Supplementary Materials for Efficient Estimation for Semiparametric Structural Equation Models With Censored Data

Kin Yau Wong, Donglin Zeng, and D. Y. Lin

Department of Biostatistics, University of North Carolina, Chapel Hill, NC 27599

S1. ADDITIONAL LEMMAS

Before proving Lemmas 1 and 2, we present the following two lemmas, which we prove in the next section.

Lemma S1. *Consider functions h and g in the space $l^{\infty}(\mathbb{R}^r \times \mathbb{R}^q)$. Assume that for any $\boldsymbol{Y} \in \mathbb{R}^r$ and $\boldsymbol{b} \in \mathbb{R}^q$, there exist $M_j > 0$, $c_j > 0$, and $\boldsymbol{N}_j \in \mathbb{R}^r$ $(j = 1, \ldots, q)$ such that*

$$\prod_{j=1}^{q} \exp(-M_j\,|b_j| + \boldsymbol{N}_j^{\mathrm{T}}\boldsymbol{Y} b_j - c_j b_j^2) \leq h\left(\boldsymbol{Y}, \boldsymbol{b}\right) \leq \prod_{j=1}^{q} \exp(M_j\,|b_j| + \boldsymbol{N}_j^{\mathrm{T}}\boldsymbol{Y} b_j - c_j b_j^2),$$

and there exists $K > 0$ such that $g(\boldsymbol{Y}, \boldsymbol{b}) \leq \exp\{K(1 + |\boldsymbol{Y}| + |\boldsymbol{b}|)\}$. Then, with $\mathbf{1}_q$ being a q-vector of ones, both

$$\frac{\int \exp\left(-e^{A_1+\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y}+\mathbf{1}_q^{\mathrm{T}}\boldsymbol{b}}\right) h\left(\boldsymbol{Y}, \boldsymbol{b}\right) g\left(\boldsymbol{Y}, \boldsymbol{b}\right)\, \mathrm{d}\boldsymbol{b}}{\int \exp\left(-e^{A_2+\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y}+\mathbf{1}_q^{\mathrm{T}}\boldsymbol{b}}\right) h\left(\boldsymbol{Y}, \boldsymbol{b}\right)\, \mathrm{d}\boldsymbol{b}} \tag{S.1}$$

and

$$\frac{\int h\left(\boldsymbol{Y}, \boldsymbol{b}\right) g\left(\boldsymbol{Y}, \boldsymbol{b}\right)\, \mathrm{d}\boldsymbol{b}}{\int h\left(\boldsymbol{Y}, \boldsymbol{b}\right)\, \mathrm{d}\boldsymbol{b}} \tag{S.2}$$

are bounded by $e^{O(1+|\boldsymbol{Y}|)}$ for any $A_1 \in \mathbb{R}$, $A_2 \in \mathbb{R}$, and $\boldsymbol{B} \in \mathbb{R}^r$.

Lemma S2. *The following classes are Donsker:*

$$\begin{aligned}
\mathcal{C}_1 &= \left\{ \log \Psi(\mathcal{O}_i; \boldsymbol{\theta}, \mathcal{A}) : \|\Lambda_m\|_{V[0,\tau]} \leq c, m = 1, \ldots, K, \boldsymbol{\theta} \in \Theta \right\}, \\
\mathcal{C}_2 &= \left\{ \frac{\dot{\Psi}_\theta(\mathcal{O}_i; \boldsymbol{\theta}, \mathcal{A})}{\Psi(\mathcal{O}_i; \boldsymbol{\theta}, \mathcal{A})} : \|\Lambda_m\|_{V[0,\tau]} \leq c, m = 1, \ldots, K, \boldsymbol{\theta} \in \Theta \right\},
\end{aligned}$$

and

$$\mathcal{C}_{3k} = \left\{ \frac{\dot{\Psi}_k(\mathcal{O}_i; \boldsymbol{\theta}, \mathcal{A})\,[H]}{\Psi(\mathcal{O}_i; \boldsymbol{\theta}, \mathcal{A})} : \|\Lambda_m\|_{V[0,\tau]} \leq c, m = 1, \ldots, K, \boldsymbol{\theta} \in \Theta, \|H\|_{V[0,\tau]} \leq c \right\}$$

for $k = 1, \ldots, K$ and any fixed $c > 0$, where Ψ, $\dot{\Psi}_\theta$, and $\dot{\Psi}_k$ are defined in the proof of Theorem 2.

S2. PROOFS OF LEMMAS

We now prove Lemmas 1-2 and S1-S2.

Proof of Lemma 1. Assume that there are two sets of density functions $(f_{Y|\eta}, f_{\eta_2|\eta_1})$ and $(\tilde{f}_{Y|\eta}, \tilde{f}_{\eta_2|\eta_1})$ such that the marginal densities are identical for all $\boldsymbol{X}$ and $\boldsymbol{Y}$. That is,

$$\begin{aligned} & \int f_{X|\eta_1}(\boldsymbol{X} \mid \boldsymbol{\eta}_1) f_{Y|\eta}(\boldsymbol{Y} \mid \boldsymbol{\eta}) f_{\eta_2|\eta_1}(\boldsymbol{\eta}_2 \mid \boldsymbol{\eta}_1) f_{\eta_1}(\boldsymbol{\eta}_1)\, \mathrm{d}\boldsymbol{\eta} \\ = & \int f_{X|\eta_1}(\boldsymbol{X} \mid \boldsymbol{\eta}_1) \tilde{f}_{Y|\eta}(\boldsymbol{Y} \mid \boldsymbol{\eta}) \tilde{f}_{\eta_2|\eta_1}(\boldsymbol{\eta}_2 \mid \boldsymbol{\eta}_1) f_{\eta_1}(\boldsymbol{\eta}_1)\, \mathrm{d}\boldsymbol{\eta}, \end{aligned}$$

where $\boldsymbol{\eta} = (\boldsymbol{\eta}_1, \boldsymbol{\eta}_2)$. Thus, with $f_{\eta_1|X} = F'_{\eta_1|X}$,

$$\int \left\{ \int f_{Y|\eta}(\boldsymbol{Y} \mid \boldsymbol{\eta}) f_{\eta_2|\eta_1}(\boldsymbol{\eta}_2 \mid \boldsymbol{\eta}_1) - \tilde{f}_{Y|\eta}(\boldsymbol{Y} \mid \boldsymbol{\eta}) \tilde{f}_{\eta_2|\eta_1}(\boldsymbol{\eta}_2 \mid \boldsymbol{\eta}_1)\, \mathrm{d}\boldsymbol{\eta}_2 \right\} f_{\eta_1|X}(\boldsymbol{\eta}_1 \mid \boldsymbol{X})\, \mathrm{d}\boldsymbol{\eta}_1 = 0$$

for all $\boldsymbol{X}$ and $\boldsymbol{Y}$. Because $\boldsymbol{\eta}_1$ is complete sufficient in $F_{\eta_1|X}$,

$$\int f_{Y|\eta}(\boldsymbol{Y} \mid \boldsymbol{\eta}_1, \boldsymbol{\eta}_2) f_{\eta_2|\eta_1}(\boldsymbol{\eta}_2 \mid \boldsymbol{\eta}_1)\, \mathrm{d}\boldsymbol{\eta}_2 = \int \tilde{f}_{Y|\eta}(\boldsymbol{Y} \mid \boldsymbol{\eta}_1, \boldsymbol{\eta}_2) \tilde{f}_{\eta_2|\eta_1}(\boldsymbol{\eta}_2 \mid \boldsymbol{\eta}_1)\, \mathrm{d}\boldsymbol{\eta}_2 \quad \forall \boldsymbol{Y}, \boldsymbol{\eta}_1.$$

By assumption, $(f_{Y|\eta}, f_{\eta_2|\eta_1}) = (\tilde{f}_{Y|\eta}, \tilde{f}_{\eta_2|\eta_1})$. Therefore, Model A is identifiable.

