\gamma_1\{H_{10}(s)\}}{B(s)\gamma_1\{H_{10}(t)\}} - \zeta'(t)\boldsymbol{\Sigma}^{-1}\mathcal{D}_1(s),$$

$$\omega_2(s,t) = \zeta'(t)\boldsymbol{\Sigma}^{-1}\mathcal{D}_2(s) + I(s \le t)\left[\frac{\mathcal{A}(s)\gamma_1\{H_{10}(s)\}}{B(s)\mathcal{K}(s)\gamma_1\{H_{10}(t)\}}\right.$$
$$\left. + \frac{\lambda_2\{H_{20}(s)\}}{\mathcal{K}(s)\gamma_1\{H_{10}(t)\}}\int_s^t \frac{\mathcal{A}(v)}{B(v)}\frac{\gamma_1\{H_{10}(v)\}}{\gamma_2\{H_{20}(v)\}}d\frac{\gamma_2\{H_{20}(v)\}}{\lambda_2\{H_{20}(v)\}}\right],$$

$$\Sigma_1(t) = E\left\{-\zeta'(t)\boldsymbol{\Sigma}^{-1}\frac{\partial \log \mathcal{L}_i(\boldsymbol{\Theta}_0; H_{10}, H_{20})}{\partial \boldsymbol{\Theta}}\right.$$
$$\left. + \int_{t_0}^{\tau}\omega_1(s,t)dM_i(s) - \int_{t_0}^{\tau}\omega_2(s,t)dM_{2i}(s)\right\}^2,$$

and

$$\Sigma_2(t) = \frac{1}{\lambda_2^2\{H_{20}(t)\}}E\left[\int_{t_0}^t \frac{\lambda_2\{H_{20}(s)\}}{\mathcal{K}(s)}dM_{2i}(s) - \mu'(t)\boldsymbol{\Sigma}^{-1}\frac{\partial \log \mathcal{L}_i(\boldsymbol{\Theta}_0; H_{10}, H_{20})}{\partial \boldsymbol{\Theta}}\right.$$
$$\left. - \mu'(t)\boldsymbol{\Sigma}^{-1}\int_{t_0}^{\tau}\mathcal{D}_1(s)dM_i(s) - \mu'(t)\boldsymbol{\Sigma}^{-1}\int_{t_0}^{\tau}\mathcal{D}_2(s)dM_{2i}(s)\right]^2.$$

Supplementary Material B: Proofs

The consistency and asymptotic normality stated in Theorems 1 and 2 are proved using arguments similar to those of Chen *et al.* (2002), so we highlight only the steps that are different.

Proof of Theorems 1 and 2.

Step 1. Using similar arguments to Step A1 of Chen *et al.* (2002), it can be shown that $d\{\widehat{H}_2(\cdot, \boldsymbol{\Theta}_0), H_{20}(\cdot)\} \to 0$ almost surely, where $\widehat{H}_2(\cdot, \boldsymbol{\Theta})$ is the function implicitly defined as the unique solution of (2.2) for fixed $\boldsymbol{\Theta}$ and $d(G_1, G_2) = \sup_{t\in[t_0,\tau]}\Big|\exp\{G_1(t)\} - \exp\{G_2(t)\}\Big|$.

Now we show that $D\{\widehat{H}_1(\cdot,\mathbf{\Theta}_0), H_{10}(\cdot)\} \to 0$ almost surely, where $\widehat{H}_1(\cdot,\mathbf{\Theta}_0) \in \mathcal{H}_1$ is the function implicitly defined as the unique solution of (2.3) with $\mathbf{\Theta} = \mathbf{\Theta}_0$, $H_2(\cdot) = \widehat{H}_2(\cdot,\mathbf{\Theta}_0)$ and $D(G_1,G_2) = \sup_{t\in[t_0,\tau]} \Big| E\Big(\log [S_{\rho_0} \{G_1(t\wedge U_{1i}) - \mathbf{X}_i'\boldsymbol{\beta}_0, H_{20}(t\wedge U_{1i}) - \mathbf{X}_i'\boldsymbol{\alpha}_0\}] - \log [S_{\rho_0} \{G_2(t\wedge U_{1i}) - \mathbf{X}_i'\boldsymbol{\beta}_0, H_{20}(t\wedge U_{1i}) - \mathbf{X}_i'\boldsymbol{\alpha}_0\}] \Big)\Big|$, for any two nondecreasing functions G_1 and G_2 on $[t_0,\tau]$ such that $G_1(t_0) = G_2(t_0) = -\infty$. Denote $\mathcal{H}_1 = \{H_1 :$ H_1 is nondecreasing step functions on $[t_0,\tau]$ with $H_1(t_0) = -\infty$ and with jumps only at the observed failure times $t_1,\cdots,t_M\}$, and A is a mapping defined by $A(H_1)(t) = \frac{1}{n}\sum_{i=1}^n \int_{t_0}^t \Big(dN_i(t) + Y_i(t) d\log\Big[S_{\rho_0}\Big\{H_1(t) - \mathbf{X}_i'\boldsymbol{\beta}_0, \widehat{H}_2(t,\mathbf{\Theta}_0) - \mathbf{X}_i'\boldsymbol{\alpha}_0\Big\}\Big]\Big)$, where ρ_0, $\boldsymbol{\beta}_0$ and $\boldsymbol{\alpha}_0$ are the true values of ρ, $\boldsymbol{\beta}$ and $\boldsymbol{\alpha}$. For an arbitrary but fixed $\epsilon > 0$, consider G_1 and G_2 such that $D(G_1,G_2) > \epsilon$, then there exists a $t^* \in [t_0,\tau]$ such that

