Efficient Estimation of the Accelerated Failure Time Model by Incorporating Auxiliary Aggregate Information

Huijuan Ma[1], Manli Cheng[1], Yukun Liu[1], Donglin Zeng[2], and Yong Zhou[1]

[1]*East China Normal University and* [2]*University of Michigan*

Supplementary Material

This supplementary material consists of seven sections. In Section S1, we introduce notations that are used throughout this supplementary material. Section S2 contains our proof of Theorem 1. Section S3 studies the large-sample properties of $\hat{\beta}_{\text{au}}$. In Section S4, we investigate the relationship between $(\hat{\beta}_{\text{au}}, \hat{\rho}_{\text{au}})$ and $(\tilde{\beta}, \tilde{\rho})$. Section S5 contains our proof of Theorem 2. Additional simulation results are given in Section S6. Section S7 studies the heterogeneity in covariate distribution and uncertainty in the auxiliary information.

S1 Notations

We give the specific forms of the ordinary and functional derivatives of $l(\beta, g; O)$ and $\Psi(\beta, \rho, g; Z)$. For any fixed function $g \in \mathcal{G}$, let $\{g_\eta : \eta \text{ in a neighborhood of } 0 \in \mathbb{R}\}$ be a smooth curve in $\mathcal{G}$ running through g at $\eta = 0$, that is, $g_\eta|_{\eta=0} = g$. Define

$$\mathcal{H} = \{h : h = \frac{\partial g_\eta}{\partial \eta}\Big|_{\eta=0}, g_\eta \in \mathcal{G}\},$$

which is general enough to include $\mathcal{G}$, for example, when $g_\eta = (1+\eta)g$, $h = \partial g_\eta/\partial\eta|_{\eta=0} = g$. Define $N_i(t) = \delta_i I(\log(X_i) \le t)$. The first and second ordinary and functional derivatives of $l(\beta, g; O)$ are

$$
\begin{aligned}
\dot{l}_\beta(\beta, g; O) &= \frac{\partial}{\partial \beta} l(\beta, g; O) \\
&= -Z\Big\{ \int \dot{g}(t - Z^\top \beta) dN(t) - \int I(\log(X) \ge t) \exp\{g(t - Z^\top\beta)\} \dot{g}(t - Z^\top \beta) dt \Big\} \\
&= -Z\Big\{ \int \dot{g}(t) dN(t, \beta) - \int Y(t, \beta) \exp\{g(t)\} \dot{g}(t) dt \Big\}, \\
\dot{l}_g(\beta, g; O)[h] &= \lim_{\eta \to 0} \eta^{-1}\{l(\beta, g + \eta h; O) - l(\beta, g; O)\} \\
&= \int h(t - Z^\top\beta) dN(t) - \int I(\log(X) \ge t) \exp\{g(t - Z^\top\beta)\} h(t - Z^\top\beta) dt \\
&= \int h(t) dN(t, \beta) - \int Y(t, \beta) \exp\{g(t)\} h(t) dt, \\
\ddot{l}_{\beta\beta}(\beta, g; O) &= \frac{\partial^2}{\partial\beta\partial\beta^\top} l(\beta, g; O) = ZZ^\top \Big\{ \int \ddot{g}(t - Z^\top\beta) dN(t) \\
&\quad - \int I(\log(X) \ge t) \exp\{g(t - Z^\top\beta)\}\{\ddot{g}(t - Z^\top\beta) + \dot{g}^2(t - Z^\top\beta)\} dt \Big\} \\
&= ZZ^\top \Big\{ \int \ddot{g}(t) dN(t, \beta) - \int Y(t, \beta) \exp\{g(t)\}\{\ddot{g}(t) + \dot{g}^2(t)\} dt \Big\}, \\
\ddot{l}_{\beta g}(\beta, g; O)[h] &= \lim_{\eta \to 0} \eta^{-1}\{\dot{l}_\beta(\beta, g + \eta h; O) - \dot{l}_\beta(\beta, g; O)\} \\
&= -Z\Big\{ \int \dot{h}(t - Z^\top\beta) dN(t) - \int I(\log(X) \ge t) \exp\{g(t - Z^\top\beta)\} \\
&\qquad \{h(t - Z^\top\beta)\dot{g}(t - Z^\top\beta) + \dot{h}(t - Z^\top\beta)\} dt \Big\} \\
&= -Z\Big\{ \int \dot{h}(t) dN(t, \beta) - \int Y(t, \beta) \exp\{g(t)\}\{h(t)\dot{g}(t) + \dot{h}(t)\} dt \Big\}, \\
\ddot{l}_{g\beta}(\beta, g; O)[h] &= \frac{\partial}{\partial\beta} \dot{l}_g(\beta, g; O)[h] = \{\ddot{l}_{\beta g}(\beta, g; O)[h]\}^\top, \\
\ddot{l}_{gg}(\beta, g; O)[h_1, h_2] &= \lim_{\eta \to 0} \eta^{-1}\{\dot{l}_g(\beta, g + \eta h_2; O)[h_1] - \dot{l}_g(\beta, g; O)[h_1]\} \\
&= -\int I(\log(X) \ge t) \exp\{g(t - Z^\top\beta)\} h_1(t - Z^\top\beta) h_2(t - Z^\top\beta) dt
\end{aligned}
$$

$$= -\int Y(t,\beta)\exp\{g(t)\}h_1(t)h_2(t)dt,$$

where $h, h_1, h_2 \in \mathcal{H}$, $\dot{g}(\cdot)$ and $\ddot{g}(\cdot)$ are the first and second derivatives of $g(\cdot)$, respectively, $\dot{h}(\cdot)$ is the first derivative of $h(\cdot)$. The ordinary and functional derivatives of $\Psi(\beta,\rho,g;Z)$ are

$$\dot{\Psi}_\beta(\beta,\rho,g;Z) = (\dot{\Psi}_{1,\beta}(\beta,\rho,g;Z),\ldots,\dot{\Psi}_{J,\beta}(\beta,\rho,g;Z)),$$
$$\dot{\Psi}_\rho(\beta,\rho,g;Z) = (\dot{\Psi}_{1,\rho}(\beta,\rho,g;Z),\ldots,\dot{\Psi}_{J,\rho}(\beta,\rho,g;Z))^\top,$$
$$\dot{\Psi}_g(\beta,\rho,g;Z)[h] = (\dot{\Psi}_{1,g}(\beta,\rho,g;Z)[h],\ldots,\dot{\Psi}_{J,g}(\beta,\rho,g;Z)[h])^\top,$$

where

$$\dot{\Psi}_{j,\beta}(\beta,\rho,g;Z) = \frac{\partial}{\partial\beta}\dot{\Psi}_j(\beta,\rho,g;Z), \qquad \dot{\Psi}_{j,\rho}(\beta,\rho,g;Z) = \frac{\partial}{\partial\rho}\dot{\Psi}_j(\beta,\rho,g;Z),$$
$$\dot{\Psi}_{j,g}(\beta,\rho,g;Z)[h] = \lim_{\eta\to 0}\eta^{-1}\{\Psi_j(\beta,\rho,g+\eta h;Z) - \Psi_j(\beta,\rho,g;Z)\}, \quad j = 1,\ldots,J.$$

When

$$\Psi_j(\beta,\rho,g;Z) = I(Z\in\Omega_j)\Big[\exp\Big\{-\int I(s\le \log t_j^* - Z^\top\beta)\exp\{g(s)\}ds\Big\} - \zeta_j^{1/\rho}\Big],$$

we have

$$\dot{\Psi}_{j,\beta}(\beta,\rho,g;Z) = I(Z\in\Omega_j)ZS(\log t_j^* - Z^\top\beta)\exp\{g(\log t_j^* - Z^\top\beta)\},$$
$$\dot{\Psi}_{j,\rho}(\beta,\rho,g;Z) = I(Z\in\Omega_j)\rho^{-2}\zeta_j^{1/\rho}\ln(\zeta_j),$$

$$\dot{\Psi}_{j,g}(\beta,\rho,g;Z)[h] = -I(Z\in\Omega_j)S(\log t_j^* - Z^\top\beta)\int I(s\le \log t_j^* - Z^\top\beta)e^{g(s)}h(s)ds,$$

where $S(\cdot) = 1 - F(\cdot), j = 1,\ldots,J$. For a r-dimensional vector of functions $\tilde{h} = (\tilde{h}_1,\ldots,\tilde{h}_r)^\top \in \mathcal{H}^r$ and a m-dimensional vector of functions $\bar{h} = (\bar{h}_1,\ldots,\bar{h}_m)^\top \in \mathcal{H}^m$, denote

$$\begin{aligned}
\dot{l}_g(\beta,g;O)[\tilde{h}] =& (\dot{l}_g(\beta,g;O)[\tilde{h}_1],\ldots,\dot{l}_g(\beta,g;O)[\tilde{h}_r])^\top,\\
\ddot{l}_{g\beta}(\beta,g;O)[\tilde{h}] =& (\ddot{l}_{g\beta}(\beta,g;O)[\tilde{h}_1]^\top,\ldots,\ddot{l}_{g\beta}(\beta,g;O)[\tilde{h}_r]^\top),\\
\ddot{l}_{gg}(\beta,g;O)[\tilde{h},h] =& (\ddot{l}_{gg}(\beta,g;O)[\tilde{h}_1,h],\ldots,\ddot{l}_{gg}(\beta,g;O)[\tilde{h}_r,h])^\top,\\
\ddot{l}_{gg}(\beta,g;O)[\tilde{h},\bar{h}] =& (\ddot{l}_{gg}(\beta,g;O)[\tilde{h},\bar{h}_1],\ldots,\ddot{l}_{gg}(\beta,g;O)[\tilde{h},\bar{h}_m]),\\
\Psi_g(\beta,\rho,g;Z)[\tilde{h}] =& (\Psi_g(\beta,\rho,g;Z)[\tilde{h}_1],\ldots,\Psi_g(\beta,\rho,g;Z)[\tilde{h}_r]).
\end{aligned}$$

S2 Proof of Theorem 1

S2.1 Lemmas

Before proving Theorem 1, we give two lemmas. For any function $f(\cdot)$, let $Pf = \int f(\cdot)dP(\cdot)$ and $\mathbb{P}_n(f) = n^{-1}\sum_{i=1}^n f(O_i)$, where P is the probability measure of a generic observation O and $\mathbb{P}_n$ is the empirical measure of $\{O_1,\ldots,O_n\}$.