To show the complete sufficiency of $\boldsymbol{\eta}_1$ under the sufficient condition, note that the density of $\boldsymbol{\eta}_1 \mid \boldsymbol{X}$ is of the form

$$f_{\eta_1|X}(\boldsymbol{\eta}_1 \mid \boldsymbol{X}) \;\propto\; f_{X|\eta_1}(\boldsymbol{X} \mid \boldsymbol{\eta}_1) f_{\eta_1}(\boldsymbol{\eta}_1) \propto \exp\left\{ \sum_{j=1}^{q} X_j s_j(\boldsymbol{\eta}_1) \right\} f^*(\boldsymbol{\eta}_1),$$

where f^* is a function of $\boldsymbol{\eta}_1$ that does not involve $\boldsymbol{X}$. Thus, as a property of the exponential family, $s(\boldsymbol{\eta}_1) \equiv (s_1(\boldsymbol{\eta}_1), \ldots, s_q(\boldsymbol{\eta}_1))$ is complete sufficient under the model with parameter $\boldsymbol{X} \in \mathcal{X}$. Because s is a one-to-one function, $\boldsymbol{\eta}_1$ is complete sufficient. $\square$

Proof of Lemma 2. With an abuse of notation, we use $\boldsymbol{\nu}$ to denote all parameters in F_Y and F_η and drop the parameter vector in the arguments of the density functions. We consider the one-dimensional submodel along $(\boldsymbol{h}_\vartheta, \boldsymbol{h}_\beta, \boldsymbol{h}_\alpha, \boldsymbol{h}_\phi, \boldsymbol{h}_\nu, h_1(\cdot), \ldots, h_K(\cdot))$ for parameters $(\boldsymbol{\vartheta}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\phi}, \boldsymbol{\nu}, \Lambda_1, \ldots, \Lambda_K)$, where $\boldsymbol{\vartheta} = (\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_K)$, and $\boldsymbol{h}_\vartheta \equiv (\boldsymbol{h}_{\vartheta 1}, \ldots, \boldsymbol{h}_{\vartheta K})$ is partitioned accordingly. We define $\boldsymbol{\beta}$, $\boldsymbol{\alpha}$, $\boldsymbol{\phi}$, $\boldsymbol{h}_\beta$, $\boldsymbol{h}_\alpha$, and $\boldsymbol{h}_\phi$ in the same way. A one-dimensional submodel along the direction $h \equiv (\boldsymbol{h}_\theta, h_1(\cdot), \ldots, h_K(\cdot)) \in \mathbb{R}^d \times BV[0,\tau]^K$ indexed by ϵ is constructed by setting $\boldsymbol{\theta} = \boldsymbol{\theta}_0 + \epsilon \boldsymbol{h}_\theta$ for the vector of all Euclidean parameters $\boldsymbol{\theta}$ and $\Lambda_k(\cdot) = \int_0^{(\cdot)} \{1 + \epsilon h_k(s)\}\, \mathrm{d}\Lambda_k(s)$ for $k = 1, \ldots, K$. By the arguments in the proof of Theorem 1, we can consider the likelihood with the survival times being right censored at any values within $[0, \tau]$. For an observation with the K survival times right censored at $(t_1, \ldots, t_K)$, the likelihood is given by the left-hand side of (A.1). For simplicity of description, assume that m_k is the Lebesgue measure. If the score is zero almost surely, then

$$
\begin{aligned}
&\int\!\!\int \exp\{-H(\boldsymbol{t})\}\, g(\boldsymbol{s})\, \frac{\partial}{\partial \boldsymbol{\nu}} \{f_Y(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{\eta})\, f_\eta(\boldsymbol{\eta} \mid \boldsymbol{Z})\}^{\mathrm{T}} \boldsymbol{h}_\nu \,\mathrm{d}\boldsymbol{s}\,\mathrm{d}\boldsymbol{\eta} \qquad \text{(S.3)}\\
&- \sum_{k=1}^{K} \int\!\!\int \exp\{-H(\boldsymbol{t})\}\, g(\boldsymbol{s})\, f_Y(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{\eta})\, f_\eta(\boldsymbol{\eta} \mid \boldsymbol{Z})\, s_k e^{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\vartheta}_k + \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}_k + \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_k + \boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\phi}_k}\\
&\quad \times \int_0^{t_k} \boldsymbol{W}^{\mathrm{T}}\boldsymbol{h}_{\vartheta k} + \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{h}_{\beta k} + \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{h}_{\alpha k} + \boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{h}_{\phi k} + h_k(\omega)\, \mathrm{d}\Lambda_k(\omega)\, \mathrm{d}\boldsymbol{s}\,\mathrm{d}\boldsymbol{\eta} \;=\; 0
\end{aligned}
$$

for all $t_1, \ldots, t_K$, $\boldsymbol{W}$, $\boldsymbol{Z}$, and $\boldsymbol{Y}$, where $H(\boldsymbol{t}) = \sum_{k=1}^K \Lambda_k(t_k) s_k e^{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\vartheta}_k + \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}_k + \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_k + \boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\phi}_k}$, $\boldsymbol{s} = (s_1, \ldots, s_K)^{\mathrm{T}}$, and $g(\boldsymbol{s}) = \prod_{k=1}^K g_k(s_k)$. For $k = 1, \ldots, K_1$, we differentiate (S.3) with respect to t_k and then set $t_l \to 0$ for $l = 1, \ldots, K$. Thus,

$$
\int\!\!\int g(\boldsymbol{s})\, \frac{\partial}{\partial \boldsymbol{\nu}} \{f_Y(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{\eta})\, f_\eta(\boldsymbol{\eta} \mid \boldsymbol{Z})\}^{\mathrm{T}} \boldsymbol{h}_\nu s_k e^{\eta_{1k}}\, \mathrm{d}\boldsymbol{\eta}\, \mathrm{d}\boldsymbol{s}
$$

$$-\int\int s_k e^{\eta_{1k}} g(\boldsymbol{s}) f_Y(\boldsymbol{Y} \mid \boldsymbol{Z}, \boldsymbol{\eta}) f_\eta(\boldsymbol{\eta} \mid \boldsymbol{Z})\, \mathrm{d}\boldsymbol{\eta}\, \mathrm{d}\boldsymbol{s} \left\{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{h}_{\vartheta k} + \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{h}_{\beta k} + \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{h}_{\alpha k} + h_k(0)\right\} = 0.$$

By linear independence of $\boldsymbol{W}$, $\boldsymbol{h}_{\vartheta k} = \boldsymbol{0}$.