$$\Big| E\Big[\log\{S_{\rho_0}(G_1(t^*\wedge U_{1i}) - \mathbf{X}_i'\boldsymbol{\beta}_0, H_{20}(t^*\wedge U_{1i}) - \mathbf{X}_i'\boldsymbol{\alpha}_0)\} - \log\{S_{\rho_0}(G_2(t^*\wedge U_{1i}) - \mathbf{X}_i'\boldsymbol{\beta}_0, H_{20}(t^*\wedge U_{1i}) - \mathbf{X}_i'\boldsymbol{\alpha}_0)\}\Big]\Big| \ge \epsilon/2.$$

Hence, coupling with $d\{\widehat{H}_2(\cdot,\mathbf{\Theta}_0), H_{20}(\cdot)\} \to 0$ almost surely, we have

$$\begin{aligned}\sup_{t\in[t_0,\tau]} &\Big| A(G_1)(t) - A(G_2)(t) \Big| \\ &\ge \frac{1}{n}\Big| \sum_{i=1}^n \Big(\log\Big[S_{\rho_0}\Big\{G_1(t^*\wedge U_{1i}) - \mathbf{X}_i'\boldsymbol{\beta}_0, \widehat{H}_2(t^*\wedge U_{1i},\mathbf{\Theta}_0) - \mathbf{X}_i'\boldsymbol{\alpha}_0\Big\}\Big] \\ &\quad - \log\Big[S_{\rho_0}\Big\{G_2(t^*\wedge U_{1i}) - \mathbf{X}_i'\boldsymbol{\beta}_0, \widehat{H}_2(t^*\wedge U_{1i},\mathbf{\Theta}_0) - \mathbf{X}_i'\boldsymbol{\alpha}_0\Big\}\Big]\Big)\Big| \ge \epsilon/2\end{aligned}$$

when n is large enough. Choose $H_{10}^* \in \mathcal{H}_1$ such that $H_{10}^*(t_i) = H_{10}(t_i)$ for $i = 1,\cdots,M$. By the law of large numbers, the continuity of H_{10} and $d\{\widehat{H}_2(\cdot,\mathbf{\Theta}_0), H_{20}(\cdot)\} \to 0$ implies that $\sup\{A(H_{10}^*)(t) : t\in[t_0,\tau]\} \to 0$ almost surely. It follows that $\sup_{t\in[t_0,\tau]} \Big| A(H_{10}^*)(t) - A\{\widehat{H}_1(\cdot,\mathbf{\Theta}_0)\}(t)\Big| \to 0$ almost surely because, by definition of $\widehat{H}_1(\cdot,\mathbf{\Theta}_0)$, $A\{\widehat{H}_1(\cdot,\mathbf{\Theta}_0)\}(t) = 0$ for all $t\in[t_0,\tau]$. Then, with probability 1, $\widehat{H}_1(\cdot,\mathbf{\Theta}_0)$ is in the neighborhood of H_{10}^* of radius ϵ under the metric $D(\cdot,\cdot)$. Therefore, $D(\widehat{H}_1(\cdot,\mathbf{\Theta}_0),H_{10})\to 0$ almost surely, since $\epsilon>0$ can be arbitrarily small and $\widehat{H}_1(\cdot,\mathbf{\Theta}_0)$ and H_{10} are monotone.

Step 2. Constructing the expressions of $\widehat{H}_2(t;\mathbf{\Theta}_0)$ and $\widehat{H}_1(t;\mathbf{\Theta}_0)$. Let $a>0$, let b be fixed finite numbers and define $\mathcal{K}_1(t)=\int_a^t E[Y_{2i}(s)\lambda_{2i}^{(1)}\{H_{20}(s)\}]dH_{20}(s)$, $\Gamma_1(x)=\int_b^x\gamma_1(s)ds$, $\Lambda_2(x)=\int_b^x\lambda_2(s)ds$, for $t>t_0$ and $x\in(-\infty,\infty)$. We choose finite $a>0$ and b as the lower limits of the integration to ensure that the integrals are finite. Similar to Step A2 in Chen *et al.* (2002), we have, uniformly for $t\in[t_0,\tau]$,

$$\Lambda_2\{\widehat{H}_2(t;\mathbf{\Theta}_0)\}-\Lambda_2\{H_{20}(t)\}=\frac{1}{n}\sum_{i=1}^n\int_{t_0}^t\frac{\lambda_2\{H_{20}(s)\}}{\mathcal{K}(s)}dM_{2i}(s)+o_p(n^{-1/2}). \tag{S.1}$$