Lemma 1. *For any $\delta_n \to 0$, define $Q_n = \{(\beta,\rho,g)\in\mathcal{B}\times\mathcal{R}\times\mathcal{F} : \|\beta-\beta_0\|\le \delta_n, |\rho-\rho_0|\le\delta_n, \|g^*-g_0^*\|_{\mathcal{G}}\le\delta_n$ with $g^* = g, \dot{g}$, and $\ddot{g}\}$, where $\|\cdot\|_{\mathcal{G}}$ denotes the metric on $\mathcal{G}$. For $h\in L_2(P)$, which is the space of square-integrable functions with* *respect to P, we have*

$$\sup_{Q_n} \|\sqrt{n}(\mathbb{P}_n - P)\{\dot{l}_\beta(\beta, g; O) - \dot{l}_\beta(\beta_0, g_0; O)\}\| = o_p(1),$$

$$\sup_{Q_n} \|\sqrt{n}(\mathbb{P}_n - P)\{\dot{l}_g(\beta, g; O)[h] - \dot{l}_g(\beta_0, g_0; O)[h]\}\| = o_p(1),$$

$$\sup_{Q_n} \|\sqrt{n}(\mathbb{P}_n - P)\{\Psi(\beta, \rho, g; Z) - \Psi(\beta_0, \rho_0, g_0; Z)\}\| = o_p(1),$$

$$\sup_{Q_n} \|\sqrt{n}(\mathbb{P}_n - P)\{\dot{\Psi}_\beta(\beta, \rho, g; Z) - \dot{\Psi}_\beta(\beta_0, \rho_0, g_0; Z)\}\| = o_p(1),$$

$$\sup_{Q_n} \|\sqrt{n}(\mathbb{P}_n - P)\{\dot{\Psi}_\rho(\beta, \rho, g; Z) - \dot{\Psi}_\rho(\beta_0, \rho_0, g_0; Z)\}\| = o_p(1),$$

$$\sup_{Q_n} \|\sqrt{n}(\mathbb{P}_n - P)\{\dot{\Psi}_g(\beta, \rho, g; Z)[h] - \dot{\Psi}_g(\beta_0, \rho_0, g_0; Z)[h]\}\| = o_p(1).$$

Proof of Lemma 1: Note that $\exp(x)$ is a Lipschitz continuous function for x in a bounded set, $1/x^2$ is a Lipschitz continuous function for x in a bounded set away from 0, the functional spaces of indicator functions and bounded variation functions are Donsker classes (van der Vaart and Wellner, 1996). We assume that the true parameter of ρ, ρ_0, is bounded away from 0. According to the regularity conditions C3(a) and C7, Q_n is uniformly bounded. By the permanence property of Donsker class, the function classes $\{\dot{l}_\beta(\beta, g; O), (\beta, \rho, g) \in Q_n\}$, $\{\dot{l}_g(\beta, g; O)[h], (\beta, \rho, g) \in Q_n, h \in L_2(P)\}$, $\{\Psi(\beta, \rho, g; Z), (\beta, \rho, g) \in Q_n\}$, $\{\dot{\Psi}_\beta(\beta, \rho, g; Z), (\beta, \rho, g) \in Q_n\}$, $\{\dot{\Psi}_\rho(\beta, \rho, g; Z), (\beta, \rho, g) \in Q_n\}$, and $\{\dot{\Psi}_g(\beta, \rho, g; Z)[h], (\beta, \rho, g) \in Q_n, h \in L_2(P)\}$ are also Donsker. The desired results follow from the equi-continuous property of the Donsker class.

Lemma 2. *For any $(\beta, \rho, g) \in \mathcal{B} \times \mathcal{P} \times \mathcal{G}$ satisfies $\|\beta - \beta_0\| = O_p(n^{-1/2})$, $|\rho - \rho_0| =$*

$O_p(n^{-1/2})$, $\|g^* - g_0^*\|_\infty \doteq \sup_{t\in[\tau_l,\tau_u]} |g^*(t) - g_0^*(t)| = O_p(n^{-1/2})$ with $g^* = g, \dot{g}$, and $\ddot{g}$, and $h \in L_2(P)$, we have

$$\begin{aligned} &\|P\{\dot{l}_\beta(\beta, g; O) - \dot{l}_\beta(\beta_0, g_0; O) - \ddot{l}_{\beta\beta}(\beta_0, g_0; O)(\beta - \beta_0) \\ &\qquad - \ddot{l}_{\beta g}(\beta_0, g_0; O)[g - g_0]\}\| = o_p(n^{-1/2}), \\ &\|P\{\dot{l}_g(\beta, g; O)[h] - \dot{l}_g(\beta_0, g_0; O)[h] - \ddot{l}_{g\beta}(\beta_0, g_0; O)[h](\beta - \beta_0) \\ &\qquad - \ddot{l}_{gg}(\beta_0, g_0; O)[h, g - g_0]\}\| = o_p(n^{-1/2}), \\ &\|P\{\Psi(\beta, \rho, g; Z) - \Psi(\beta_0, \rho_0, g_0; Z) - \dot{\Psi}_\beta(\beta_0, \rho_0, g_0; Z)(\beta - \beta_0) \\ &\qquad - \dot{\Psi}_\rho(\beta_0, \rho_0, g_0; Z)(\rho - \rho_0) - \dot{\Psi}_g(\beta_0, \rho_0, g_0; O)[g - g_0]\}\| = o_p(n^{-1/2}). \end{aligned}$$

Proof of Lemma 2: We only prove the first equality because the proofs of the other two equalities are similar. Applying Taylor expansion for $\dot{l}_\beta(\beta, g; O)$ at (β_0, g_0) yields

$$\dot{l}_\beta(\beta, g; O) - \dot{l}_\beta(\beta_0, g_0; O) = \ddot{l}_{\beta\beta}(\bar{\beta}, \bar{g}; O)(\beta - \beta_0) - \ddot{l}_{\beta g}(\bar{\beta}, \bar{g}; O)[g - g_0],$$

where $(\bar{\beta}, \bar{g}) \in \mathcal{B} \times \mathcal{G}$ lies between (β, g) and (β_0, g_0). Therefore, it suffices to show that

$$P\big[\{\ddot{l}_{\beta\beta}(\bar{\beta}, \bar{g}; O) - \ddot{l}_{\beta\beta}(\beta_0, g_0; O)\}(\beta - \beta_0)\big] = o_p(n^{-1/2}),$$

and

$$P\big\{\ddot{l}_{\beta g}(\bar{\beta}, \bar{g}; O)[g - g_0] - \ddot{l}_{\beta g}(\beta_0, g_0; O)[g - g_0]\big\} = o_p(n^{-1/2}).$$

Recall that

$$\ddot{l}_{\beta\beta}(\beta, g; O) = ZZ^\top \int \ddot{g}(t - Z^\top\beta)dN(t)$$
$$- ZZ^\top \int I(\log(X) \geq t)\exp\{g(t - Z^\top\beta)\}\{\ddot{g}(t - Z^\top\beta) - \dot{g}^2(t - Z^\top\beta)\}dt.$$

This implies

$$P\Big|\ddot{l}_{\beta\beta}(\bar{\beta}, \bar{g}; O) - \ddot{l}_{\beta\beta}(\beta_0, g_0; O)\Big| \leq I_1 + I_2 + I_3,$$

where

$$\begin{aligned}
I_1 &= P\Big\{ZZ^\top|\ddot{\bar{g}}(\log(X) - Z^\top\bar{\beta}) - \ddot{g}_0(\log(X) - Z^\top\beta_0)|\Big\}, \\
I_2 &= P\Big\{ZZ^\top\Big|\int \big\{\exp\{\bar{g}(t - Z^\top\bar{\beta})\}\ddot{\bar{g}}(t - Z^\top\bar{\beta}) - \exp\{g_0(t - Z^\top\beta_0)\}\ddot{g}_0(t - Z^\top\beta_0)\big\}dt\Big|\Big\}, \\
I_3 &= P\Big\{ZZ^\top\Big|\int \big\{\exp\{\bar{g}(t - Z^\top\bar{\beta})\}\dot{\bar{g}}^2(t - Z^\top\bar{\beta}) - \exp\{g_0(t - Z^\top\beta_0)\}\dot{g}_0^2(t - Z^\top\beta_0)\big\}dt\Big|\Big\}.
\end{aligned}$$

By Condition (C2)(b) and straightforward calculation, we have

$$I_1 = O(\|\beta - \beta_0\| + \|\ddot{g} - \ddot{g}\|_\infty) = O_p(n^{-1/2}).$$

Similarly, $I_2 = O_p(n^{-1/2})$ and $I_3 = O_p(n^{-1/2})$. Thus,

$$P[\{\ddot{l}_{\beta\beta}(\bar{\beta}, \bar{g}; O) - \ddot{l}_{\beta\beta}(\beta_0, g_0; O)\}(\beta - \beta_0)] = o_p(n^{-1/2}).$$

Similarly, we can show that

$$P\{\ddot{l}_{\beta g}(\bar{\beta}, \bar{g}; O)[g - g_0] - \ddot{l}_{\beta g}(\beta_0, g_0; O)[g - g_0]\} = o_p(n^{-1/2}).$$

This completes the proof of Lemma 2.

S2.2 Proof of Theorem 1(i)

First, we show with probability tending to one, the maximum conditional likelihood estimators $\tilde{\beta}$ and $\tilde{F}$ exist and are unique. Because $\tilde{F}(\cdot|\beta)$ is a smoothed version of $\breve{F}(\cdot|\beta)$, the existence and uniqueness of the former is guaranteed by those of the latter. For each given β, $\breve{F}(\cdot|\beta)$ assigns positive probability masses on all censored and uncensored residual points $e_j(\beta), j = 1, \dots, n$, and the complete-data log conditional likelihood function given the observed data is strictly concave in $p_j(\beta), j = 1, \dots, n$. By arguments similar to those in Vardi (1989), we can show that, given β and the supporting points $e_j(\beta), j = 1, \dots, n$, the maximizer of the complete-data log conditional likelihood of $p_j(\beta)$'s is unique, and that the first step of our expectation-maximization Algorithm 1 produces the unique maximizer $\breve{F}(\cdot|\beta)$ since the set of all feasible distributions $F(x|\beta) = \sum_{i=1}^n p_i(\beta) I(e_i(\beta) \le x)$ is convex. Here $p_i(\beta)$'s satisfy $p_i(\beta) \ge 0$ and $\sum_{i=1}^n p_i(\beta) = 1$. The existence of $\tilde{\beta}$ follows from the compactness of $\mathcal{B}$ and the continuity of $\ell_n(\beta, \tilde{F}(\cdot|\beta))$. It then follows from Condition (C6) that with probability tending to one, $\tilde{\beta}$ is unique. This also guarantees the uniqueness of $\tilde{F}$.

Second, we show the identifiability of (β_0, F_0). Suppose we have $P_{\beta,F} = P_{\beta_0,F_0}$

almost everywhere under P_{β_0,F_0}, where $P_{\beta,F}$ is the probability measure under (β, F). Consider the densities on $\delta = 0$, we can see that $S\{\log(t) - Z^\top\beta\} = S_0\{\log(t) - Z^\top\beta_0\}$ for every t, where $S(\cdot) = 1 - F(\cdot)$. Thus, there exists a monotone and differentiable function ϱ such that $T\exp\{-Z^\top\beta\} = \varrho\{T\exp\{-Z^\top\beta_0\}\}$. Differentiating both sides with respect to T yields $Z^\top(\beta - \beta_0) = -\log\{\varrho'(e^\epsilon)\}$. It then follows from condition (C5) that $\beta = \beta_0$, and thus $F(\cdot) = F_0(\cdot)$.

Finally, we show the consistency of $(\tilde{\beta}, \tilde{F})$. It is sufficient to show every convergent subsequence of $(\tilde{\beta}, \tilde{F})$ converges to the same limit (β_0, F_0) for any $t \in [\tau_l, \tau_u]$. Since $(\tilde{\beta}, \tilde{F})$ is bounded, according to Helly's selection theorem, for any subsequence of $(\tilde{\beta}, \tilde{F})$, there exists a further convergent subsequence. We show that any convergent subsequence converges to (β_0, F_0). For any subsequence $\{(\tilde{\beta}_{n_k}, \tilde{F}_{n_k}) : k = 1, 2, \ldots\}$, denote its limit as (β^*, F^*). It suffices to show $(\beta^*, F^*) = (\beta_0, F_0)$. Choose $\breve{\beta} = \beta_0$, and define

$$\breve{F}(t) = \frac{1}{n}\sum_{j=1}^{n}\big[\delta_j + \sum_{i=1}^{n}(1-\delta_i)\mathbb{E}_0\{I(\varepsilon_i^*(\beta_0) = e_j(\beta_0))\}\big]I(e_j(\beta_0) \le t),$$

where $\mathbb{E}_0$ is the expectation under (β_0, F_0). If (β_0, F_0) was used as the initial value in our expectation-maximization algorithm, $\breve{F}(t)$ is simply the one-step estimator of F. Applying the Glivenko-Cantelli theorem and a standard argument of Donsker class, we can show that $\breve{F}(t)$ converges to $F_0(t)$ almost surely and uniformly in $[\tau_l, \tau_u]$. Under Condition (C8), we have $\sup_{u\in[\tau_l,\tau_u]}\sup_{\beta\in\mathcal{B}}|\tilde{F}(u|\beta) - \breve{F}(u|\beta)| = o(1)$ and $\sup_{u\in[\tau_l,\tau_u]}\sup_{\beta\in\mathcal{B}}|d\tilde{F}(u|\beta) - d\breve{F}(u|\beta)| = o(1)$ almost surely. This, combined with the strong law of large numbers for empirical processes, implies that $n_k^{-1}\{\ell_{n_k}(\tilde{\beta}_{n_k}, \tilde{F}_{n_k}) - \ell_{n_k}(\check{\beta}, \check{F})\}$ converges almost surely to the negative Kullback-Leibler distance between P_{β^*,F^*} and P_{β_0,F_0}. On the other hand, since ℓ_{n_k} is maximized at $(\tilde{\beta}_{n_k}, \tilde{F}_{n_k})$, we have $\ell_{n_k}(\tilde{\beta}_{n_k}, \tilde{F}_{n_k}) - \ell_{n_k}(\check{\beta}, \check{F}) \geq 0$. Hence $P_{\beta^*,F^*} = P_{\beta_0,F_0}$ almost surely. Thus by model identifiability, we have $\beta^* = \beta_0$, $F^* = F_0$. The continuity and monotonicity of F_0 ensures the uniform convergence of $\tilde{F}$ in $[\tau_l, \tau_u]$. This completes the proof of Theorem 1(i).