Consider the first survival time T_1. Because $e^{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\vartheta}_1}$ takes at least two distinct values by conditions (C1) and (C3), we assume, without loss of generality, that it takes 1 and c with $c < 1$. Let $U_1 = s_1 e^{\eta_{11}}$, and let f_{Y,U_1} be the density of $(\boldsymbol{Y}, U_1)$ given $\boldsymbol{Z}$. Setting $t_2, \ldots, t_K \to 0$ and $e^{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\vartheta}_1} = 1$ in (S.3), we have

$$\begin{aligned}\int_0^{t_1} h_1(\omega)\, \mathrm{d}\Lambda_1(\omega) = & \left[\int \exp\left\{-\Lambda_1(t_1) U_1 e^{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}_1 + \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_1}\right\} U_1 e^{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}_1 + \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_1} f_{Y,U_1}(\boldsymbol{Y}, U_1 \mid \boldsymbol{Z})\, \mathrm{d}U_1\right]^{-1} \\ & \int \exp\left\{-\Lambda_1(t_1) U_1 e^{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}_1 + \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_1}\right\} \left\{\frac{\partial}{\partial \boldsymbol{\nu}} f_{Y,U_1}(\boldsymbol{Y}, U_1 \mid \boldsymbol{Z})^{\mathrm{T}} \boldsymbol{h}_\nu \right. \\ & \left. - \Lambda_1(t_1)\left(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{h}_{\beta 1} + \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{h}_{\alpha 1}\right) U_1 e^{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}_1 + \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_1} f_{Y,U_1}(\boldsymbol{Y}, U_1 \mid \boldsymbol{Z})\right\} \mathrm{d}U_1 \\ \equiv\ & a\left\{\Lambda_1(t_1)\right\}. \end{aligned} \tag{S.4}$$

Likewise, setting $t_2, \ldots, t_K \to 0$ and $e^{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\vartheta}_1} = c$ in (S.3), we have $ca\{\Lambda_1(t_1)\} = a\{c\Lambda_1(t_1)\}$. Thus, for all $v \in [0, \Lambda_1(\tau)]$, $a'(c^n v) = a'(v)$ for any integer n. It follows that $a'(v) = a'(0)$, such that a is a linear function. Let $a(v) = \kappa_1 v$. Then, with $v = \Lambda_1(t_1)$, (S.4) becomes

$$\begin{aligned} & \kappa_1 v \int e^{-vU_1 e^{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}_1 + \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_1}} U_1 e^{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}_1 + \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_1} f_{Y,U_1}(\boldsymbol{Y}, U_1 \mid \boldsymbol{Z})\, \mathrm{d}U_1 \\ = & \int e^{-vU_1 e^{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}_1 + \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_1}} \left\{\frac{\partial}{\partial \boldsymbol{\nu}} f_{Y,U_1}(\boldsymbol{Y}, U_1 \mid \boldsymbol{Z})^{\mathrm{T}} \boldsymbol{h}_\nu \right. \\ & \left. -v\left(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{h}_{\beta 1} + \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{h}_{\alpha 1}\right) U_1 e^{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}_1 + \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_1} f_{Y,U_1}(\boldsymbol{Y}, U_1 \mid \boldsymbol{Z})\right\} \mathrm{d}U_1. \end{aligned}$$

Therefore,

$$\int e^{-vU_1 e^{\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}_1 + \boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_1}} \left[\frac{\partial}{\partial \boldsymbol{\nu}} f_{Y,U_1}(\boldsymbol{Y}, U_1 \mid \boldsymbol{Z})^{\mathrm{T}} \boldsymbol{h}_\nu \right.$$

$$- \left(\kappa_1 + \boldsymbol{Z}^{\mathrm{T}} \boldsymbol{h}_{\beta 1} + \boldsymbol{Y}^{\mathrm{T}} \boldsymbol{h}_{\alpha 1}\right) \frac{\partial}{\partial U_1} \left\{U_1 f_{Y,U_1}\left(\boldsymbol{Y}, U_1 \mid \boldsymbol{Z}\right)\right\} \Bigg] \mathrm{d}U_1 = 0$$

for all $v \in [0, \Lambda_1(\tau)]$, $\boldsymbol{Z}$, and $\boldsymbol{Y}$. By the uniqueness of the Laplace transform, the term in the square brackets in the above integral is zero for all U_1, $\boldsymbol{Z}$, and $\boldsymbol{Y}$. Let $f_{Y,\bar{\eta}_1}$ be the density of $(\boldsymbol{Y}, \eta_{11})$ given $\boldsymbol{Z}$. Then,

$$\int \Bigg\{ \frac{1}{U_1} \frac{\partial}{\partial \boldsymbol{\nu}} f_{Y,\bar{\eta}_1}\left(\boldsymbol{Y}, \log U_1 - \log s \mid \boldsymbol{Z}\right)^{\mathrm{T}} \boldsymbol{h}_\nu$$
$$- \left(\kappa_1 + \boldsymbol{Z}^{\mathrm{T}} \boldsymbol{h}_{\beta 1} + \boldsymbol{Y}^{\mathrm{T}} \boldsymbol{h}_{\alpha 1}\right) \frac{\partial}{\partial U_1} f_{Y,\bar{\eta}_1}\left(\boldsymbol{Y}, \log U_1 - \log s \mid \boldsymbol{Z}\right) \Bigg\} g_1(s)\, \mathrm{d}s = 0.$$

By the arguments for convolution in the proof of Theorem 1,

$$\frac{\partial}{\partial \boldsymbol{\nu}} f_{Y,\bar{\eta}_1}\left(\boldsymbol{Y}, \eta_{11} \mid \boldsymbol{Z}\right)^{\mathrm{T}} \boldsymbol{h}_\nu - \left(\kappa_1 + \boldsymbol{Z}^{\mathrm{T}} \boldsymbol{h}_{\beta 1} + \boldsymbol{Y}^{\mathrm{T}} \boldsymbol{h}_{\alpha 1}\right) \frac{\partial}{\partial \eta_{11}} f_{Y,\bar{\eta}_1}\left(\boldsymbol{Y}, \eta_{11} \mid \boldsymbol{Z}\right) = 0$$

for all η_{11}. Multiplying both sides of the above equation by η_{11} and then integrating with respect to $(\boldsymbol{Y}, \eta_{11})$ at $\boldsymbol{Z} = \boldsymbol{0}$, we obtain

$$\frac{\partial}{\partial \boldsymbol{\nu}} \mathrm{E}\left(\eta_{11} \mid \boldsymbol{Z} = \boldsymbol{0}\right)^{\mathrm{T}} \boldsymbol{h}_\nu + \mathrm{E}\left(\boldsymbol{Y}^{\mathrm{T}} \boldsymbol{h}_{\alpha 1} \mid \boldsymbol{Z} = \boldsymbol{0}\right) + \kappa_1 = 0.$$

It then follows from condition (C2) that $\kappa_1 = 0$. Thus, $h_1(\cdot) = 0$.