Now we consider the representative of $\widehat{H}_1(t;\mathbf{\Theta}_0)$. Denote

$$B_1(t)=\int_a^t E[Y_i(s)d\gamma_{1i}\{H_{10}(s),H_{20}(s)\}],\ \ \Gamma_2(x)=\int_b^x\gamma_2(s)ds,$$

$$\mathcal{A}_1(t)=\int_a^t E[Y_i(s)d\gamma_{2i}\{H_{10}(s),H_{20}(s)\}],\ \ \mathcal{A}(t)=E[Y_i(t)\gamma_{2i}\{H_{10}(t),H_{20}(t)\}].$$

Thus, it is easy to see that $d\gamma_1\{H_{10}(t)\}=[\gamma_1\{H_{10}(t)\}/B(t)]dB_1(t)$ and $d\gamma_2\{H_{20}(t)\}=[\gamma_2\{H_{20}(t)\}/\mathcal{A}(t)]d\mathcal{A}_1(t)$ and therefore:

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^n M_i(t)&=\frac{1}{n}\sum_{i=1}^n\int_{t_0}^t Y_i(s)d\left(\frac{\gamma_{1i}\{H_{10}(s),H_{20}(s)\}}{\gamma_1\{H_{10}(s)\}}\left[\Gamma_1\{\widehat{H}_1(s;\mathbf{\Theta}_0)\}-\Gamma_1\{H_{10}(s)\}\right]\right)\\&\quad+\frac{1}{n}\sum_{i=1}^n\int_{t_0}^t Y_i(s)d\left(\frac{\gamma_{2i}\{H_{10}(s),H_{20}(s)\}}{\gamma_2\{H_{20}(s)\}}\left[\Gamma_2\{\widehat{H}_2(s;\mathbf{\Theta}_0)\}-\Gamma_2\{H_{20}(s)\}\right]\right)+o_p(n^{-1/2})\\&=\int_{t_0}^t\frac{B(s)}{\gamma_1\{H_{10}(s)\}}d\left[\Gamma_1\{\widehat{H}_1(s;\mathbf{\Theta}_0)\}-\Gamma_1\{H_{10}(s)\}\right]\\&\quad+\int_{t_0}^t\frac{\mathcal{A}(s)}{\gamma_2\{H_{20}(s)\}}d\left[\Gamma_2\{\widehat{H}_2(s;\mathbf{\Theta}_0)\}-\Gamma_2\{H_{20}(s)\}\right]+o_p(n^{-1/2}).\end{aligned} \tag{S.2}$$

Denote $\Upsilon(t)=\Gamma_2\{\Lambda_2^{-1}(t)\}$. By (S.1), uniformly for $t\in[t_0,\tau]$, we have

$$\begin{aligned}\Gamma_2\{\widehat{H}_2(t;\mathbf{\Theta}_0)\}-\Gamma_2\{H_{20}(t)\}&=\Upsilon[\Lambda_2\{\widehat{H}_2(t;\mathbf{\Theta}_0)\}]-\Upsilon[\Lambda_2\{H_{20}(t)\}]\\&=\frac{\gamma_2\{H_{20}(t)\}}{n\lambda_2\{H_{20}(t)\}}\sum_{i=1}^n\int_{t_0}^t\frac{\lambda_2\{H_{20}(s)\}}{\mathcal{K}(s)}dM_{2i}(s)+o_p(n^{-1/2}).\end{aligned} \tag{S.3}$$

Substituting it into (S.2), we get

$$\Gamma_1\{\widehat{H}_1(t;\mathbf{\Theta}_0)\} - \Gamma_1\{H_{10}(t)\} = \frac{1}{n}\sum_{i=1}^{n}\int_{t_0}^{t}\frac{\gamma_1\{H_{10}(s)\}}{B(s)}dM_i(s) - \frac{1}{n}\sum_{i=1}^{n}\int_{t_0}^{t}\left[\frac{\mathcal{A}(s)}{B(s)}\frac{\gamma_1\{H_{10}(s)\}}{\mathcal{K}(s)}\right.$$
$$\left. + \frac{\lambda_2\{H_{20}(s)\}}{\mathcal{K}(s)}\int_s^t \frac{\mathcal{A}(v)}{B(v)}\frac{\gamma_1\{H_{10}(v)\}}{\gamma_2\{H_{20}(v)\}}d\frac{\gamma_2\{H_{20}(v)\}}{\lambda_2\{H_{20}(v)\}}\right] dM_{2i}(s) + o_p(n^{-1/2})$$
$$\widehat{=}\frac{1}{n}\sum_{i=1}^{n}\int_{t_0}^{t}\Psi_1(s)dM_i(s) - \frac{1}{n}\sum_{i=1}^{n}\int_{t_0}^{t}\Psi_2(s,t)dM_{2i}(s) + o_p(n^{-1/2}), \tag{S.4}$$

uniformly over $t \in [t_0, \tau]$.