S2.3 Proof of Theorem 1(ii)

Let $\tilde{g}(t) = \log[\{d\tilde{F}(t|\tilde{\beta})/dt\}/\{1 - \tilde{F}(t|\tilde{\beta})\}]$. We prove the asymptotic normality of $n^{1/2}(\tilde{\beta} - \beta_0, \tilde{g} - g_0)$ by applying the Z-theorem for the infinite-dimensional estimating equations (Theorem 3.3.1 in van der Vaart and Wellner (1996)). We would verify the three main conditions: the Fréchet differentiability of the score functions, weak convergence of estimating equations, and the stochastic approximation of estimating equations.

The ordinary derivative of $n^{-1}\ell_n(\beta, g)$ with respect to β is $\mathbb{P}_n\{\dot{l}_\beta(\beta, g; O)\}$, where

$$\dot{l}_\beta(\beta, g; O) = -Z\Big\{\int \dot{g}(t)dN(t, \beta) - \int Y(t, \beta)\exp\{g(t)\}\dot{g}(t)dt\Big\}.$$

To obtain the score equation for g, we consider a submodel defined by $g_\eta(u) = g(u) + \eta\tilde{h}_t(u)$, where $\tilde{h}_t(\cdot) = I(\cdot \leq t)$. Differentiating $n^{-1}\ell_n(\beta, g_\eta)$ with respect to η and evaluating the partial derivative at $\eta = 0$ give the partial derivative of

$n^{-1}\ell_n(\beta, g_\eta)$ with respect to g along the direction $\tilde{h}_t$, i.e. $\mathbb{P}_n\{\dot{l}_g(\beta, g; O)[\tilde{h}_t]\}$, where

$$\dot{l}_g(\beta, g; O)[\tilde{h}_t] = \int_{-\infty}^{t} dN(u, \beta) - \int_{-\infty}^{t} Y(u, \beta)\exp\{g(u)\}du.$$

Define $U_n(\beta, g)(t) = (U_{1n}^\top(\beta, g), U_{2n}(\beta, g)(t))^\top$ with $U_{1n}(\beta, g) = \mathbb{P}_n\{\dot{l}_\beta(\beta, g; O)\}$ and $U_{2n}(\beta, g)(t) = \mathbb{P}_n\{\dot{l}_g(\beta, g; O)[\tilde{h}_t]\}$, and similarly define $U(\beta, g)(t) = (U_1^\top(\beta, g), U_2(\beta, g)(t))^\top$ with $U_1(\beta, g) = P\{\dot{l}_\beta(\beta, g; O)\}$ and $U_2(\beta, g)(t) = P\{\dot{l}_g(\beta, g; O)[\tilde{h}_t]\}$. Both the score function U_n and its expectation U are defined on $\mathcal{B} \times \mathcal{G}$, where $\mathcal{B}$ is assumed to be compact in $\mathbb{R}^p$, and $\mathcal{G}$ consists of functions of bounded variations.

We first show that U is Frechét differentiable at (β_0, g_0) and its Frechét derivative is continuously invertible. Consider submodels $(\beta_\eta, g_\eta) = (\beta_0 + \eta\beta, g_0 + \eta g)$, the Gâteaux derivative of U at (β_0, g_0) can be obtained by taking the derivative of $U(\beta_\eta, g_\eta, t)$ with respect to η and evaluating at $\eta = 0$. Then, the Gâteaux derivative of U at (β_0, g_0) is given by

$$\dot{U}_0(\beta, g)(t) = \begin{pmatrix} \sigma_{11}(\beta) + \sigma_{12}(g) \\ \sigma_{21}(\beta)(t) + \sigma_{22}(g)(t) \end{pmatrix},$$

where

$$\begin{aligned}
\sigma_{11}(\beta) &= \frac{\partial}{\partial \eta} U_1(\beta_\eta, g_0)\Big|_{\eta=0} = P\{\ddot{l}_{\beta\beta}(\beta_0, g_0; O)\}\beta \\
&= \mathbb{E}\Big[-ZZ^\top \int_{\tau_l}^{\tau_u} Y(t, \beta_0)\exp\{g_0(t)\}\dot{g}_0^2(t)dt\Big]\beta \equiv J_0\beta, \\
\sigma_{12}(g) &= \frac{\partial}{\partial \eta} U_1(\beta_0, g_\eta)\Big|_{\eta=0} = P\{\ddot{l}_{\beta g}(\beta_0, g_0; O)[g]\},
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\Big[Z\int_{\tau_l}^{\tau_u} Y(t,\beta_0)\exp\{g_0(t)\}\dot{g}_0(t)g(t)dt\Big],\\
\sigma_{21}(\beta)(t) &= \frac{\partial}{\partial\eta}U_2(\beta_\eta, g_0, t)\Big|_{\eta=0} = \mathbb{E}\Big[\delta Z^\top \bar{f}(t) + Z^\top\int_{\tau_l}^{t}\bar{f}(u)\exp\{g_0(u)\}du\Big]\beta,\\
\sigma_{22}(g)(t) &= \frac{\partial}{\partial\eta}U_2(\beta_0, g_\eta, t)\Big|_{\eta=0} = \mathbb{E}\Big[-\int_{\tau_l}^{t}\{1-\bar{F}(u)\}\exp\{g_0(u)\}g(u)du\Big],
\end{aligned}
$$

with $\bar{F}(\cdot)$ and $\bar{f}(\cdot)$ being the distribution and density functions, respectively, of $\log(X) - Z^\top\beta_0$. In deriving $\sigma_{11}(\beta)$ and $\sigma_{12}(g)$, we use the fact $\mathbb{E}\{dM(t,\beta_0)\} = \mathbb{E}\{dN(t,\beta_0) - Y(t,\beta_0)\exp\{g_0(t)\}dt\} = 0$. In deriving $\sigma_{21}(\beta)$ and $\sigma_{22}(g)$, we have used the fact that

$$
U_2(\beta,g)(t) = \mathbb{E}\Big[\delta\bar{F}(t+Z^\top(\beta-\beta_0)) - \int_{\tau_l}^{t}\{1-\bar{F}(u+Z^\top(\beta-\beta_0))\}\exp\{g(u)\}du\Big].
$$

The Gâteaux derivative of U at (β, g) is continuous. By similar arguments to those in the proof of Lemma 15.8 in Kosorok (2008), we can show that U is Frechét differentiable, and its derivative at (β_0, g_0) is $\dot{U}_0$. Note that the operator $\dot{U}_0$ is a linear continuous operator defined on the product space of $\mathbb{R}^p$ and the Banach space $L_2[\tau_l,\tau_u]$, where $L_2[\tau_l,\tau_u]$ is the space of functions on $[\tau_l,\tau_u]$ with finite L_2 norm. By Banach's continuous inverse theorem (Zeidler, 1995), if the inverse operator $\dot{U}_0$ exists, then it must be continuous. Therefore, to prove the continuous invertibility of U_0, it suffices to show that the inverse operator of $\dot{U}_0$ exists. Actually, if σ_{11} and

$\Phi = \sigma_{22} - \sigma_{21}\sigma_{11}^{-1}\sigma_{12}$ are invertible, then the inverse of $\dot{U}_0$ is

$$\dot{U}_0^{-1} = \begin{pmatrix} \sigma_{11}^{-1} + \sigma_{11}^{-1}\sigma_{12}\Phi^{-1}\sigma_{21}\sigma_{11}^{-1} & -\sigma_{11}^{-1}\sigma_{12}\Phi^{-1} \\ -\Phi^{-1}\sigma_{21}\sigma_{11}^{-1} & \Phi^{-1} \end{pmatrix}.$$

Under condition (C6), the matrix J_0 in σ_{11} is invertible, so is σ_{11}. The operator Φ has the following form:

$$\Phi(g)(t) = \int_{\tau_l}^{t} Q(u)g(u)du + \int_{\tau_l}^{\tau_u} R(t,u)g(u)du,$$

where $Q(u) = \mathbb{E}[-\{1 - \bar{F}(u)\}\exp\{g_0(u)\}]$ and

$$R(t,u) = \mathbb{E}\Big[\delta Z^\top \bar{f}(t) + Z^\top \int_{\tau_l}^{t} \bar{f}(u)\exp\{g_0(u)\}du\Big] J_0^{-1}\mathbb{E}[ZY(u,\beta_0)\exp\{g_0(u)\}\dot{g}_0(u)].$$

Following Qin et al. (2011) and Huang et al. (2015), it can be shown that Φ has an inverse, i.e.

$$\Phi^{-1}(g)(t) = \int_{\tau_l}^{t} \frac{g(u)}{Q(u)}du - \int_{\tau_l}^{\tau_u} \Big(\int_{\tau_l}^{t} \frac{H(u,v)}{Q(u)}du\Big) g(v)dv,$$

where

$$H(u,v) = -\frac{\dot{R}(u,v)}{Q(v)} - \int_{\tau_l}^{\tau_u} H(u,s)\frac{\dot{R}(s,v)}{Q(v)}ds$$

with $\dot{R}(t,u) = \partial R(t,u)/\partial t$.

Second, we show the weak convergence of $n^{1/2}U_n(\beta_0, g_0)(t), t \in [\tau_l, \tau_u]$. Write

$$U_n(\beta_0, g_0)(t) = U_n(\beta_0, g_0)(t) - U(\beta_0, g_0)(t) = (\mathbb{P}_n - P) \begin{pmatrix} \dot{l}_\beta(\beta_0, g_0; O) \\ \dot{l}_g(\beta_0, g_0; O)[\tilde{h}_t]) \end{pmatrix},$$

which, up to a constant $1/n$, is the sum of zero-mean independently and identically distributed variables for each t, the multivariate central limit theorem implies that $n^{1/2}U_n(\beta_0, g_0)(t)$ converges in finite-dimensional distribution to a zero-mean Gaussian process for $t \in [\tau_l, \tau_u]$. Since $\dot{l}_\beta(\beta_0, g_0; O)$ does not depend on t, $\dot{l}_g(\beta_0, g_0; O)[\tilde{h}_t]$ is the difference of two increasing functions $\int_{-\infty}^t dN(u, \beta)$ and $\int_{-\infty}^t Y(u, \beta) \exp\{g(u)\} du$, where the counting process $N(u, \beta)$ and the integrand $Y(u, \beta) \exp\{g(u)\}$ are nonnegative and bounded above, it is easy to show that they are asymptotic equicontinuous and are thus manageable (Pollard, 1990). Therefore, $n^{1/2}\{U_n(\beta_0, g_0)(\cdot) - U(\beta_0, g_0)(\cdot)\}$ is tight and converges weakly to a Gaussian process $\mathbb{W}$.