Assume that $h_{k-1}(\cdot)$ has been shown to be a zero function for some $k = 2, \ldots, K_1$. Let $\bar{\boldsymbol{\eta}}_k = (\eta_{11}, \ldots, \eta_{1k})$, and let $f_{Y,\bar{\eta}_k}$ be the density of $(\boldsymbol{Y}, \bar{\boldsymbol{\eta}}_k)$ given $\boldsymbol{Z}$. Setting $t_{k+1}, \ldots, t_K \to 0$ in (S.3), we have

$$\int \int \exp\left\{ - \sum_{l=1}^{k} \Lambda_l\big(t_l\big) s_l e^{\boldsymbol{W}^{\mathrm{T}} \boldsymbol{\vartheta}_l + \boldsymbol{Z}^{\mathrm{T}} \boldsymbol{\beta}_l + \boldsymbol{Y}^{\mathrm{T}} \boldsymbol{\alpha}_l + \eta_{1l}} \right\} g\left(\boldsymbol{s}\right) \left[\frac{\partial}{\partial \boldsymbol{\nu}} \left\{ f_{Y,\bar{\eta}_k}\left(\boldsymbol{Y}, \bar{\boldsymbol{\eta}}_k \mid \boldsymbol{Z}\right) \right\}^{\mathrm{T}} \boldsymbol{h}_\nu \right.$$
$$- \sum_{l=1}^{k-1} \Lambda_l\left(t_l\right) \left(\boldsymbol{Z}^{\mathrm{T}} \boldsymbol{h}_{\beta l} + \boldsymbol{Y}^{\mathrm{T}} \boldsymbol{h}_{\alpha l}\right) s_l e^{\boldsymbol{W}^{\mathrm{T}} \boldsymbol{\vartheta}_l + \boldsymbol{Z}^{\mathrm{T}} \boldsymbol{\beta}_l + \boldsymbol{Y}^{\mathrm{T}} \boldsymbol{\alpha}_l + \eta_{1l}} f_{Y,\bar{\eta}_k}\left(\boldsymbol{Y}, \bar{\boldsymbol{\eta}}_k \mid \boldsymbol{Z}\right)$$

$$-s_k e^{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\vartheta}_k+\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}_k+\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_k+\eta_{1k}} f_{Y,\bar{\eta}_k}\left(\boldsymbol{Y},\bar{\boldsymbol{\eta}}_k \mid \boldsymbol{Z}\right) \int_0^{t_k} \boldsymbol{Z}^{\mathrm{T}}\boldsymbol{h}_{\beta k}+\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{h}_{\alpha k}+h_k(\omega)\,\mathrm{d}\Lambda_k(\omega)\Bigg]\,\mathrm{d}\bar{\boldsymbol{\eta}}_k\,\mathrm{d}\boldsymbol{s}$$

equals zero for all $t_1,\ldots,t_k$, $\boldsymbol{W}$, $\boldsymbol{Z}$, and $\boldsymbol{Y}$. By the uniqueness of the Laplace transform, we conclude that $h_k(\omega)=\kappa_k\omega$ for some $\kappa_k\in\mathbb{R}$ and

$$\begin{aligned}\frac{\partial}{\partial\boldsymbol{\nu}} f_{Y,\bar{\eta}_k}\left(\boldsymbol{Y},\bar{\boldsymbol{\eta}}_k \mid \boldsymbol{Z}\right)^{\mathrm{T}}\boldsymbol{h}_\nu+\kappa_k\frac{\partial}{\partial\eta_{1k}} f_{Y,\bar{\eta}_k}\left(\boldsymbol{Y},\bar{\boldsymbol{\eta}}_k \mid \boldsymbol{Z}\right)\\ -\sum_{l=1}^{k}\left(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{h}_{\beta l}+\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{h}_{\alpha l}\right)\frac{\partial}{\partial\eta_{1l}} f_{Y,\bar{\eta}_k}\left(\boldsymbol{Y},\bar{\boldsymbol{\eta}}_k \mid \boldsymbol{Z}\right)=0.\end{aligned}$$

Multiplying both sides of the above equation by $(\eta_{11}\times\cdots\times\eta_{1k})$ and then integrating with respect to $(\boldsymbol{Y},\bar{\boldsymbol{\eta}}_k)$ at $\boldsymbol{Z}=\boldsymbol{0}$, we have $\kappa_k=0$ and $h_k(\cdot)=0$. By induction,

$$\frac{\partial}{\partial\boldsymbol{\nu}} f_{Y,\eta_1}\left(\boldsymbol{Y},\boldsymbol{\eta}_1 \mid \boldsymbol{Z}\right)^{\mathrm{T}}\boldsymbol{h}_\nu-\sum_{k=1}^{K_1}\left(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{h}_{\beta k}+\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{h}_{\alpha k}\right)\frac{\partial}{\partial\eta_{1k}} f_{Y,\eta_1}\left(\boldsymbol{Y},\boldsymbol{\eta}_1 \mid \boldsymbol{Z}\right)=0 \quad \forall\boldsymbol{Z},\boldsymbol{Y},\boldsymbol{\eta}_1.$$

It then follows from condition (D5) that $\boldsymbol{h}_\nu=\boldsymbol{0}$, $\boldsymbol{h}_{\beta k}=\boldsymbol{0}$, and $\boldsymbol{h}_{\alpha k}=\boldsymbol{0}$ for $k=1,\ldots,K_1$.

Consider the left-hand side of (S.3). The first term and the first K_1 terms in the summation of the second term have been shown to be zero. Thus, the left-hand side of (S.3) can be viewed as Laplace transforms with arguments $\Lambda_1(t_1),\ldots,\Lambda_{K_1}(t_{K_1})$. By the properties of the Laplace transform and function convolution,

$$\begin{aligned}-\sum_{k=K_1+1}^{K}\int\int\exp\left\{-\sum_{l=K_1+1}^{K}\Lambda_l(t_l)\,s_l e^{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\vartheta}_l+\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}_l+\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_l+\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\phi}_l}\right\}\\ \times\left\{\prod_{l=K_1+1}^{K} g_l(s_l)\right\} f_Y\left(\boldsymbol{Y} \mid \boldsymbol{Z},\boldsymbol{\eta}\right) f_\eta\left(\boldsymbol{\eta} \mid \boldsymbol{Z}\right) s_k e^{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\vartheta}_k+\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}_k+\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_k+\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\phi}_k}\,\mathrm{d}\left(s_{K_1+1},\ldots,s_K\right)\\ \times\left\{\int_0^{t_k}\boldsymbol{W}^{\mathrm{T}}\boldsymbol{h}_{\vartheta k}+\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{h}_{\beta k}+\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{h}_{\alpha k}+\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{h}_{\phi k}+h_k(\omega)\,\mathrm{d}\Lambda_k(\omega)\right\}\mathrm{d}\boldsymbol{\eta}_2=0\end{aligned}$$