Step 3. Denote $\mathcal{L}(\mathbf{\Theta}; H_1, H_2) = \prod_{i=1}^{n}\mathcal{L}_i(\mathbf{\Theta}; H_1, H_2)$ and $U(\mathbf{\Theta}; H_1, H_2) = \frac{\partial \log \mathcal{L}(\mathbf{\Theta}; H_1, H_2)}{\partial \mathbf{\Theta}}$. In the step, we compute $V(\mathbf{\Theta}) = \frac{\partial U(\mathbf{\Theta}; \widehat{H}_1(\cdot;\mathbf{\Theta}), \widehat{H}_2(\cdot;\mathbf{\Theta}))}{n\partial \mathbf{\Theta}}$ at $\mathbf{\Theta} = \mathbf{\Theta}_0$. By differentiating both side of (2.2) with $H_2(t)$ replaced by $\widehat{H}_2(t;\mathbf{\Theta})$, respect to $\mathbf{\Theta}$, we obtain the identity

$$\sum_{i=1}^{n}\int_{t_0}^{t} Y_{2i}(s)d\left[\lambda_{2i}\{\widehat{H}_2(s;\mathbf{\Theta})\}\left\{\frac{\partial \widehat{H}_2(s;\mathbf{\Theta})}{\partial \mathbf{\Theta}} - \frac{\partial \mathbf{X}_i'\boldsymbol{\alpha}}{\partial \mathbf{\Theta}}\right\}\right]\Bigg|_{\mathbf{\Theta}=\mathbf{\Theta}_0} = 0. \tag{S.5}$$

Similar to Step 2, we see that

$$\frac{\partial \widehat{H}_2(t;\mathbf{\Theta}_0)}{\partial \mathbf{\Theta}} = \frac{\mu(t)}{\lambda_2\{H_{20}(t)\}} + o_p(1), \tag{S.6}$$

where $\mu(t)$ is defined in Section 3.

However, by differentiating both side of (2.3) with $H_1(t)$ and $H_2(t)$ replaced by $\widehat{H}_1(t;\mathbf{\Theta})$ and $\widehat{H}_2(t;\mathbf{\Theta})$, respect to $\mathbf{\Theta}$, we obtain the identity

$$\sum_{i=1}^{n} Y_i(t)d\left[\gamma_{1i}\{\widehat{H}_1(t,\mathbf{\Theta}), \widehat{H}_2(t,\mathbf{\Theta})\}\left\{\frac{\partial \widehat{H}_1(t,\mathbf{\Theta})}{\partial \mathbf{\Theta}} - \frac{\partial \mathbf{X}_i'\boldsymbol{\beta}}{\partial \mathbf{\Theta}}\right\}\right.$$
$$\left. + \gamma_{2i}\{\widehat{H}_1(t,\mathbf{\Theta}), \widehat{H}_2(t,\mathbf{\Theta})\}\left\{\frac{\partial \widehat{H}_2(t,\mathbf{\Theta})}{\partial \mathbf{\Theta}} - \frac{\partial \mathbf{X}_i'\boldsymbol{\alpha}}{\partial \mathbf{\Theta}}\right\} + \gamma_{3i}\{\widehat{H}_1(t,\mathbf{\Theta}), \widehat{H}_2(t,\mathbf{\Theta})\}\frac{\partial \rho}{\partial \mathbf{\Theta}}\right] = 0,$$

where $\gamma_{3i}(x,y) = -\frac{\dot{S}_\rho(x - \mathbf{X}_i'\boldsymbol{\beta}, y - \mathbf{X}_i'\boldsymbol{\alpha})}{S_\rho(x - \mathbf{X}_i'\boldsymbol{\beta}, y - \mathbf{X}_i'\boldsymbol{\alpha})}$ and $\dot{S}_\rho(x,y) = \frac{\partial S_\rho(x,y)}{\partial \rho}$. Similar to Step 2, we now see that

$$\frac{\partial \widehat{H}_1(t,\mathbf{\Theta})}{\partial \mathbf{\Theta}}\gamma_1\{H_{10}(t)\} + \int_{t_0}^{t} \frac{\mathcal{A}(s)\gamma_1\{H_{10}(s)\}}{B(s)\gamma_2\{H_{20}(s)\}} d\left[\gamma_2\{H_{20}(s)\}\frac{\partial \widehat{H}_2(s,\mathbf{\Theta})}{\partial \mathbf{\Theta}}\right]$$
$$= \int_{t_0}^{t} \gamma_1\{H_{10}(s)\}\frac{dC(s)}{B(s)} + o_p(1),$$

for $\mathbf{\Theta} = \mathbf{\Theta}_0$, where $C(s)$ is defined in Section 3. Then, using (S.6), we get for $\mathbf{\Theta} = \mathbf{\Theta}_0$,