Finally, we establish the stochastic approximation $\|\sqrt{n}\{U_n(\tilde{\beta}, \tilde{g})(\cdot) - U(\tilde{\beta}, \tilde{g})(\cdot)\} - \sqrt{n}\{U_n(\beta_0, g_0)(\cdot) - U(\beta_0, g_0)(\cdot)\}\| = o_p(1)$. The function is defined on $\mathcal{B} \times \mathcal{G}$, where $\mathcal{B}$ is a compact set that contains β_0 and $\mathcal{G}$ is a set of functions with bounded variation. Let $\bar{\mathcal{G}}$ be the closed linear space generated by $\mathcal{G}$, endowed with the total variation norm $\|\cdot\|_v$. Define $\|(\beta, g)\|_{\mathcal{B}\times\mathcal{G}} = \|\beta\| + \|g\|_v$, where $\|\beta\|$ is the Euclidean norm of β. Write $\vartheta(\beta, g, t; O) = (\dot{l}_\beta(\beta, g; O)^\top, \dot{l}_g(\beta, g; O)[\tilde{h}_t]))$. We show that the class of functions $\{\vartheta(\beta, g, t; O) - \vartheta(\beta_0, g_0, t; O) : t \in [\tau_l, \tau_u], \beta \in \mathcal{B}, g \in \mathcal{G}, \|(\beta, g) - (\beta_0, g_0)\|_{\mathcal{B}\times\mathcal{G}} < \iota\}$ is a Donsker class, where ι is a fixed small number. We claim that $\mathcal{B}$ and $\mathcal{G}$ are Donsker classes because $\mathcal{B}$ is a compact subset of $\mathbb{R}^p$ and $\mathcal{G}$ consists of functions with bounded variation. The class of functions $\{\vartheta(\beta, g, t; O) - \vartheta(\beta_0, g_0, t; O) : (\beta, g) \in \mathcal{B} \times \mathcal{G}, t \in [\tau_l, \tau_u]\}$ is a Donsker class, as the summation, production and Lipschitz transformations of Donsker classes are also Donsker. Furthermore, as $\|(\beta, g) - (\beta_0, g_0)\|_{\mathcal{B}\times\mathcal{G}} \to 0$, we have $\sup_{t\in[\tau_l,\tau_u]} \mathbb{E}\|\vartheta(\beta, g, t; O) - \vartheta(\beta_0, g_0, t; O)\|^2 \to 0$. By result (i) and condition (C8), $\|(\tilde{\beta}, \tilde{g}) - (\beta_0, g_0)\|_{\mathcal{B}\times\mathcal{G}} \to 0$. It follows from Lemma 3.3.5 of van der Vaart and Wellner (1996) that $\|\sqrt{n}\{U_n(\tilde{\beta}, \tilde{g}, t) - U(\tilde{\beta}, \tilde{g}, t)\} - \sqrt{n}\{U_n(\beta_0, g_0, t) - U(\beta_0, g_0, t)\}\| = o_p(1)$. According to Theorem 3.3.1 of van der Vaart and Wellner (1996), $n^{1/2}\{(\tilde{\beta}, \tilde{g}) - (\beta_0, g_0)\}$ converges weakly to the mean zero Gaussian process $-\dot{U}_0^{-1}(\mathbb{W})$.

Let ψ be the transformation from (β, g) to (β, F) with $\psi(\beta_0, g_0) = (\beta_0, F_0)$. It is easy to check that the mapping is Hadamard differentiable. Following the functional delta method, we can show that $n^{1/2}\{(\tilde{\beta}, \tilde{F}) - (\beta_0, F_0)\}$ converges weakly to a tight mean zero Gaussian process $\psi_0'\{-\dot{U}_0^{-1}(\mathbb{W})\}$, where ψ_0' is the Hadamard derivative of ψ at (β_0, g_0). This completes the proof of Theorem 1(ii).

S2.4 Proof of Theorem 1(iii)

Without incorporating the auxiliary information, the estimator $(\tilde{\beta}, \tilde{g})$ satisfies

$$\mathbb{P}_n\Big\{\dot{l}_\beta(\tilde{\beta}, \tilde{g}; O)\Big\} = 0,$$

$$\mathbb{P}_n\Big\{\dot{l}_g(\tilde{\beta}, \tilde{g}; O)[h]\Big\} = 0,$$

for any $h \in L_2(P)$. It follows from Lemmas 1 and 2 and the consistency of $(\tilde{\beta}, \tilde{g})$ that

$$\begin{aligned}
-\mathbb{P}_n\big\{\dot{l}_\beta(\beta_0, g_0; O)\big\} =& P\big\{\ddot{l}_{\beta\beta}(\beta_0, g_0; O)\big\}(\tilde{\beta} - \beta_0) + P\big\{\ddot{l}_{\beta g}(\beta_0, g_0; O)[\tilde{g} - g_0]\big\} + o_p(n^{-1/2}), \\
-\mathbb{P}_n\{\dot{l}_g(\beta_0, g_0; O)[h_1^*]\} =& P\big\{\ddot{l}_{g\beta}(\beta_0, g_0; O)[h_1^*]\big\}(\tilde{\beta} - \beta_0) + P\big\{\ddot{l}_{gg}(\beta_0, g_0; O)[h_1^*, \tilde{g} - g_0]\big\} \\
& + o_p(n^{-1/2}) \\
=& P\big\{\ddot{l}_{g\beta}(\beta_0, g_0; O)[h_1^*]\big\}(\tilde{\beta} - \beta_0) + P\big\{\ddot{l}_{\beta g}(\beta_0, g_0; O)[\tilde{g} - g_0]\big\} \\
& + o_p(n^{-1/2}),
\end{aligned}$$

where we have used equality (S3.6). The difference between the above two equations yields

$$n^{1/2}(\tilde{\beta} - \beta_0) = \Sigma^{-1} \cdot n^{1/2}\mathbb{P}_n\big\{\iota(\beta_0, g_0; O)\big\} + o_p(1),$$

where

$$\iota(\beta_0, g_0; O) = \dot{l}_\beta(\beta_0, g_0; O) - \dot{l}_g(\beta_0, g_0; O)[h_1^*], \tag{S2.1}$$

because

$$\Sigma \equiv P\big\{\iota(\beta_0, g_0; O)^{\otimes 2}\big\} = P\big\{ -\ddot{l}_{\beta\beta}(\beta_0, g_0; O) + \ddot{l}_{g\beta}(\beta_0, g_0; O)[h_1^*]\big\},$$

which will be proved in the proof of Theorem 2(ii). Therefore, $n^{1/2}(\tilde{\beta} - \beta_0)$ converges in distribution to a normal distribution with mean 0 and variance Σ^{-1}.

S3 Asymptotic results of $\hat{\beta}_{\mathrm{au}}$

S3.1 Asymptotic properties of $\hat{\beta}_{\mathrm{au}}$

Theorem S1 presents the consistency and asymptotic normality of $\hat{\beta}_{\mathrm{au}}$.

Theorem S1. *Suppose that conditions (C1)–(C8) and (D1)–(D2) are satisfied. As $n \to \infty$, (i) $\hat{\beta}_{\mathrm{au}}$ is consistent to β_0, and (ii) $n^{1/2}(\hat{\beta}_{\mathrm{au}} - \beta_0)$ converges to a normal distribution with mean zero and covariance $\{\Sigma + BQ^{-1}B^\top - BQ^{-1}A(A^\top Q^{-1}A)^{-1}A^\top Q^{-1}B^\top\}^{-1}$, provided the matrices Σ and $A^\top Q^{-1}A$ are nonsingular. (iii) $\hat{\beta}_{\mathrm{au}}$ is asymptotically more efficient than $\tilde{\beta}$.*

S3.2 Proof of Theorem S1(i)

Our proof follows from the proof of Theorem 1 in Qin and Lawless (1994). The difference is that Qin and Lawless (1994) handled finite-dimensional parameters whereas we deal with the finite-dimensional parameters and infinite-dimensional nuisance function simultaneously. Let $\hat{g}_{\mathrm{au}}(t) = \log[\{d\hat{F}_{\mathrm{au}}(t)/dt\}/\{1 - \hat{F}_{\mathrm{au}}(t)\}]$ and define

$$L(\beta, \rho, g) = -\sum_{i=1}^{n} l(\beta, g; O_i) + \sum_{i=1}^{n} \log\left\{1 + \nu^\top(\beta, \rho, g)\Psi(\beta, \rho, g; Z_i)\right\},$$

where we write the Lagrangian multiplier ν as a function of (β, ρ, g) to highlight its dependence on the latter. Note that

$$(\hat{\beta}_{\mathrm{au}}, \hat{\rho}_{\mathrm{au}}, \hat{g}_{\mathrm{au}}) = \mathrm{argmin}_{\beta,\rho,g}\mathrm{max}_\nu L(\beta, \rho, g).$$

Keep in mind that $L(\beta,\rho,g)$ is continuous with respect to (β,ρ,g).

Let $S_n = \{(\beta,\rho,g) : \|(\beta,\rho,g)-(\beta_0,\rho_0,g_0)\|_{\mathcal{B}\times\mathbb{R}^+\times\mathcal{G}} = n^{-1/3}\}$. For $(\beta,\rho,g)\in S_n$, as $\mathbb{E}\{\|\Psi(\beta,\rho,g;Z_i)\|^3\}<\infty$, similar to the proof of Qin and Lawless (1994), we can show that

$$\sum_{i=1}^{n}\log\left\{1+\nu(\beta,\rho,g)^\top\Psi(\beta,\rho,g;Z_i)\right\}\geq c\cdot n^{1/3}$$

for some positive constant c, and

$$\sum_{i=1}^{n}\log\left\{1+\nu(\beta_0,\rho_0,g_0)^\top\Psi(\beta_0,\rho_0,g_0;Z_i)\right\}=O(\log\log(n)).$$

Meanwhile

$$\begin{aligned}-\sum_{i=1}^{n}l(\beta,g;O_i) &= -\sum_{i=1}^{n}l(\beta_0,g_0;O_i)-\sum_{i=1}^{n}l_\beta^\top(\beta_0,g_0;O_i)(\beta-\beta_0)\\ &\quad -\sum_{i=1}^{n}l_g(\beta_0,g_0;O_i)[g-g_0]+O_p(n^{1/3})\\ &\geq -\sum_{i=1}^{n}l(\beta_0,g_0;O_i)+O_p(n^{1/3}).\end{aligned}$$

In summary, we have

$$L(\beta,\rho,g)\geq L(\beta_0,\rho_0,g_0)+O_p(n^{1/3}).$$

It follows that as n is large, the minimizer $(\hat\beta_{\rm au},\hat\rho_{\rm au},\hat g_{\rm au})$ should lie within the ball S_n. This implies the consistency of $(\hat\beta_{\rm au},\hat\rho_{\rm au},\hat g_{\rm au})$.

S3.3 Proof of Theorem S1(ii)

Define $L_n(\beta, \rho, \nu, g) = \sum_{i=1}^n l(\beta, g; O_i) - \sum_{i=1}^n \log\{1 + \nu^\top \Psi(\beta, \rho, g; Z_i)\}$. For any $h \in L_2(P)$, the estimator $(\hat{\beta}_{\mathrm{au}}, \hat{\nu}_{\mathrm{au}}, \hat{\rho}_{\mathrm{au}}, \hat{g}_{\mathrm{au}})$ satisfies

$$
\begin{aligned}
\mathbb{P}_n\Big\{\dot{l}_\beta(\hat{\beta}_{\mathrm{au}}, \hat{g}_{\mathrm{au}}; O) - \frac{\hat{\nu}_{\mathrm{au}}^\top \dot{\Psi}_\beta(\hat{\beta}_{\mathrm{au}}, \hat{\rho}_{\mathrm{au}}, \hat{g}_{\mathrm{au}}; Z)}{1 + \hat{\nu}_{\mathrm{au}}^\top \Psi(\hat{\beta}_{\mathrm{au}}, \hat{\rho}_{\mathrm{au}}, \hat{g}_{\mathrm{au}}; Z)}\Big\} &= 0, \\
\mathbb{P}_n\Big\{\dot{l}_g(\hat{\beta}_{\mathrm{au}}, \hat{g}_{\mathrm{au}}; O)[h] - \frac{\hat{\nu}_{\mathrm{au}}^\top \dot{\Psi}_g(\hat{\beta}_{\mathrm{au}}, \hat{\rho}_{\mathrm{au}}, \hat{g}_{\mathrm{au}}; Z)[h]}{1 + \hat{\nu}_{\mathrm{au}}^\top \Psi(\hat{\beta}_{\mathrm{au}}, \hat{\rho}_{\mathrm{au}}, \hat{g}_{\mathrm{au}}; Z)}\Big\} &= 0, \\
\mathbb{P}_n\Big\{\frac{\Psi(\hat{\beta}_{\mathrm{au}}, \hat{\rho}_{\mathrm{au}}, \hat{g}_{\mathrm{au}}; Z)}{1 + \hat{\nu}_{\mathrm{au}}^\top \Psi(\hat{\beta}_{\mathrm{au}}, \hat{\rho}_{\mathrm{au}}, \hat{g}_{\mathrm{au}}; Z)}\Big\} &= 0, \\
\mathbb{P}_n\Big\{\frac{\hat{\nu}_{\mathrm{au}}^\top \dot{\Psi}_\rho(\hat{\beta}_{\mathrm{au}}, \hat{\rho}_{\mathrm{au}}, \hat{g}_{\mathrm{au}}; Z)}{1 + \hat{\nu}_{\mathrm{au}}^\top \Psi(\hat{\beta}_{\mathrm{au}}, \hat{\rho}_{\mathrm{au}}, \hat{g}_{\mathrm{au}}; Z)}\Big\} &= 0.
\end{aligned}
$$