for all $t_{K_1+1},\ldots,t_K$, $\boldsymbol{\eta}_1$, $\boldsymbol{W}$, $\boldsymbol{Z}$, and $\boldsymbol{Y}$. For $k=K_1+1,\ldots,K$, setting $t_l\to 0$ for $l\neq k$ in the above equation yields

$$\int\int e^{-\Lambda_k(t_k)s_k e^{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\vartheta}_k+\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}_k+\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_k+\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\phi}_k}} s_k e^{\boldsymbol{W}^{\mathrm{T}}\boldsymbol{\vartheta}_k+\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{\beta}_k+\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_k+\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\phi}_k} f_{\eta_2^{(k)}|Y,\eta_1}\left(\boldsymbol{\eta}_2^{(k)} \mid \boldsymbol{Z},\boldsymbol{Y},\boldsymbol{\eta}_1\right)$$
$$\times g_k\left(s_k\right)\left\{\int_0^{t_k} \boldsymbol{W}^{\mathrm{T}}\boldsymbol{h}_{\vartheta k}+\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{h}_{\beta k}+\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{h}_{\alpha k}+\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{h}_{\phi k}+h_k\left(\omega\right)\,\mathrm{d}\Lambda_k\left(\omega\right)\right\}\mathrm{d}\boldsymbol{\eta}_2^{(k)}\,\mathrm{d}s_k=0$$

for all t_k, $\boldsymbol{\eta}_1$, $\boldsymbol{W}$, $\boldsymbol{Z}$, and $\boldsymbol{Y}$, where $f_{\eta_2^{(k)}|Y,\eta_1}$ is the density of $\boldsymbol{\eta}_2^{(k)}$ given $(\boldsymbol{Z},\boldsymbol{Y},\boldsymbol{\eta}_1)$. By condition (C5), $\boldsymbol{\eta}_2^{(k)}$ is complete sufficient in $f_{\eta_2^{(k)}|Y,\eta_1}$ with $(\boldsymbol{Y}^{-(k)},\boldsymbol{\eta}^{-(k)})$ as parameters. Because $\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{\alpha}_k+\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\phi}_k$ and $\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{h}_{\alpha k}+\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{h}_{\phi k}$ do not depend on $(\boldsymbol{Y}^{-(k)},\boldsymbol{\eta}^{-(k)})$, the property of complete sufficient statistics implies that

$$\int_0^{t_k} \boldsymbol{W}^{\mathrm{T}}\boldsymbol{h}_{\vartheta k}+\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{h}_{\beta k}+\boldsymbol{Y}^{\mathrm{T}}\boldsymbol{h}_{\alpha k}+\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{h}_{\phi k}+h_k\left(\omega\right)\,\mathrm{d}\Lambda_k\left(\omega\right)=0 \quad \forall t_k\in[0,\tau]\,,\boldsymbol{\eta},\boldsymbol{W},\boldsymbol{Z},\boldsymbol{Y}.$$

Thus, $\boldsymbol{h}_{\vartheta k}=\mathbf{0}$, $\boldsymbol{h}_{\beta k}=\mathbf{0}$, $\boldsymbol{h}_{\alpha k}=\mathbf{0}$, and $\boldsymbol{h}_{\phi k}=\mathbf{0}$. Therefore, $h_k(\cdot)=0$.

We have shown that the information operator is one-to-one. Using the arguments of Zeng and Lin (2010), we can show that it is also a Fredholm operator. Therefore, the invertibility of the information operator follows. □

Proof of Lemma S1. Without loss of generality, assume that Y is a scalar. Clearly,

$$h\left(Y,\boldsymbol{b}\right)g\left(Y,\boldsymbol{b}\right) \;\leq\; \prod_{j=1}^{q}\exp\{K\left(1+|Y|\right)+\left(M_j+K\right)|b_j|+N_jYb_j-c_jb_j^2\}.$$

Note that $(M+K)|b_j|+NYb_j-cb_j^2$ is bounded by

$$-c\left(\left|b_j-\frac{NY}{2c}\right|-\frac{M+K}{2c}\right)^2+\frac{\left(M+K\right)^2}{4c}+\left(M+K\right)\left|\frac{NY}{2c}\right|+\frac{N^2Y^2}{4c}.$$

Similarly,

$$-M\,|b_j| + NYb_j - cb_j^2 \;\;\geq\;\; -c\left(\left|b_j - \frac{NY}{2c}\right| + \frac{M}{2c}\right)^2 - M\left|\frac{NY}{2c}\right| + \frac{N^2Y^2}{4c}.$$

Therefore, (S.2) is bounded by

$$e^{O(1+|Y|)}\frac{\int\prod_{j=1}^{q} e^{-c_j\left(\left|b_j-\frac{N_jY}{2c_j}\right|-\frac{M_j+K}{2c_j}\right)^2}\,\mathrm{d}\boldsymbol{b}}{\int\prod_{j=1}^{q} e^{-c_j\left(\left|b_j-\frac{N_jY}{2c_j}\right|+\frac{M_j}{2c_j}\right)^2}\,\mathrm{d}\boldsymbol{b}} = e^{O(1+|Y|)}\frac{\int\prod_{j=1}^{q} e^{-c_j\left(|b_j|-\frac{M_j+K}{2c_j}\right)^2}\,\mathrm{d}\boldsymbol{b}}{\int\prod_{j=1}^{q} e^{-c_j\left(|b_j|+\frac{M_j}{2c_j}\right)^2}\,\mathrm{d}\boldsymbol{b}} \leq e^{O(1+|Y|)}.$$

Likewise, (S.1) is bounded by

$$e^{O(1+|Y|)}\frac{\int \exp\left\{-e^{A_1+\left(B+\sum_j N_j/2c_j\right)Y+\mathbf{1}_q^{\mathrm{T}}\boldsymbol{b}}\right\}\prod_{j=1}^{q} e^{-c_j\left(|b_j|+\frac{M_j+K}{2c_j}\right)^2}\,\mathrm{d}\boldsymbol{b}}{\int \exp\left\{-e^{A_2+\left(B+\sum_j N_j/2c_j\right)Y+\mathbf{1}_q^{\mathrm{T}}\boldsymbol{b}}\right\}\prod_{j=1}^{q} e^{-c_j\left(|b_j|+\frac{M_j}{2c_j}\right)^2}\,\mathrm{d}\boldsymbol{b}}. \tag{S.5}$$

For any $w > 0$ and $a \in \mathbb{R}$,

$$\int_{b\in\mathbb{R}} \exp\left(-we^{b}\right) e^{-(|b|-a)^2}\,\mathrm{d}b \;\;\leq\;\; 2\int_{b\in\mathbb{R}} \exp\left(-we^{-a}e^{-b}\right) e^{-b^2}\,\mathrm{d}b.$$

Thus, the numerator in (S.5) is bounded above by

$$2^q\int \exp\left\{-e^{A_1+\sum_j -(M_j+K)/2c_j+(B+N_j/2c_j)Y-\mathbf{1}_q^{\mathrm{T}}\boldsymbol{b}}\right\}\prod_{j=1}^{q} e^{-c_jb_j^2}\,\mathrm{d}\boldsymbol{b}.$$