$$\frac{\partial \widehat{H}_1(t,\mathbf{\Theta})}{\partial \mathbf{\Theta}} = \frac{1}{\gamma_1\{H_{10}(t)\}}\int_{t_0}^{t} \frac{\gamma_1\{H_{10}(s)\}}{B(s)} dC(s) - \frac{1}{\gamma_1\{H_{10}(t)\}}\int_{t_0}^{t}\left(\frac{\mathcal{A}(s)\gamma_1\{H_{10}(s)\}}{B(s)\lambda_2\{H_{20}(s)\}}\right.$$
$$\left. + \int_{s}^{t} \frac{\gamma_1\{H_{10}(v)\}\mathcal{A}(v)}{\lambda_2\{H_{20}(v)\}B(v)}\left[\frac{d\mathcal{A}_1(v)}{\mathcal{A}(v)} - \frac{d\lambda_2\{H_{20}(v)\}}{\lambda_2\{H_{20}(v)\}}\right]\right) d\mu(s) + o_p(1). \qquad \text{(S.7)}$$

It follows from the law of large numbers that

$$V(\mathbf{\Theta}_0) = \frac{1}{n}\frac{\partial^2 \log \mathcal{L}(\mathbf{\Theta}_0; H_{10}, H_{20})}{\partial \mathbf{\Theta}\partial \mathbf{\Theta}'} + \frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log \mathcal{L}_i(\mathbf{\Theta}_0; H_{10}, H_{20})}{\partial \mathbf{\Theta}\partial H_1(U_{1i})}\frac{\partial \widehat{H}_1(U_{1i};\mathbf{\Theta}_0)}{\partial \mathbf{\Theta}'}$$
$$+\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log \mathcal{L}_i(\mathbf{\Theta}_0; H_{10}, H_{20})}{\partial \mathbf{\Theta}\partial H_2(U_{2i})}\frac{\partial \widehat{H}_2(U_{2i};\mathbf{\Theta}_0)}{\partial \mathbf{\Theta}'} = \mathbf{\Sigma} + o_p(1),$$

where $\mathcal{L}_i(\mathbf{\Theta}; H_1, H_2)$ is the contribution of subject i to the likelihood function $\mathcal{L}(\mathbf{\Theta}; H_1, H_2)$.

Step 4. In the step, we show the asymptotic normality of $U(\mathbf{\Theta}_0; \widehat{H}_1(\cdot,\mathbf{\Theta}_0), \widehat{H}_2(\cdot,\mathbf{\Theta}_0))$. Using the results of Steps 1 and 2 and some empirical process approximation techniques, we can write

$$\frac{1}{n}U\{\mathbf{\Theta}_0; \widehat{H}_1(\cdot,\mathbf{\Theta}_0), \widehat{H}_2(\cdot,\mathbf{\Theta}_0)\} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log \mathcal{L}_i(\mathbf{\Theta}_0; H_{10}, H_{20})}{\partial \mathbf{\Theta}}$$
$$+\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log \mathcal{L}_i(\mathbf{\Theta}_0; H_{10}, H_{20})}{\partial \mathbf{\Theta}\partial H_1(U_{1i})}\frac{1}{\gamma_1\{H_{10}(U_{1i})\}}\left[\Gamma_1\{\widehat{H}_1(U_{1i};\mathbf{\Theta}_0)\} - \Gamma_1\{H_{10}(U_{1i})\}\right]$$
$$+\frac{1}{n}\sum_{i=1}^{n} \frac{\partial^2 \log \mathcal{L}_i(\mathbf{\Theta}_0; H_{10}, H_{20})}{\partial \mathbf{\Theta}\partial H_2(U_{2i})}\frac{1}{\gamma_2\{H_{20}(U_{2i})\}}\left[\Gamma_2\{\widehat{H}_2(U_{2i};\mathbf{\Theta}_0)\} - \Gamma_2\{H_{20}(U_{2i})\}\right] + o_p(n^{-1/2})$$

Substituting (S.4) and (S.3) into it, we get

$$
\begin{aligned}
&\frac{1}{n}U(\boldsymbol{\Theta}_0; \widehat{H}_1(\cdot,\boldsymbol{\Theta}_0), \widehat{H}_2(\cdot,\boldsymbol{\Theta}_0)) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{\partial \log \mathcal{L}_i(\boldsymbol{\Theta}_0; H_{10}, H_{20})}{\partial \boldsymbol{\Theta}} + \int_{t_0}^{\tau} \mathcal{D}_1(s)dM_i(s) + \int_{t_0}^{\tau} \mathcal{D}_2(s)dM_{2i}(s)\right\} + o_p(n^{-1/2}),
\end{aligned}
$$

where $\mathcal{D}_1(s)$ and $\mathcal{D}_2(s)$ are defined in Section 3. It then follows that

$$
\frac{1}{\sqrt{n}}U\{\boldsymbol{\Theta}_0; \widehat{H}_1(\cdot,\boldsymbol{\Theta}_0), \widehat{H}_2(\cdot,\boldsymbol{\Theta}_0)\} \to N(0,\Delta),
$$

where Δ is defined in Section 3. The rest of the proof essentially proceeds along the lines of Chen *et al.* (2002) and is omitted here.