By Lemmas 1 and 2 and the consistency of $(\hat{\beta}_{\mathrm{au}}, \hat{\rho}_{\mathrm{au}}, \hat{g}_{\mathrm{au}})$, we can show that $\hat{\nu}_{\mathrm{au}} = o_p(1)$. Further, using first-order Taylor series approximations, we can rewrite the above equations as

$$
\begin{aligned}
-\mathbb{P}_n\big\{\dot{l}_\beta(\beta_0, g_0; O)\big\} =& P\big\{\ddot{l}_{\beta\beta}(\beta_0, g_0; O)(\hat{\beta}_{\mathrm{au}} - \beta_0) + \ddot{l}_{\beta g}(\beta_0, g_0; O)[\hat{g}_{\mathrm{au}} - g_0]\big\} \\
& - P\big\{\dot{\Psi}_\beta(\beta_0, \rho_0, g_0; O)\big\}\hat{\nu}_{\mathrm{au}} + o_p(n^{-1/2}),
\end{aligned} \tag{S3.2}
$$

$$
\begin{aligned}
-\mathbb{P}_n\{\dot{l}_g(\beta_0, g_0; O)[h]\} =& P\big\{\ddot{l}_{g\beta}(\beta_0, g_0; O)[h]\big\}(\hat{\beta}_{\mathrm{au}} - \beta_0) + P\big\{\ddot{l}_{gg}(\beta_0, g_0; O)[h, \hat{g}_{\mathrm{au}} - g_0]\big\} \\
& - P\big\{\dot{\Psi}_g(\beta_0, \rho_0, g_0; Z)[h]\big\}^\top \hat{\nu}_{\mathrm{au}} + o_p(n^{-1/2}),
\end{aligned} \tag{S3.3}
$$

$$
\begin{aligned}
-\mathbb{P}_n\{\Psi(\beta_0, \rho_0, g_0; Z)\} =& P\big\{\dot{\Psi}_\beta(\beta_0, \rho_0, g_0; Z)\big\}^\top (\hat{\beta}_{\mathrm{au}} - \beta_0) \\
& + P\{\dot{\Psi}_g(\beta_0, \rho_0, g_0; Z)[\hat{g}_{\mathrm{au}} - g_0]\} - P\big\{\Psi(\beta_0, \rho_0, g_0; Z)^{\otimes 2}\big\}\hat{\nu}_{\mathrm{au}} \\
& + P\big\{\dot{\Psi}_\rho(\beta_0, \rho_0, g_0; Z)\big\}(\hat{\rho}_{\mathrm{au}} - \rho_0) + o_p(n^{-1/2}),
\end{aligned} \tag{S3.4}
$$

$$0 = P\Big\{\dot{\Psi}_{\rho}(\beta_0, \rho_0, g_0; Z)\Big\}^{\top}\hat{\nu}_{\text{au}} + o_p(n^{-1/2}). \tag{S3.5}$$

To derive an asymptotic representation of $(\hat{\beta}_{\text{au}}, \hat{\rho}_{\text{au}})$, we need to find the least favorable directions h_1^* and h_2^*, which can be used to profile out the infinite-dimensional parameter g. Let $h_1^* = (h_{11}^*, \ldots, h_{1p}^*)^{\top}$ be the direction that satisfies

$$P\Big\{\ddot{l}_{\beta g}(\beta_0, g_0; O)[h] - \ddot{l}_{gg}(\beta_0, g_0; O)[h_1^*, h]\Big\} = 0 \tag{S3.6}$$

for all $h \in \mathcal{H}$, where $h_{1j}^* \in \mathcal{H}, j = 1, \ldots, p$. Let $h_2^* = (h_{21}^*, \ldots, h_{2J}^*)^{\top}$ be the direction that satisfies

$$P\Big\{\ddot{l}_{gg}(\beta_0, g_0; O)[h_2^*, h]\Big\} = P\Big\{\dot{\Psi}_g(\beta_0, \rho_0, g_0; Z)[h]\Big\} \tag{S3.7}$$

for all $h \in \mathcal{H}$, where $h_{2j}^* \in \mathcal{H}, j = 1, \ldots, J$. By direct calculation,

$$\begin{aligned}
&P\Big\{\ddot{l}_{\beta g}(\beta_0, g_0; O)[h] - \ddot{l}_{gg}(\beta_0, g_0; O)[h_1^*, h]\Big\}\\
=&P\Big\{ - Z\int \dot{h}(t)dN(t;\beta_0) + Z\int Y(t;\beta_0)\exp\{g_0(t)\}\{h(t)\dot{g}_0(t) + \dot{h}(t)\}dt\\
&\quad + \int Y(t;\beta_0)\exp\{g_0(t)\}h_1^*(t)h(t)dt\Big\}\\
=&P\Big\{\int ZY(t;\beta_0)\exp\{g_0(t)\}h(t)\dot{g}_0(t)dt + \int Y(t;\beta_0)\exp\{g_0(t)\}h(t)h_1^*(t)dt\Big\}\\
=&P\Big\{\int Y(t;\beta_0)\exp\{g_0(t)\}h(t)\{Z\dot{g}_0(t) + h_1^*(t)\}dt\Big\},
\end{aligned}$$

where the second equality holds because

$$P\{\dot{l}_g(\beta_0, g_0; O)[\dot{h}]\} = P\Big\{\int \dot{h}(t)dN(t;\beta_0) - \int Y(t;\beta_0)\exp\{g_0(t)\}\dot{h}(t)dt\Big\} = 0.$$

To guarantee (S3.6), we choose h_1^* to be

$$h_1^*(t,\beta_0) = -\dot{g}_0(t)\mathbb{E}\{Z|Y(t;\beta_0)\} = -\dot{g}_0(t)\frac{\mathbb{E}\{ZI(\log(X) - Z^\top\beta_0 \geq t)\}}{\mathbb{E}\{I(\log(X) - Z^\top\beta_0 \geq t)\}}.$$

When the aggregate information is the subgroup survival probabilities, one obvious choice of h_2^* is $h_2^*(t,\beta) = (h_{21}^*(t,\beta), \ldots, h_{2J}^*(t,\beta))^\top$ with

$$h_{2j}^*(t,\beta_0) = \frac{\mathbb{E}[I(Z\in\Omega_j)I(t \leq \log t_j^* - Z^\top\beta_0)\{1 - F(\log t_j^* - Z^\top\beta_0)\}]}{\mathbb{E}\{Y(t,\beta_0)\}}, \quad j = 1,\ldots,J.$$

It follows from (S3.6) and (S3.7) that

$$P\Big\{\ddot{l}_{\beta g}(\beta_0, g_0; O)[\hat{g}_{\rm au} - g_0] - \ddot{l}_{gg}(\beta_0, g_0; O)[h_1^*, \hat{g}_{\rm au} - g_0]\Big\} = 0, \tag{S3.8}$$

$$P\Big\{\ddot{l}_{gg}(\beta_0, g_0; O)[h_2^*, \hat{g}_{\rm au} - g_0]\Big\} = P\Big\{\dot{\Psi}_g(\beta_0, \rho_0, g_0; Z)[\hat{g}_{\rm au} - g_0]\Big\}. \tag{S3.9}$$

And it follows from (S3.3) that

$$\begin{aligned}-\mathbb{P}_n\{\dot{l}_g(\beta_0, g_0; O)[h_1^*]\} =& P\Big\{\ddot{l}_{g\beta}(\beta_0, g_0; O)[h_1^*]\Big\}(\hat{\beta}_{\rm au} - \beta_0) + P\Big\{\ddot{l}_{gg}(\beta_0, g_0; O)[h_1^*, \hat{g}_{\rm au} - g_0]\Big\} \\ &- P\Big\{\dot{\Psi}_g(\beta_0, \rho_0, g_0; Z)[h_1^*]\Big\}^\top\hat{\nu}_{\rm au} + o_p(n^{-1/2}), \qquad (S3.10)\end{aligned}$$

$$-\mathbb{P}_n\{\dot{l}_g(\beta_0, g_0; O)[h_2^*]\} = P\Big\{\ddot{l}_{g\beta}(\beta_0, g_0; O)[h_2^*]\Big\}(\hat{\beta}_{\rm au} - \beta_0) + P\Big\{\ddot{l}_{gg}(\beta_0, g_0; O)[h_2^*, \hat{g}_{\rm au} - g_0]\Big\}$$

$$- P\big\{\dot{\Psi}_g(\beta_0,\rho_0,g_0;Z)[h_2^*]\big\}^\top \hat{\nu}_{\rm au} + o_p(n^{-1/2}). \tag{S3.11}$$

Subtracting (S3.2) from (S3.10) and using (S3.8), we have

$$\begin{aligned}\mathbb{P}_n\big\{\iota(\beta_0,g_0;O)\big\} =& P\big\{ - \ddot{l}_{\beta\beta}(\beta_0,g_0;O) + \ddot{l}_{g\beta}(\beta_0,g_0;O)[h_1^*]\big\}(\hat{\beta}_{\rm au}-\beta_0) \\ &+ P\big\{\dot{\Psi}_\beta(\beta_0,\rho_0,g_0;Z) - \{\dot{\Psi}_g(\beta_0,\rho_0,g_0;Z)[h_1^*]\}^\top\big\}\hat{\nu}_{\rm au} + o_p(n^{-1/2}),\end{aligned} \tag{S3.12}$$

where $\iota(\beta_0,g_0;O)$ is defined in (S2.1). Subtracting (S3.4) from (S3.11) and using (S3.9), we have

$$\begin{aligned}\mathbb{P}_n\big\{\chi(\beta_0,\rho_0,g_0;O)\big\} =& P\big\{\ddot{l}_{g\beta}(\beta_0,g_0;O)[h_2^*] - \{\dot{\Psi}_\beta(\beta_0,\rho_0,g_0;Z)\}^\top\big\}(\hat{\beta}_{\rm au}-\beta_0) \\ &+ P\big\{\Psi(\beta_0,\rho_0,g_0;Z)^{\otimes 2} - \{\dot{\Psi}_g(\beta_0,\rho_0,g_0;Z)[h_2^*]\}^\top\big\}\hat{\nu}_{\rm au} \\ &- P\big\{\dot{\Psi}_\rho(\beta_0,\rho_0,g_0;Z)\big\}(\hat{\rho}_{\rm au}-\rho_0) + o_p(n^{-1/2}),\end{aligned} \tag{S3.13}$$

where $\chi(\beta_0,\rho_0,g_0;O) = \Psi(\beta_0,\rho_0,g_0;Z) - \dot{l}_g(\beta_0,g_0;O)[h_2^*]$.