In addition, if $a > 0$, then

$$\int_{b\in\mathbb{R}} \exp\left(-we^{b}\right) e^{-(|b|+a)^2}\,\mathrm{d}b \;\;\geq\;\; \frac{1}{2}K_a^{-1}\int_{b\in\mathbb{R}} \exp\left(-we^{a}e^{-b}\right) e^{-b^2}\,\mathrm{d}b,$$

where $K_a = \int_0^\infty e^{-b^2}\,\mathrm{d}b/\int_a^\infty e^{-b^2}\,\mathrm{d}b$. Thus, the denominator in (S.5) is bounded below by

$$\frac{1}{2^q}\prod_{j=1}^{q} K_{M_j/2c_j}^{-1}\int \exp\left\{-e^{A_2+\sum_j M_j/2c_j+(B+N_j/2c_j)Y-\mathbf{1}_q^{\mathrm{T}}\boldsymbol{b}}\right\}\prod_{j=1}^{q} e^{-c_j b_j^2}\,\mathrm{d}\boldsymbol{b}.$$

The fraction in (S.5) is bounded by

$$4^q\prod_{j=1}^{q} K_{M_j/2c_j}\frac{\int \exp\left\{-w_1 e^{-\left(\boldsymbol{c}^{-1/2}\right)^{\mathrm{T}}\boldsymbol{b}}\right\} e^{-|\boldsymbol{b}|^2}\,\mathrm{d}\boldsymbol{b}}{\int \exp\left\{-w_2 e^{-\left(\boldsymbol{c}^{-1/2}\right)^{\mathrm{T}}\boldsymbol{b}}\right\} e^{-|\boldsymbol{b}|^2}\,\mathrm{d}\boldsymbol{b}},$$

where $\boldsymbol{c}^{-1/2}=(c_1^{-1/2},\ldots,c_q^{-1/2})^{\mathrm{T}}$, and $w_k=e^{A_k+\sum_j(-1)^k(M_j+K)/2c_j+(B+N_j/2c_j)Y}$ for $k=1,2$. Therefore, the fraction in (S.5) is finite if $\sum_j(B+N_j/2c_j)Y\to-\infty$. If $\sum_j(B+N_j/2c_j)Y\to\infty$, then we use the approximation that

$$\begin{aligned}
&\int_{b\in\mathbb{R}}\exp\left(-\frac{b^2}{2}-we^{\mu b}\right)\mathrm{d}b\\
=\;&\left[\frac{2\pi\{1+o(1)\}}{\log w}\right]^{1/2}\exp\Bigg(-\frac{1}{2\mu^2}\left[\log w\left\{1-\frac{\log\log w-\log\mu^2}{\log w}+o\left(\frac{1}{\log w}\right)\right\}\right]^2\\
&-\frac{\log w}{\mu^2}\left\{1-\frac{\log\log w-\log\mu^2}{\log w}+o\left(\frac{1}{\log w}\right)\right\}\Bigg)
\end{aligned}$$

as $w\to\infty$ (Evans and Swartz 2000). It follows that

$$\begin{aligned}
&\frac{\int \exp\left\{-w_1 e^{-\left(\boldsymbol{c}^{-1/2}\right)^{\mathrm{T}}\boldsymbol{b}}\right\} e^{-|\boldsymbol{b}|^2}\,\mathrm{d}\boldsymbol{b}}{\int \exp\left\{-w_2 e^{-\left(\boldsymbol{c}^{-1/2}\right)^{\mathrm{T}}\boldsymbol{b}}\right\} e^{-|\boldsymbol{b}|^2}\,\mathrm{d}\boldsymbol{b}}\\
=\;&\frac{\int \exp\left(-w_1 e^{-\left|\boldsymbol{c}^{-1/2}\right|b}\right) e^{-b^2}\,\mathrm{d}b}{\int \exp\left(-w_2 e^{-\left|\boldsymbol{c}^{-1/2}\right|b}\right) e^{-b^2}\,\mathrm{d}b}\\
=\;&O(1)\left(\frac{\log w_2}{\log w_1}\right)^{1/2}\exp\Bigg[-\frac{1}{2\left|\boldsymbol{c}^{-1/2}\right|^2}\left\{\log w_1-\log\log w_1+2\log\left|\boldsymbol{c}^{-1/2}\right|+o(1)\right\}^2\\
&-\frac{1}{\left|\boldsymbol{c}^{-1}\right|^2}\left\{\log w_1-\log\log w_1+2\log\left|\boldsymbol{c}^{-1/2}\right|+o(1)\right\}
\end{aligned}$$

$$+\frac{1}{2\left|\boldsymbol{c}^{-1/2}\right|^{2}}\left\{\log w_2-\log\log w_2+2\log\left|\boldsymbol{c}^{-1/2}\right|+o\left(1\right)\right\}^{2}$$
$$\left.+\frac{1}{\left|\boldsymbol{c}^{-1}\right|^{2}}\left\{\log w_2-\log\log w_2+2\log\left|\boldsymbol{c}^{-1/2}\right|+o\left(1\right)\right\}\right].$$

The Y^2 terms in the exponent cancel out; therefore, (S.1) is bounded by $e^{O(1+|Y|)}$. □

Proof of Lemma S2. We use $\boldsymbol{Z}$ to denote both $\boldsymbol{W}$ and $\boldsymbol{Z}$ with $\boldsymbol{\beta}_k$ ($k=1,\ldots,K$) as the corresponding vector of regression parameters. Note that

$$\begin{aligned}
&\frac{\partial}{\partial\boldsymbol{\beta}}\log\Psi\big(\mathcal{O}_i;\boldsymbol{\theta},\mathcal{A}\big) \qquad\qquad \text{(S.6)}\\
=&\int\Omega_{ki}\left(\boldsymbol{\eta}\right)\Bigg(\Delta_{ki}\bigg[1+\frac{G_k''\left\{q_{ki}\left(\boldsymbol{\theta},\mathcal{A}\right)\right\}}{G_k'\left\{q_{ki}\left(\boldsymbol{\theta},\mathcal{A}\right)\right\}}\Lambda_k\big(\tilde{T}_{ki}\big)e^{\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\beta}_k+\boldsymbol{Y}_i^{\mathrm{T}}\boldsymbol{\alpha}_k+\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\phi}_k}\bigg]\\
&-G_k'\left\{q_{ki}\left(\boldsymbol{\theta},\mathcal{A}\right)\right\}\Lambda_k\big(\tilde{T}_{ki}\big)e^{\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\beta}_k+\boldsymbol{Y}_i^{\mathrm{T}}\boldsymbol{\alpha}_k+\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\phi}_k}\Bigg)\boldsymbol{Z}_i f_Y\big(\boldsymbol{Y}_i\mid\boldsymbol{Z}_i,\boldsymbol{\eta};\boldsymbol{\psi}\big)f_\eta\big(\boldsymbol{\eta}\mid\boldsymbol{Z}_i;\boldsymbol{\nu}\big)\,\mathrm{d}\boldsymbol{\eta}\\
&\times\left\{\int\Omega_{ki}\left(\boldsymbol{\eta}\right)f_Y\left(\boldsymbol{Y}_i\mid\boldsymbol{Z}_i,\boldsymbol{\eta};\boldsymbol{\psi}\right)f_\eta\left(\boldsymbol{\eta}\mid\boldsymbol{Z}_i;\boldsymbol{\nu}\right)\,\mathrm{d}\boldsymbol{\eta}\right\}^{-1},
\end{aligned}$$