Proof of Theorem 3.

By Taylor series expansions, (S.3), (S.6) and Theorem 2, we get,

$$
\begin{aligned}
&\Lambda_2\{\widehat{H}_2(t;\widehat{\boldsymbol{\Theta}})\} - \Lambda_2\{H_{20}(t)\} \\
&\quad= \Lambda_2\{\widehat{H}_2(t;\widehat{\boldsymbol{\Theta}})\} - \Lambda_2\{\widehat{H}_2(t;\boldsymbol{\Theta}_0)\} + \Lambda_2\{\widehat{H}_2(t;\boldsymbol{\Theta}_0)\} - \Lambda_2\{H_{20}(t)\} \\
&\quad= \lambda_2\{H_{20}(t)\}\frac{\partial \widehat{H}_2(t;\boldsymbol{\Theta}_0)}{\partial \boldsymbol{\Theta}}\left(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0\right) + \Lambda_2\{\widehat{H}_2(t;\boldsymbol{\Theta}_0)\} - \Lambda_2\{H_{20}(t)\} + o_p(n^{-1/2}) \\
&\quad= \frac{1}{n}\sum_{i=1}^{n}\left[\int_{t_0}^{t}\frac{\lambda_2\{H_{20}(s)\}}{\mathcal{K}(s)}dM_{2i}(s) - \mu'(t)\Sigma^{-1}\frac{\partial \log \mathcal{L}_i(\boldsymbol{\Theta}_0; H_{10}, H_{20})}{\partial \boldsymbol{\Theta}}\right. \\
&\qquad\qquad \left. -\mu'(t)\Sigma^{-1}\int_{t_0}^{\tau}\mathcal{D}_1(s)dM_i(s) - \mu'(t)\Sigma^{-1}\int_{t_0}^{\tau}\mathcal{D}_2(s)dM_{2i}(s)\right] + o_p(n^{-1/2}).
\end{aligned}
$$

Then the first part of Theorem 3 follows.

By Taylor series expansions, we get,

$$
\begin{aligned}
&\Gamma_1\{\widehat{H}_1(t)\} - \Gamma_1\{H_{10}(t)\} \\
&\quad= \Gamma_1\{\widehat{H}_1(t;\widehat{\boldsymbol{\Theta}})\} - \Gamma_1\{\widehat{H}_1(t;\boldsymbol{\Theta}_0)\} + \Gamma_1\{\widehat{H}_1(t;\boldsymbol{\Theta}_0)\} - \Gamma_1\{H_{10}(t)\} \\
&\quad= \gamma_1\{H_{10}(t)\}\frac{\partial \widehat{H}_1(t;\boldsymbol{\Theta}_0)}{\partial \boldsymbol{\Theta}}\left(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0\right) + \Gamma_1\{\widehat{H}_1(t;\boldsymbol{\Theta}_0)\} - \Gamma_1\{H_{10}(t)\}.
\end{aligned}
$$

Then the second part of Theorem 3 follows from Theorem 2, (S.4) and (S.7).

Proof of the Proposition.

Denote $U^*(\mathbf{\Theta}; H_1^*, H_2^*) = \sum_{i=1}^n \xi_i \frac{\partial \mathcal{L}_i(\mathbf{\Theta}; H_1^*, H_2^*)}{\partial \mathbf{\Theta}}$. Then, for $\|\mathbf{\Theta}_1 - \mathbf{\Theta}_2\| = o(1)$, we have

$$\frac{1}{n}U^*(\mathbf{\Theta}_1; H_1^*, H_2^*) - \frac{1}{n}U^*(\mathbf{\Theta}_2; H_1^*, H_2^*) \\ = V^*(\mathbf{\Theta}_2)(\mathbf{\Theta}_1 - \mathbf{\Theta}_2) + o(\|\mathbf{\Theta}_1 - \mathbf{\Theta}_2\|), \quad \text{(S.8)}$$

where $V^*(\mathbf{\Theta}) = -\frac{\partial U^*(\mathbf{\Theta}; H_1^*, H_2^*)}{n\partial \mathbf{\Theta}}$. Similarly to Step 3 in the proof of Theorem 1 and Theorem 2, combining with $E(\xi) = 1$ and $Var(\xi) = 1$, it follows from the strong law of large numbers that

$$d\{H_1^*(\cdot, \mathbf{\Theta}), \widehat{H}_1(\cdot, \mathbf{\Theta})\} \to 0, \ \ d\{H_2^*(\cdot, \mathbf{\Theta}), \widehat{H}_2(\cdot, \mathbf{\Theta})\} \to 0,$$

$$\frac{\partial H_1^*(t; \widehat{\mathbf{\Theta}})}{\partial \mathbf{\Theta}} \to \zeta(t), \ \ \frac{\partial H_2^*(t; \widehat{\mathbf{\Theta}})}{\partial \mathbf{\Theta}} \to \frac{\mu(t)}{\lambda_2\left\{H_{20}(t)\right\}},$$