We now show that

$$\Sigma \equiv P\big\{\iota(\beta_0,g_0;O)^{\otimes 2}\big\} = P\big\{ - \ddot{l}_{\beta\beta}(\beta_0,g_0;O) + \ddot{l}_{g\beta}(\beta_0,g_0;O)[h_1^*]\big\}, \tag{S3.14}$$

$$Q \equiv P\big\{\chi(\beta_0,\rho_0,g_0;O)^{\otimes 2}\big\} = P\big\{\Psi(\beta_0,\rho_0,g_0;Z)^{\otimes 2} - \{\dot{\Psi}_g(\beta_0,\rho_0,g_0;Z)[h_2^*]\}^\top\big\}, \tag{S3.15}$$

$$-B^\top = P\big\{\ddot{l}_{g\beta}(\beta_0,g_0;O)[h_2^*] - \{\dot{\Psi}_\beta(\beta_0,\rho_0,g_0;O)\}^\top\big\}. \tag{S3.16}$$

The property of the score function implies that

$$
\begin{aligned}
P\{\ddot{l}_{\beta\beta}(\beta_0, g_0; O)\} &= -P\{\dot{l}_\beta(\beta_0, g_0; O)\dot{l}_\beta^\top(\beta_0, g_0; O)\}, \\
P\{\ddot{l}_{\beta g}(\beta_0, g_0; O)[h_1^*]\} &= -P\{\dot{l}_\beta(\beta_0, g_0; O)\dot{l}_g^\top(\beta_0, g_0; O)[h_1^*]\}, \\
P\{\ddot{l}_{g\beta}(\beta_0, g_0; O)[h_1^*]\} &= -P\{\dot{l}_g(\beta_0, g_0; O)[h_1^*]\dot{l}_\beta^\top(\beta_0, g_0; O)\}, \\
P\{\ddot{l}_{gg}(\beta_0, g_0; O)[h_1^*, h_1^*]\} &= -P\{\dot{l}_g(\beta_0, g_0; O)[h_1^*]\dot{l}_g^\top(\beta_0, g_0; O)[h_1^*]\}.
\end{aligned}
$$

This, together with the fact

$$
P\{\ddot{l}_{\beta g}(\beta_0, g_0; O)[h_1^*] - \ddot{l}_{gg}(\beta_0, g_0; O)[h_1^*, h_1^*]\} = 0,
$$

implies that

$$
\begin{aligned}
\Sigma &= P\{\dot{l}_\beta(\beta_0, g_0; O) - \dot{l}_g(\beta_0, g_0; O)[h_1^*]\}^{\otimes 2} \\
&= P\{-\ddot{l}_{\beta\beta}(\beta_0, g_0; O) + \ddot{l}_{g\beta}(\beta_0, g_0; O)[h_1^*] + \ddot{l}_{\beta g}(\beta_0, g_0; O)[h_1^*] - \ddot{l}_{gg}(\beta_0, g_0; O)[h_1^*, h_1^*]\} \\
&= P\{-\ddot{l}_{\beta\beta}(\beta_0, g_0; O) + \ddot{l}_{g\beta}(\beta_0, g_0; O)[h_1^*]\}.
\end{aligned}
$$

The fact that

$$
P\big\{\Psi(\beta_0, \rho_0, g_0; Z)\dot{l}_g^\top(\beta_0, g_0; O)[h_2^*]\big\} = P\big\{\Psi(\beta_0, \rho_0, g_0; Z)\mathbb{E}\{\dot{l}_g^\top(\beta_0, g_0; O)[h_2^*]|Z\}\big\} = 0
$$

implies that

$$
\begin{aligned}
Q =& P\big\{\Psi(\beta_0,\rho_0,g_0;Z) - \dot{l}_g(\beta_0,g_0;O)[h_2^*]\big\}^{\otimes 2}\\
=& P\{\Psi(\beta_0,\rho_0,g_0;Z)^{\otimes 2} + (\dot{l}_g(\beta_0,g_0;O)[h_2^*])^{\otimes 2}\}\\
=& P\big\{\Psi(\beta_0,\rho_0,g_0;Z)^{\otimes 2} - \ddot{l}_{gg}(\beta_0,g_0;O)[h_2^*,h_2^*]\big\}\\
=& P\big\{\Psi(\beta_0,\rho_0,g_0;Z)^{\otimes 2} - \{\dot{\Psi}_g(\beta_0,\rho_0,g_0;Z)[h_2^*]\}^\top\big\}.
\end{aligned}
$$

Since

$$
\begin{aligned}
& P\{\ddot{l}_{g\beta}(\beta_0,g_0;O)[h_2^*]\} = P\{\ddot{l}_{\beta g}(\beta_0,g_0;O)[h_2^*]\}^\top = P\{\ddot{l}_{gg}(\beta_0,g_0;O)[h_1^*,h_2^*]\}^\top\\
=& P\{\ddot{l}_{gg}(\beta_0,g_0;O)[h_2^*,h_1^*]\} = P\{\dot{\Psi}_g(\beta_0,\rho_0,g_0;O)[h_1^*]\},
\end{aligned}
$$

we have

$$
P\big\{\ddot{l}_{g\beta}(\beta_0,g_0;O)[h_2^*] - \{\dot{\Psi}_\beta(\beta_0,\rho_0,g_0;O)\}^\top\big\} = -B^\top.
$$

Recall that $A = -P\big\{\dot{\Psi}_\rho(\beta_0,\rho_0,g_0;Z)\big\}$. Therefore, it follows from (S3.12), (S3.13), and (S3.5) that

$$
\begin{pmatrix} \Sigma & B & 0\\ -B^\top & Q & A\\ 0 & -A^\top & 0 \end{pmatrix}
\begin{pmatrix} n^{1/2}(\hat{\beta}_{\mathrm{au}}-\beta_0)\\ n^{1/2}\hat{\nu}_{\mathrm{au}}\\ n^{1/2}(\hat{\rho}_{\mathrm{au}}-\rho_0) \end{pmatrix}
=
\begin{pmatrix} n^{1/2}\mathbb{P}_n\{\iota(\beta_0,g_0;O)\}\\ n^{1/2}\mathbb{P}_n\{\chi(\beta_0,\rho_0,g_0;O)\}\\ 0 \end{pmatrix} + o_p(1),
$$

or equivalently,

$$\begin{pmatrix} \Sigma & \bar{B} \\ -\bar{B}^\top & \bar{Q} \end{pmatrix} \begin{pmatrix} n^{1/2}(\hat{\beta}_{\text{au}} - \beta_0) \\ n^{1/2}\hat{\nu}_{\text{au}} \\ n^{1/2}(\hat{\rho}_{\text{au}} - \rho_0) \end{pmatrix} = \begin{pmatrix} n^{1/2}\mathbb{P}_n\{\iota(\beta_0, g_0; O)\} \\ n^{1/2}\mathbb{P}_n\{\chi(\beta_0, \rho_0, g_0; O)\} \\ 0 \end{pmatrix} + o_p(1), \qquad \text{(S3.17)}$$

where

$$\bar{B} = (B,\ 0), \quad \bar{Q} = \begin{pmatrix} Q & A \\ -A^\top & 0 \end{pmatrix}.$$

Under the condition (D2)(iii) and the fact $Q = P\{\Psi(\beta_0, \rho_0, g_0; Z)^{\otimes 2} + (\dot{l}_g(\beta_0, g_0; O)[h_2^*])^{\otimes 2}\}$, the matrix Q is positive definite. Suppose that Σ and $A^\top Q^{-1} A$ are positive definite. It follows from (S3.17) that

$$n^{1/2}(\hat{\beta}_{\text{au}} - \beta_0) = (I_{p\times p} \ \ 0_{p\times(J+1)}) \begin{pmatrix} \Sigma & \bar{B} \\ -\bar{B}^\top & \bar{Q} \end{pmatrix}^{-1} \begin{pmatrix} n^{1/2}\mathbb{P}_n\{\iota(\beta_0, g_0; O)\} \\ n^{1/2}\mathbb{P}_n\{\chi(\beta_0, \rho_0, g_0; O)\} \\ 0 \end{pmatrix} + o_p(1),$$

where

$$\begin{pmatrix} \Sigma & \bar{B} \\ -\bar{B}^\top & \bar{Q} \end{pmatrix}^{-1} = \begin{pmatrix} \Gamma & -\Gamma\bar{B}\bar{Q}^{-1} \\ \bar{Q}^{-1}\bar{B}^\top\Gamma & \bar{Q}^{-1} - \bar{Q}^{-1}\bar{B}^\top\Gamma\bar{B}\bar{Q}^{-1} \end{pmatrix},$$

with $\Gamma = (\Sigma + \bar{B}\bar{Q}^{-1}\bar{B}^\top)^{-1}$. Straightforward algebra yields

$$\bar{Q}^{-1} = \begin{pmatrix} Q & A \\ -A^\top & 0 \end{pmatrix}^{-1} = \begin{pmatrix} Q^{-1} - Q^{-1}A(A^\top Q^{-1}A)^{-1}A^\top Q^{-1} & -Q^{-1}A(A^\top Q^{-1}A)^{-1} \\ (A^\top Q^{-1}A)^{-1}A^\top Q^{-1} & (A^\top Q^{-1}A)^{-1} \end{pmatrix}$$

and

$$\bar{B}\bar{Q}^{-1} \begin{pmatrix} 0 & A \\ -A^\top & 0 \end{pmatrix} \{\bar{Q}^{-1}\}^\top \bar{B}^\top = 0.$$

Therefore, the asymptotic variance of $n^{1/2}(\hat{\beta}_{\rm au} - \beta_0)$ is

$$\begin{aligned} & \begin{pmatrix} \Gamma & -\Gamma\bar{B}\bar{Q}^{-1} \end{pmatrix} \begin{pmatrix} \Sigma & 0 & 0 \\ 0 & Q & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \Gamma \\ -\bar{Q}^{-1}\bar{B}^\top\Gamma \end{pmatrix} \\ = & \begin{pmatrix} \Gamma & -\Gamma\bar{B}\bar{Q}^{-1} \end{pmatrix} \begin{pmatrix} \Sigma & 0 \\ 0 & \bar{Q} \end{pmatrix} \begin{pmatrix} \Gamma \\ -\bar{Q}^{-1}\bar{B}^\top\Gamma \end{pmatrix} \\ = & \Gamma\Sigma\Gamma + \Gamma\bar{B}\bar{Q}^{-1}\bar{B}^\top\Gamma \\ = & \Gamma. \end{aligned}$$

S3.4 Proof of Theorem S1(iii)

Theorem 1 shows that the asymptotic variance of the maximum conditional likelihood estimator $\tilde{\beta}$ is Σ^{-1}. The asymptotic variance of $\hat{\beta}_{\rm au}$ is $\Gamma = (\Sigma + \bar{B}\bar{Q}^{-1}\bar{B}^\top)^{-1}$. Since

$$\Gamma^{-1} - \Sigma = \bar{B}\bar{Q}^{-1}\bar{B}^\top$$

$$
\begin{aligned}
&= (B,\ 0)\begin{pmatrix} Q & A \\ -A^\top & 0 \end{pmatrix}^{-1}\begin{pmatrix} B^\top \\ 0 \end{pmatrix} \\
&= B\{Q^{-1} - Q^{-1}A(A^\top Q^{-1}A)^{-1}A^\top Q^{-1}\}B^\top,
\end{aligned}
$$

and $\bar{B}\bar{Q}^{-1}\bar{B}^\top$ is a nonnegative definition matrix, we conclude that $\Gamma^{-1} - \Sigma$ is nonnegative definite, which implies that $\hat{\beta}_{\mathrm{au}}$ is asymptotically more efficient than the maximum conditional likelihood estimator $\tilde{\beta}$.

S4 The relationship between $(\hat{\beta}_{\mathrm{au}}, \hat{\rho}_{\mathrm{au}})$ and $(\tilde{\beta}, \tilde{\rho})$

We shall derive first-order linear approximations of $(\tilde{\beta}, \tilde{\rho})$ and $(\hat{\beta}_{\mathrm{au}}, \hat{\rho}_{\mathrm{au}})$, and then compare the approximations.