where $q_{ki}\left(\boldsymbol{\theta},\mathcal{A}\right)=e^{\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\beta}_k+\boldsymbol{Y}_i^{\mathrm{T}}\boldsymbol{\alpha}_k+\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\phi}_k}\Lambda_k(\tilde{T}_{ki})$, and

$$\Omega_{ki}\left(\boldsymbol{\eta}\right)=\Big[e^{\boldsymbol{Z}_i^{\mathrm{T}}\boldsymbol{\beta}_k+\boldsymbol{Y}_i^{\mathrm{T}}\boldsymbol{\alpha}_k+\boldsymbol{\eta}^{\mathrm{T}}\boldsymbol{\phi}_k}G_k'\left\{q_{ki}\left(\boldsymbol{\theta},\mathcal{A}\right)\right\}\Big]^{\Delta_{ki}}\exp\left[-G_k\left\{q_{ki}\left(\boldsymbol{\theta},\mathcal{A}\right)\right\}\right].$$

By condition (D4), the terms $G_k'\big\{q_{ki}(\boldsymbol{\theta},\mathcal{A})\big\}$ and $\big[G_k'\big\{q_{ki}(\boldsymbol{\theta},\mathcal{A})\big\}\big]^{-1}G_k''\big\{q_{ki}(\boldsymbol{\theta},\mathcal{A})\big\}$ are bounded by $e^{O(1+|\boldsymbol{Y}_i|+|\boldsymbol{\eta}|)}$. Therefore, the first integral term on the right-hand side of (S.6) is bounded above by

$$\int\Omega_{ki}\left(\boldsymbol{\eta}\right)e^{O(1+|\boldsymbol{Y}_i|+|\boldsymbol{\eta}|)}f_Y\big(\boldsymbol{Y}_i\mid\boldsymbol{Z}_i,\boldsymbol{\eta};\boldsymbol{\psi}\big)f_\eta\big(\boldsymbol{\eta}\mid\boldsymbol{Z}_i;\boldsymbol{\nu}\big)\,\mathrm{d}\boldsymbol{\eta}.$$

Condition (D3) states that either $G_k(x)/\log x\to M_k$ or $G_k(x)/x^{\rho_k}\to M_k$ as $x\to\infty$. If $G_k(x)/x^{\rho_k}\to M_k$, then we can find $M_{1k}>0$, $M_{2k}>0$, $C_{1k}\in\mathbb{R}$, and $C_{2k}\in\mathbb{R}$ such that

$M_{2k}x^{\rho_k} + C_{2k} \le G_k(x) \le M_{1k}x^{\rho_k} + C_{1k}$. Thus,

$$\exp\left(-e^{A_1+\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y}_i+\boldsymbol{C}^{\mathrm{T}}\boldsymbol{\eta}}\right) e^{-O(1+|\boldsymbol{Y}_i|+|\boldsymbol{\eta}|)} \le \Omega_{ki}\left(\boldsymbol{\eta}\right) \le \exp\left(-e^{A_2+\boldsymbol{B}^{\mathrm{T}}\boldsymbol{Y}_i+\boldsymbol{C}^{\mathrm{T}}\boldsymbol{\eta}}\right) e^{O(1+|\boldsymbol{Y}_i|+|\boldsymbol{\eta}|)}$$

for some A_1, A_2, $\boldsymbol{B}$, and $\boldsymbol{C}$. After transforming $\boldsymbol{\eta}$ to $\boldsymbol{b}$ using the transformation S specified in condition (D3), we see that (S.6) is bounded by a term of the form (S.1) and is in turn bounded by $e^{O(1+|\boldsymbol{Y}_i|)}$ according to Lemma S1. If $G_k\big(x\big)/\log x \to M_k$, then (S.6) is bounded by a term of the form (S.2), which is also bounded by $e^{O(1+|\boldsymbol{Y}_i|)}$. Similarly, the derivatives of $\log\Psi(\mathcal{O}_i;\boldsymbol{\theta},\mathcal{A})$ with respect to other parameters are bounded by $e^{O(1+|\boldsymbol{Y}_i|)}$.

By Theorem 2.7.5 of van der Vaart and Wellner (1996), the bracket covering number for any bounded set in $BV\left[0,\tau\right]$ is of the order $\exp\left\{O\left(1/\epsilon\right)\right\}$. Therefore, we can construct $N_\epsilon \equiv 1/\epsilon^d \times \exp\{O(K/\epsilon)\}$ brackets for $\Theta \times BV\left[0,\tau\right]^K$, denoted by $\{(\boldsymbol{\theta}_j^L, \mathcal{A}_j^L), (\boldsymbol{\theta}_j^U, \mathcal{A}_j^U)\}$ $(j = 1,\ldots,N_\epsilon)$, where $|\boldsymbol{\theta}_j^U - \boldsymbol{\theta}_j^L| < \epsilon$, and

$$\int_0^\tau \left|\Lambda_{kj}^U(t) - \Lambda_{kj}^L(t)\right|^2 \mathrm{E}\left\{e^{O(1+|\boldsymbol{Y}_i|)}\,\mathrm{d}I(\tilde{T}_{ki} \le t)\right\} < \epsilon^2.$$

By the mean-value theorem, for any $(\boldsymbol{\theta}_1, \mathcal{A}_1)$ and $(\boldsymbol{\theta}_2, \mathcal{A}_2)$,

$$\left|\log\Psi\big(\mathcal{O}_i;\boldsymbol{\theta}_1,\mathcal{A}_1\big) - \log\Psi\big(\mathcal{O}_i;\boldsymbol{\theta}_2,\mathcal{A}_2\big)\right| \le e^{O(1+|\boldsymbol{Y}_i|)}\left\{|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2| + \sum_{k=1}^K \left|\Lambda_{1k}(\tilde{T}_{ki}) - \Lambda_{2k}(\tilde{T}_{ki})\right|\right\}.$$

The pairs of functions

$$\log\Psi\big(\mathcal{O}_i;\boldsymbol{\theta}_j^L,\mathcal{A}_j^L\big) \pm e^{O(1+|\boldsymbol{Y}_i|)}\left\{\left|\boldsymbol{\theta}_j^U - \boldsymbol{\theta}_j^L\right| + \sum_{k=1}^K \left|\Lambda_{kj}^U(\tilde{T}_{ki}) - \Lambda_{kj}^L(\tilde{T}_{ki})\right|\right\},$$

$j = 1,\ldots,N_\epsilon$, constitute a bracket cover for $\mathcal{C}_1$, where the $L_2(\mathcal{P})$-distance within each bracket pair is of the order ϵ. Therefore, the bracket entropy of $\mathcal{C}_1$ is finite, such that $\mathcal{C}_1$ is Donsker. Similarly, the classes $\mathcal{C}_2$ and $\mathcal{C}_{3k}$ can also be shown to be Donsker. □

REFERENCES

Evans, M., and Swartz, T. (2000), *Approximating Integrals via Monte Carlo and Deterministic Methods*, New York: Oxford University Press.