almost surely, and furthermore,

$$V^*(\widehat{\mathbf{\Theta}}) \xrightarrow{a.s.} \Sigma. \quad \text{(S.9)}$$

However,

$$\frac{1}{n}U^*(\widehat{\mathbf{\Theta}}; H_1^*, H_2^*) \\ = \frac{1}{n}\sum_{i=1}^n (\xi_i - 1)\frac{\partial \mathcal{L}_i(\widehat{\mathbf{\Theta}}; \widehat{H}_1, \widehat{H}_2)}{\partial \mathbf{\Theta}} + \frac{1}{n}\sum_{i=1}^n \xi_i \left\{\frac{\partial \mathcal{L}_i(\widehat{\mathbf{\Theta}}; H_1^*, H_2^*)}{\partial \mathbf{\Theta}} - \frac{\partial \mathcal{L}_i(\widehat{\mathbf{\Theta}}; \widehat{H}_1, \widehat{H}_2)}{\partial \mathbf{\Theta}}\right\},$$

where the second term is $o_p(1)$. Hence, following from the result of Step 4 in the proof of Theorem 1 and Theorem 2, we have

$$\frac{1}{\sqrt{n}}U^*\{\widehat{\mathbf{\Theta}}; H_1^*, H_2^*\} \to N(0, \Delta). \quad \text{(S.10)}$$

Finally, note that $U^*(\mathbf{\Theta}^*; H_1^*, H_2^*) = 0$, and $U^*(\mathbf{\Theta}_0; H_1^*, H_2^*) \to U^*(\mathbf{\Theta}_0; \widehat{H}_1, \widehat{H}_2) \to U(\mathbf{\Theta}_0; \widehat{H}_1, \widehat{H}_2) \to 0$ almost surely; In addition, with the existence of Σ^{-1} and $\widehat{\mathbf{\Theta}} \xrightarrow{a.s.} \mathbf{\Theta}_0$, we have $\|\mathbf{\Theta}^* - \widehat{\mathbf{\Theta}}\| = o(1)$. Combining (S.8) with (S.10), the proof of the Proposition is completed.

Supplementary Material C: Simulation 3

Simulation 3. For computational easy, we estimate the transformation function by estimating equations, which may be less efficient than the maximum likelihood estimator. To investigate the possible loss due to the using of the estimating equations instead of the maximum likelihood function, we compare the proposed method with the oracle parametric maximum likelihood (PML) method, where the transformation function is correctly specified and all of the parameters are estimated by the maximum likelihood function. We generate data from a simulation setting similar to Simulation 1, except that $H_1(t) = log(t)$ and $H_2(t) = 10\log(t)+t$ in Case 1; $H_1(t) = t+1.5$ and $H_2(t) = 3t+0.5$ in Case 2. To use the PML method, $H_1(t)$ and $H_2(t)$ are specified as $H_1(t) = \theta_1 log(t) + \theta_2$ and $H_2(t) = \theta_3 \log(t)+\theta_4 t$ for Caes 1, and $H_1(t) = \theta_1 t+\theta_2$ and $H_2(t) = \theta_3 t+\theta_4$ for Case 2. The resulting estimators are displayed in Table 5. Table 5 shows that the performance of the proposed method is close to that of the PML although the transformation function of our method is unknown. The results in Table 5 confirm the proposed method is efficient.

Table 5. The bias, SD and RMSE of estimators for Simulation 3.

	$\widehat{\alpha}$		$\widehat{\beta}$		$\widehat{\rho}$	
	Bias(SD)	RMSE	Bias(SD)	RMSE	Bias(SD)	RMSE
			Case 1			
Prop.	0.032(0.126)	0.130	0.021(0.121)	0.122	0.0064(0.0439)	0.0444
PML	0.018(0.126)	0.127	0.019(0.119)	0.121	-0.0001(0.0436)	0.0436
			Case 2			
Prop.	0.024(0.115)	0.117	0.021(0.116)	0.118	0.0082(0.0458)	0.0465
PML	0.005(0.116)	0.116	-0.001(0.116)	0.116	0.0010(0.0450)	0.0450

Supplementary Material D: Simulation 4

Simulation 4. In this simulation, we consider the data with two-dimensional covariate vector, which is a combination of continuous and discrete covariates. The setting in Simulation 4 is similar to that in Simulation 1, except that the covariate $\boldsymbol{X} = (X_1, X_2)$, $H_1(t) = t^3$ and $H_2(t) = \Phi^{-1}(t/2)$, where X_1 is generated uniformly over [-2,2], X_2 is treatment indicator in which $n/2$ subjects receive each of the two groups. The censoring random variable C is distributed uniformly on (0, 5), so that about 20% of T is censored by C and about 24% of S is censored by $C \wedge T$. The resulting estimators are displayed in Table 6, suggesting the proposed method performs well in this setting.

Table 6. The bias, empirical standard deviation and root of mean square error(RMSE) of estimators based on 500 simulations.