For $(\tilde{\beta}, \tilde{\rho}, \tilde{g})$, it satisfies

$$
\begin{aligned}
&\mathbb{P}_n\Big\{\dot{l}_\beta(\tilde{\beta}, \tilde{g}; O)\Big\} = 0, \\
&\mathbb{P}_n\Big\{\dot{l}_g(\tilde{\beta}, \tilde{g}; O)[h]\Big\} = 0, \quad \text{for any } \ h \in L_2(P), \\
&\mathbb{P}_n\Big\{1_{1\times J}\Psi(\tilde{\beta}, \tilde{\rho}, \tilde{g}; Z)\Big\} = 0.
\end{aligned}
$$

It follows from Lemma 1 that

$$
\begin{aligned}
&\sqrt{n}(\mathbb{P}_n - P)\Big\{\dot{l}_\beta(\tilde{\beta}, \tilde{g}; O) - \dot{l}_\beta(\beta_0, g_0; O)\Big\} = o_p(1), \\
&\sqrt{n}(\mathbb{P}_n - P)\Big\{\dot{l}_g(\tilde{\beta}, \tilde{g}; O)[h] - \dot{l}_g(\beta_0, g_0; O)[h]\Big\} = o_p(1),
\end{aligned}
$$

$$\sqrt{n}(\mathbb{P}_n - P)\Big\{1_{1\times J}\Psi(\tilde{\beta},\tilde{\rho},\tilde{g};Z) - 1_{1\times J}\Psi(\beta_0,\rho_0,g_0;Z)\Big\} = o_p(1).$$

By Lemma 2 and the consistency of $(\tilde{\beta},\tilde{\rho},\tilde{g})$, we have

$$\begin{aligned}
-\mathbb{P}_n\Big\{\dot{l}_\beta(\beta_0,g_0;O)\Big\} &= P\Big\{\dot{l}_\beta(\tilde{\beta},\tilde{g};O) - \dot{l}_\beta(\beta_0,g_0;O)\Big\} + o_p(n^{-1/2}) \\
&= P\Big\{\ddot{l}_{\beta\beta}(\beta_0,g_0;O)\Big\}(\tilde{\beta}-\beta_0) + P\Big\{\ddot{l}_{\beta g}(\beta_0,g_0;O)[\tilde{g}-g_0]\Big\} \\
&\quad + o_p(n^{-1/2}), \qquad \text{(S4.18)} \\
-\mathbb{P}_n\Big\{1_{1\times J}\Psi(\beta_0,\rho_0,g_0;Z)\Big\} &= P\Big\{1_{1\times J}\Psi(\tilde{\beta},\tilde{\rho},\tilde{g};Z) - 1_{1\times J}\Psi(\beta_0,\rho_0,g_0;Z)\Big\} + o_p(n^{-1/2}) \\
&= P\Big\{1_{1\times J}\{\dot{\Psi}_\beta(\beta_0,\rho_0,g_0;Z)\}^\top\Big\}(\tilde{\beta}-\beta_0) \\
&\quad + P\Big\{1_{1\times J}\dot{\Psi}_\rho(\beta_0,\rho_0,g_0;Z)\Big\}(\tilde{\rho}-\rho_0) \\
&\quad + P\Big\{1_{1\times J}\dot{\Psi}_g(\beta_0,\rho_0,g_0;Z)[\tilde{g}-g_0]\Big\} + o_p(n^{-1/2}), \qquad \text{(S4.19)} \\
-\mathbb{P}_n\Big\{\dot{l}_g(\beta_0,g_0;O)[h]\Big\} &= P\Big\{\dot{l}_g(\tilde{\beta},\tilde{g};O)[h] - \dot{l}_g(\beta_0,g_0;O)[h]\Big\} + o_p(n^{-1/2}), \\
&= P\Big\{\ddot{l}_{g\beta}(\beta_0,g_0;O)[h]\Big\}(\tilde{\beta}-\beta_0) \\
&\quad + P\Big\{\ddot{l}_{gg}(\beta_0,g_0;O)[h,\tilde{g}-g_0]\Big\} + o_p(n^{-1/2}), \qquad \text{(S4.20)}
\end{aligned}$$

for any $h \in L_2(P)$. The definitions of h_1^* and h_2^* imply that

$$\begin{aligned}
P\Big\{\ddot{l}_{gg}(\beta_0,g_0;O)[h_1^*,\tilde{g}-g_0]\Big\} &= P\Big\{\ddot{l}_{\beta g}(\beta_0,g_0;O)[\tilde{g}-g_0]\Big\}, \\
P\Big\{\ddot{l}_{gg}(\beta_0,g_0;O)[h_2^*,\tilde{g}-g_0]\Big\} &= P\Big\{\dot{\Psi}_g(\beta_0,\rho_0,g_0;Z)[\tilde{g}-g_0]\Big\}.
\end{aligned}$$

Then, replacing h with h_1^* in (S4.20) and using (S4.18), we have

$$-\mathbb{P}_n\Big\{\iota(\beta_0, g_0; O)\Big\} = P\Big\{\ddot{l}_{\beta\beta}(\beta_0, g_0; O) - \ddot{l}_{g\beta}(\beta_0, g_0; O)[h_1^*]\Big\}(\tilde{\beta} - \beta_0) + o_p(n^{-1/2}). \tag{S4.21}$$

Replacing h by h_2^* in (S4.20) and using (S4.19), we have

$$\begin{aligned}-\mathbb{P}_n\Big\{1_{1\times J}\chi(\beta_0, \rho_0, g_0; O)\Big\} =& P\Big\{1_{1\times J}\dot{\Psi}_\beta^\top(\beta_0, \rho_0, g_0; Z) - 1_{1\times J}\ddot{l}_{g\beta}(\beta_0, g_0; O)[h_2^*]\Big\}(\tilde{\beta} - \beta_0) \\ &+ P\Big\{1_{1\times J}\dot{\Psi}_\rho(\beta_0, \rho_0, g_0; Z)\Big\}(\tilde{\rho} - \rho_0) + o_p(n^{-1/2}).\end{aligned} \tag{S4.22}$$

We have shown in the proof of Theorem S1(ii) that

$$\Sigma \equiv P\big\{\iota(\beta_0, g_0; O)^{\otimes 2}\big\} = P\big\{-\ddot{l}_{\beta\beta}(\beta_0, g_0; O) + \ddot{l}_{g\beta}(\beta_0, g_0; O)[h_1^*]\big\}.$$

Thus, combining (S4.21) with (S4.22) yields

$$\begin{pmatrix} \Sigma & 0_{p\times 1} \\ -1_{1\times J}B^\top & 1_{1\times J}A \end{pmatrix} \begin{pmatrix} \tilde{\beta} - \beta_0 \\ \tilde{\rho} - \rho_0 \end{pmatrix} = \mathbb{P}_n \begin{pmatrix} \iota(\beta_0, g_0; O) \\ 1_{1\times J}\chi(\beta_0, \rho_0, g_0; O) \end{pmatrix} + o_p(n^{-1/2}).$$

This implies that

$$\begin{pmatrix} \tilde{\beta} - \beta_0 \\ \tilde{\rho} - \rho_0 \end{pmatrix} = \begin{pmatrix} \Sigma & 0_{p\times 1} \\ -1_{1\times J}B^\top & 1_{1\times J}A \end{pmatrix}^{-1} \begin{pmatrix} I_{p\times p} & 0_{p\times J} & 0_{p\times 1} \\ 0_{1\times p} & 1_{1\times J} & 0 \end{pmatrix} \begin{pmatrix} \mathbb{P}_n\{\iota(\beta_0, g_0; O)\} \\ \mathbb{P}_n\{\chi(\beta_0, \rho_0, g_0; O)\} \\ 0 \end{pmatrix} + o_p(n^{-1/2})$$

$$
\begin{aligned}
&= \begin{pmatrix} \Sigma & 0_{p\times 1} \\ -1_{1\times J}B^\top & 1_{1\times J}A \end{pmatrix}^{-1} \begin{pmatrix} I_{p\times p} & 0_{p\times J} & 0_{p\times 1} \\ 0_{1\times p} & 1_{1\times J} & 0 \end{pmatrix} \begin{pmatrix} \Sigma & B & 0 \\ -B^\top & Q & A \\ 0 & -A^\top & 0 \end{pmatrix} \\
&\quad \cdot \begin{pmatrix} \Sigma & B & 0 \\ -B^\top & Q & A \\ 0 & -A^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{P}_n\{\iota(\beta_0, g_0; O)\} \\ \mathbb{P}_n\{\chi(\beta_0, \rho_0, g_0; O)\} \\ 0 \end{pmatrix} + o_p(n^{-1/2}) \\
&= \begin{pmatrix} I_{p\times p} & \Sigma^{-1}B & 0_{p\times 1} \\ 0_{1\times p} & (1_{1\times J}A)^{-1}1_{1\times J}(B^\top\Sigma^{-1}B + Q) & 1 \end{pmatrix} \\
&\quad \cdot \begin{pmatrix} \Sigma & B & 0 \\ -B^\top & Q & A \\ 0 & -A^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{P}_n\{\iota(\beta_0, g_0; O)\} \\ \mathbb{P}_n\{\chi(\beta_0, \rho_0, g_0; O)\} \\ 0 \end{pmatrix} + o_p(n^{-1/2}), \qquad \text{(S4.23)}
\end{aligned}
$$

where we have used

$$
\begin{aligned}
&\begin{pmatrix} \Sigma & 0_{p\times 1} \\ -1_{1\times J}B^\top & 1_{1\times J}A \end{pmatrix}^{-1} \begin{pmatrix} I_{p\times p} & 0_{p\times J} & 0_{p\times 1} \\ 0_{1\times p} & 1_{1\times J} & 0 \end{pmatrix} \begin{pmatrix} \Sigma & B & 0 \\ -B^\top & Q & A \\ 0 & -A^\top & 0 \end{pmatrix} \\
&= \begin{pmatrix} \Sigma & 0_{p\times 1} \\ -1_{1\times J}B^\top & 1_{1\times J}A \end{pmatrix}^{-1} \begin{pmatrix} \Sigma & B & 0 \\ -1_{1\times J}B^\top & 1_{1\times J}Q & 1_{1\times J}A \end{pmatrix} \\
&= \begin{pmatrix} \Sigma^{-1} & 0_{p\times 1} \\ (1_{1\times J}A)^{-1}1_{1\times J}B^\top\Sigma^{-1} & (1_{1\times J}A)^{-1} \end{pmatrix} \begin{pmatrix} \Sigma & B & 0 \\ -1_{1\times J}B^\top & 1_{1\times J}Q & 1_{1\times J}A \end{pmatrix}
\end{aligned}
$$

$$= \begin{pmatrix} I_{p\times p} & \Sigma^{-1}B & 0_{p\times 1} \\ 0_{1\times p} & (1_{1\times J}A)^{-1}1_{1\times J}(B^\top\Sigma^{-1}B + Q) & 1 \end{pmatrix}.$$

For $(\hat\beta_{\mathrm{au}}, \hat\rho_{\mathrm{au}}, \hat g_{\mathrm{au}})$, we have shown in the proof of Theorem S1(ii) that

$$\begin{pmatrix} \Sigma & B & 0 \\ -B^\top & Q & A \\ 0 & -A^\top & 0 \end{pmatrix} \begin{pmatrix} \hat\beta_{\mathrm{au}} - \beta_0 \\ \hat\nu_{\mathrm{au}} \\ \hat\rho_{\mathrm{au}} - \rho_0 \end{pmatrix} = \begin{pmatrix} \mathbb{P}_n\{\iota(\beta_0, g_0; O)\} \\ \mathbb{P}_n\{\chi(\beta_0, \rho_0, g_0; O)\} \\ 0 \end{pmatrix} + o_p(n^{-1/2}),$$

which implies that

$$\begin{pmatrix} \hat\beta_{\mathrm{au}} - \beta_0 \\ \hat\rho_{\mathrm{au}} - \rho_0 \end{pmatrix} = \begin{pmatrix} I_{p\times p} & 0_{p\times J} & 0_{p\times 1} \\ 0_{1\times p} & 0_{1\times J} & 1 \end{pmatrix} \begin{pmatrix} \hat\beta_{\mathrm{au}} - \beta_0 \\ \hat\nu_{\mathrm{au}} \\ \hat\rho_{\mathrm{au}} - \rho_0 \end{pmatrix}$$

$$= \begin{pmatrix} I_{p\times p} & 0_{p\times J} & 0_{p\times 1} \\ 0_{1\times p} & 0_{1\times J} & 1 \end{pmatrix} \begin{pmatrix} \Sigma & B & 0 \\ -B^\top & Q & A \\ 0 & -A^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{P}_n\{\iota(\beta_0, g_0; O)\} \\ \mathbb{P}_n\{\chi(\beta_0, \rho_0, g_0; O)\} \\ 0 \end{pmatrix} + o_p(n^{-1/2}). \tag{S4.24}$$