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

Zeng, D., and Lin, D. Y. (2010), "A General Asymptotic Theory for Maximum Likelihood Estimation in Semiparametric Regression Models With Censored Data," *Statistica Sinica*, 20, 871–910.

Table S.1. Simulation Results for Mplus

Dep	Ind	Param	Bias	SE	SEE	CP
T	Z_1	0.100	0.004	0.075	0.073	0.95
	Z_2	−0.200	−0.004	0.18	0.182	0.95
	Y_6	0.100	0.001	0.145	0.143	0.95
	Y_7	0.200	−0.007	0.167	0.163	0.96
	η_2	0.500	0.035	0.171	0.170	0.97
	$\lambda_0(t_1)$	0.900	−0.004	1.512	0.915	0.64
	$\lambda_0(t_2)$	1.440	−0.027	2.634	1.443	0.67
	$\lambda_0(t_3)$	2.100	0.105	3.229	2.254	0.73
Y_1	Int	0.000	0.000	0.074	0.073	0.94
	Var	1.000	−0.015	0.132	0.128	0.96
Y_2	Int	0.000	0.002	0.076	0.073	0.94
	η_1	1.000	0.022	0.221	0.204	0.94
	Var	1.000	−0.013	0.13	0.129	0.96
Y_3	Int	0.000	0.001	0.076	0.073	0.95
	η_1	1.000	0.021	0.215	0.203	0.95
	Var	1.000	−0.010	0.132	0.128	0.96
Y_4	Int	0.000	−0.001	0.077	0.075	0.95
	Var	1.000	0.000	0.185	0.178	0.94
Y_5	Int	0.000	−0.002	0.078	0.076	0.94
	η_2	1.000	0.099	0.324	0.298	0.95
	Var	1.000	−0.054	0.205	0.195	0.97
Y_6	Int	0.000	0.005	0.286	0.273	0.94
	Z_1	−0.500	0.001	0.112	0.113	0.95
	Z_2	0.500	0.021	0.307	0.297	0.95
	η_2	0.000	0.008	0.233	0.226	0.96
Y_7	Int	1.000	−0.032	0.338	0.345	0.96
	Z_1	1.000	0.028	0.148	0.148	0.96
	Z_2	0.200	0.002	0.333	0.342	0.96
	Y_6	−0.200	−0.011	0.264	0.264	0.95
	η_2	0.000	0.006	0.269	0.260	0.97
η_1	Var	0.500	0.009	0.138	0.134	0.95
η_2	η_1	0.500	−0.004	0.156	0.146	0.93
	Var	0.500	−0.020	0.164	0.159	0.96

NOTE: Each row corresponds to the regression parameter of the dependent variable "Dep" on the independent variable "Ind" or some other parameter in the model of "Dep". "Int" and "Var" stand for the intercept and error variance, respectively. The parameters $\lambda_0(t_1)$, $\lambda_0(t_2)$, and $\lambda_0(t_3)$ correspond to the baseline hazard function values at the 25%, 50%, and 75% quantiles of the survival time. The true value of a parameter is given under "Param". "Bias" is the empirical bias; "SE" is the empirical standard error; "SEE" is the empirical mean of the standard error estimator; and "CP" is the empirical coverage probability of the 95% confidence interval.

Table S.2. Analysis Results for the Gene ACACA

		NPMLE ($r=1$)		NPMLE ($r=0$)		Mplus ($r=0$)	
Dep	Ind	Est	St Error	Est	St Error	Est	St Error
T	Z_1	0.229	0.086	0.114	0.054	0.114	0.054
	Z_2	0.038	0.309	0.168	0.192	0.168	0.192
	Y_6	0.760	0.284	0.436	0.139	0.436	0.139
	Y_7	0.511	1.118	0.263	0.151	0.263	0.148
	η_2	0.192	0.116	0.068	0.062	0.068	0.062
	$\Lambda_0(t_1)$	0.135	0.143	0.160	0.028	0.160	N/A
	$\Lambda_0(t_2)$	0.401	0.415	0.392	0.061	0.394	N/A
	$\Lambda_0(t_3)$	1.062	1.061	0.767	0.114	0.771	N/A
Y_1	Int	0.004	0.044	0.004	0.044	0.005	0.044
	Var	0.286	0.038	0.288	0.038	0.287	0.038
Y_2	Int	−0.013	0.044	−0.013	0.044	−0.013	0.044
	η_1	0.687	0.056	0.688	0.056	0.687	0.056
	Var	0.658	0.046	0.658	0.046	0.658	0.046
Y_3	Int	0.014	0.043	0.014	0.044	0.014	0.043
	η_1	1.019	0.062	1.022	0.062	1.021	0.062
	Var	0.276	0.039	0.274	0.039	0.275	0.039
Y_4	Int	−0.021	0.048	−0.023	0.049	−0.022	0.048
	Var	0.096	0.046	0.097	0.046	0.106	0.045
Y_5	Int	−0.009	0.048	−0.011	0.049	−0.011	0.048
	η_2	0.909	0.054	0.910	0.054	0.919	0.054
	Var	0.237	0.041	0.236	0.041	0.229	0.041
Y_6	Int	−1.733	0.128	−1.734	0.128	−1.732	0.128
	Z_1	−0.290	0.125	−0.290	0.125	−0.289	0.125
	Z_2	−0.095	0.434	−0.092	0.434	−0.094	0.434
	η_2	−0.113	0.157	−0.121	0.157	−0.123	0.158
Y_7	Int	1.977	0.149	1.977	0.149	1.996	0.151
	Z_1	0.132	0.135	0.132	0.135	0.133	0.135
	Z_2	0.001	0.461	0.002	0.461	−0.015	0.462
	Y_6	−0.055	0.356	−0.054	0.356	−0.072	0.357
	η_2	0.082	0.162	0.080	0.161	0.081	0.162
η_1	Var	0.689	0.068	0.687	0.068	0.687	0.068
η_2	η_1	0.691	0.062	0.692	0.063	0.690	0.063
	Var	0.560	0.063	0.559	0.063	0.551	0.062

NOTE: Each row corresponds to the regression parameter of the dependent variable "Dep" on the independent variable "Ind" or some other parameter in the model of "Dep". "Int" and "Var" stand for the intercept and error variance, respectively. The representation of each variable is given in Section 6. The parameters $\Lambda_0(t_1)$, $\Lambda_0(t_2)$, and $\Lambda_0(t_3)$ correspond to the cumulative baseline hazard function values at the 25%, 50%, and 75% quantiles of the progression-free survival, respectively. The point estimate of and standard error estimate of a parameter are given under "Est" and "St Error", respectively.