	$\widehat{\alpha}_1$			$\widehat{\alpha}_2$		
Method	Bias	SD	RMSE	Bias	SD	RMSE
Proposed	-0.0066	0.0625	0.0628	0.0438	0.1160	0.1240
	$\widehat{\beta}_1$			$\widehat{\beta}_2$		
Proposed	-0.0070	0.0593	0.0597	0.0449	0.1167	0.1250
	$\widehat{\rho}$					
Proposed	0.0046	0.0461	0.0463			

Supplementary Material E: Figures 3 and 4

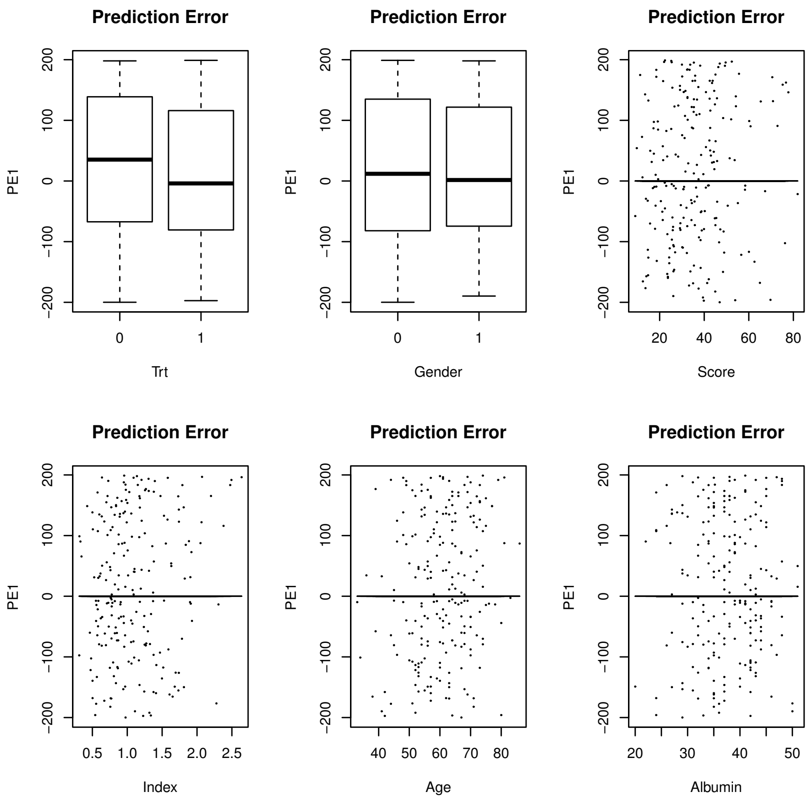

Figure 3: The plot of the prediction error PE_1 versus the covariates using the proposed method for the myeloma data.

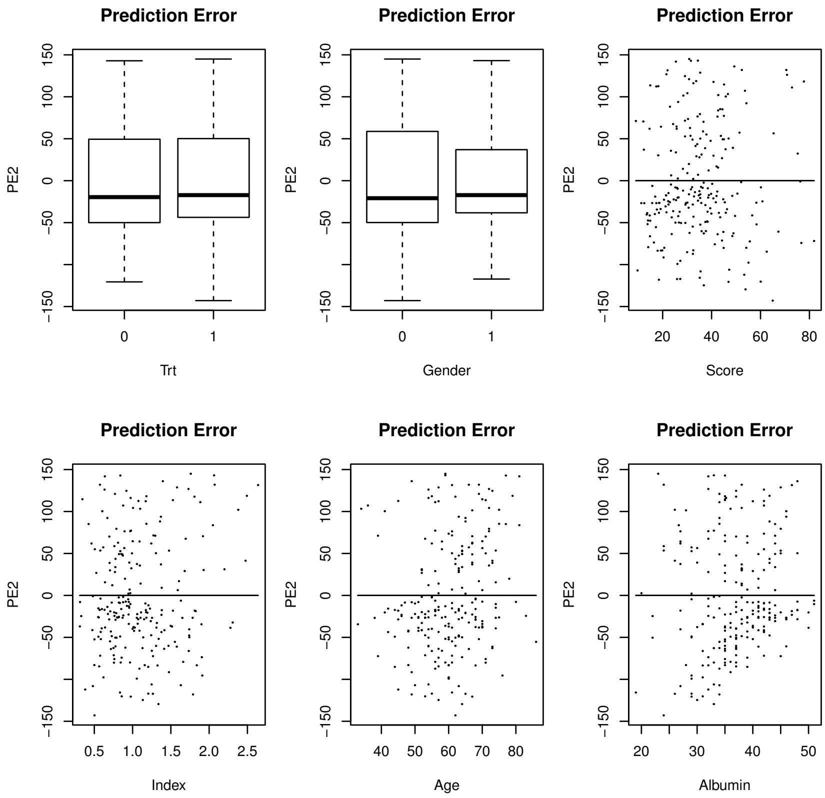

Figure 4: The plot of the prediction error PE_2 versus the covariates using the proposed method for the myeloma data.