By comparing the approximations in (S4.23) and (S4.24), we have

$$\begin{pmatrix} \hat\beta_{\mathrm{au}} \\ \hat\rho_{\mathrm{au}} \end{pmatrix} = \begin{pmatrix} \tilde\beta \\ \tilde\rho \end{pmatrix} + \begin{pmatrix} 0_{p\times p} & -\Sigma^{-1}B & 0_{p\times 1} \\ 0_{1\times p} & -(1_{1\times J}A)^{-1}1_{1\times J}(B^\top\Sigma^{-1}B + Q) & 0 \end{pmatrix}$$

$$\times \begin{pmatrix} \Sigma & B & 0 \\ -B^\top & Q & A \\ 0 & -A^\top & 0 \end{pmatrix}^{-1} \cdot \begin{pmatrix} \mathbb{P}_n\{\iota(\beta_0, g_0; O)\} \\ \mathbb{P}_n\{\chi(\beta_0, \rho_0, g_0; O)\} \\ 0 \end{pmatrix} + o_p(n^{-1/2}). \quad \text{(S4.25)}$$

S5 Proof of Theorem 2

S5.1 Proof of Theorem 2 (i)

Recall that $\tilde{\rho}$ is the solution to $\sum_{i=1}^n \sum_{j=1}^J \Psi_j(\tilde{\beta}, \rho, \tilde{g}; Z_i) = 0$, where

$$\Psi_j(\beta, \rho, g; Z) = I(Z \in \Omega_j)\Big[\exp\Big\{-\int I(s \le \log t_j^* - Z^\top \beta)\exp\{g(s)\}ds\Big\} - \zeta_j^{1/\rho}\Big].$$

For fixed β and g, $\Psi_j(\beta, \rho, g; Z)$ is a continuous and monotone function of $\rho \in (0, \infty)$, therefore, the solution to

$$\sum_{i=1}^n \sum_{j=1}^J \Psi_j(\tilde{\beta}, \rho, \tilde{g}; Z_i) = 0,$$

namely $\tilde{\rho}$, is unique and converges to ρ_0, which is the solution of $\sum_{j=1}^J E\{\Psi_j(\tilde{\beta}, \rho, \tilde{g}; Z)\} = 0$.

Recall that the one-step estimator is defined as

$$\begin{pmatrix} \hat{\beta}_{\text{os}} \\ \hat{\rho}_{\text{os}} \end{pmatrix} = \begin{pmatrix} \tilde{\beta} \\ \tilde{\rho} \end{pmatrix} + \begin{pmatrix} 0_{p\times p} & -\hat{\Sigma}^{-1}\hat{B} & 0_{p\times 1} \\ 0_{1\times p} & -(1_{1\times J}\hat{A})^{-1}1_{1\times J}(\hat{B}^\top\hat{\Sigma}^{-1}\hat{B} + \hat{Q}) & 0 \end{pmatrix}$$

$$\times \begin{pmatrix} \hat{\Sigma} & \hat{B} & 0 \\ -\hat{B}^\top & \hat{Q} & \hat{A} \\ 0 & -\hat{A}^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} n^{-1}\sum_{i=1}^n \iota(\tilde{\beta}, \tilde{g}; O_i) \\ n^{-1}\sum_{i=1}^n \chi(\tilde{\beta}, \tilde{\rho}, \tilde{g}; O_i) \\ 0 \end{pmatrix}. \qquad \text{(S5.26)}$$

It follows from the strong law of large numbers that $\hat{\Sigma}, \hat{B}, \hat{Q}$, and $\hat{A}$ converge to Σ, B, Q, and A, respectively, and $n^{-1}\sum_{i=1}^n \iota(\tilde{\beta}, \tilde{g}; O_i)$ and $n^{-1}\sum_{i=1}^n \chi(\tilde{\beta}, \tilde{\rho}, \tilde{g}; O_i)$ converge to zeros. Therefore, the one-step estimator $(\hat{\beta}_{\text{os}}^\top, \hat{\rho}_{\text{os}})^\top$ has the same limit as $(\tilde{\beta}^\top, \tilde{\rho})^\top$. As the limit of the latter is $(\beta_0^\top, \rho_0)^\top$, the one-step estimator $(\hat{\beta}_{\text{os}}^\top, \hat{\rho}_{\text{os}})^\top$ also converges to $(\beta_0^\top, \rho_0)^\top$ almost surely.

S5.2 Proof of Theorem 2 (ii)

By comparing the equalities in (S4.25) and (S5.26), we arrive at $(\hat{\beta}_{\text{os}}^\top, \hat{\rho}_{\text{os}})^\top = (\hat{\beta}_{\text{au}}^\top, \hat{\rho}_{\text{au}})^\top + o_p(n^{-1/2})$ under the conditions in Theorem S1. This implies that $\hat{\beta}_{\text{os}}$ has the same limiting distribution as $\hat{\beta}_{\text{au}}$. Hence, $n^{1/2}(\hat{\beta}_{\text{os}} - \beta_0)$ converges to a normal distribution with mean zero and covariance $\{\Sigma + BQ^{-1}B^\top - BQ^{-1}A(A^\top Q^{-1}A)^{-1}A^\top Q^{-1}B^\top\}^{-1}$, and $\hat{\beta}_{\text{os}}$ is asymptotically more efficient than the maximum conditional likelihood estimator $\tilde{\beta}$.

S5.3 Proof of Theorem 2 (iii)

Since the limiting distributions of $\hat{\beta}_{\text{os}}$ and $\hat{\beta}_{\text{au}}$ are the same, the proof of Theorem 2 (iii) follows from the proof of Theorem S1 (iii). More Specifically, the asymptotic variance of $\hat{\beta}_{\text{os}}$ is $\Gamma = (\Sigma + \bar{B}\bar{Q}^{-1}\bar{B}^\top)^{-1}$, where Σ^{-1} is the asymptotic variance of $\tilde{\beta}$.

Since

$$
\begin{aligned}
\Gamma^{-1} - \Sigma &= \bar{B}\bar{Q}^{-1}\bar{B}^\top \\
&= (B,\ 0)\begin{pmatrix} Q & A \\ -A^\top & 0 \end{pmatrix}^{-1}\begin{pmatrix} B^\top \\ 0 \end{pmatrix} \\
&= B\{Q^{-1} - Q^{-1}A(A^\top Q^{-1}A)^{-1}A^\top Q^{-1}\}B^\top,
\end{aligned}
$$

and $\bar{B}\bar{Q}^{-1}\bar{B}^\top$ is a nonnegative definition matrix, we conclude that $\Gamma^{-1} - \Sigma$ is nonnegative definite, which implies that $\hat{\beta}_{\text{os}}$ is asymptotically more efficient than $\tilde{\beta}$.

S6 Additional simulation studies

S6.1 Simulation results under the homogeneous scenario

In Section 4 of the main paper, we consider two simulation settings: a heterogeneous scenario ($\rho_0 = 0.9$ and unknown) and a homogeneous scenario ($\rho_0 = 1$ and known). We have reported the simulation results for the heterogeneous scenario in Table 1 of the main paper. Those for the homogeneous scenario are presented in Table S1. Our general findings from Table S1 are the same as those from Table 1 of the main paper.

S6.2 Comparison of our estimator without auxiliary information and existing methods

We conducted simulations to compare our proposed estimator (Proposed for short) in the absence of auxiliary aggregate information with Zeng and Lin (2007)'s estimator (ZL) and Lin and Chen (2013)'s estimator (LC) under the simulation configurations (Cases I–IV) in the main paper. Our method involves a bandwidth $\sigma = csn^{-1/3}$, where $c > 0$ and s is the sample standard deviation of $\log(X) - Z^\top\beta$ (with β replaced by an initial parameter value) among all subjects.

To investigate the sensitivity of the tuning parameter σ, we allow the constant c in σ to vary from 0.1 to 3.5 with step length 0.2. Based on 1000 simulated samples of size $n = 100$, we calculate the empirical biases, standard deviations, and computation time of the three estimators (Proposed, ZL, LC) when data were generated from Cases I–IV with different c. The corresponding simulation results are displayed in Figures S1–S5, respectively. In each figure, the plots in the upper panel display empirical biases and standard deviations of the three estimators (Proposed, ZL, LC) for β_1, β_2, and β_3 for different c values, and the plot in the lower panel displays the computation times of the three methods for different c.

From Figures S1–S5, we draw the following two conclusions. First, in terms of computational cost, the proposed estimator is slightly inferior to and comparable with LC, and both of them are much better than ZL. Second, the proposed estimator is insensitive to the tuning parameter c or σ, and it has smallest standard deviations among the three estimators under comparison and negligible biases. Therefore, the proposed estimator is computationally efficient, insensitive to tuning parameters and shows advantages in estimation accuracy.

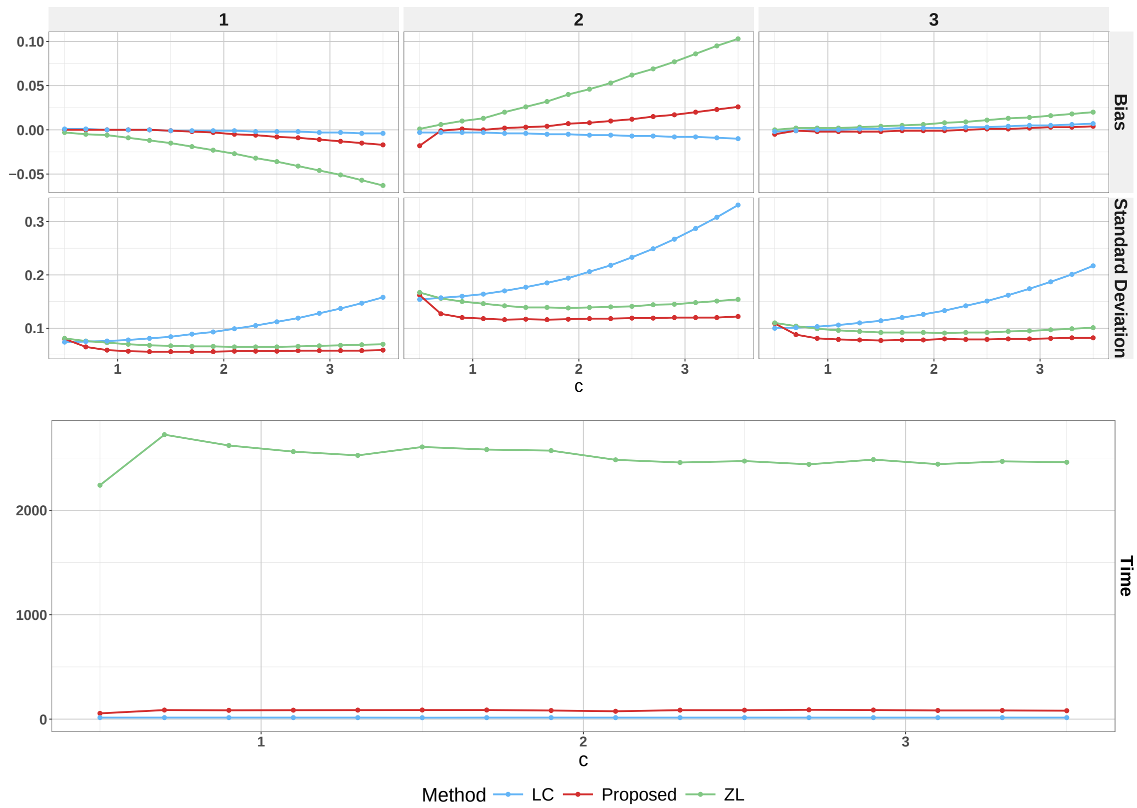

Figure S1: Simulation results when data were generated from Case I, where the error term follows a norm distribution. The three plots (from left to right) in the upper panel display empirical biases and standard deviations of the estimators for β_1, β_2, and β_3, respectively, when c varies. The lower panel displays computation times of the three methods.

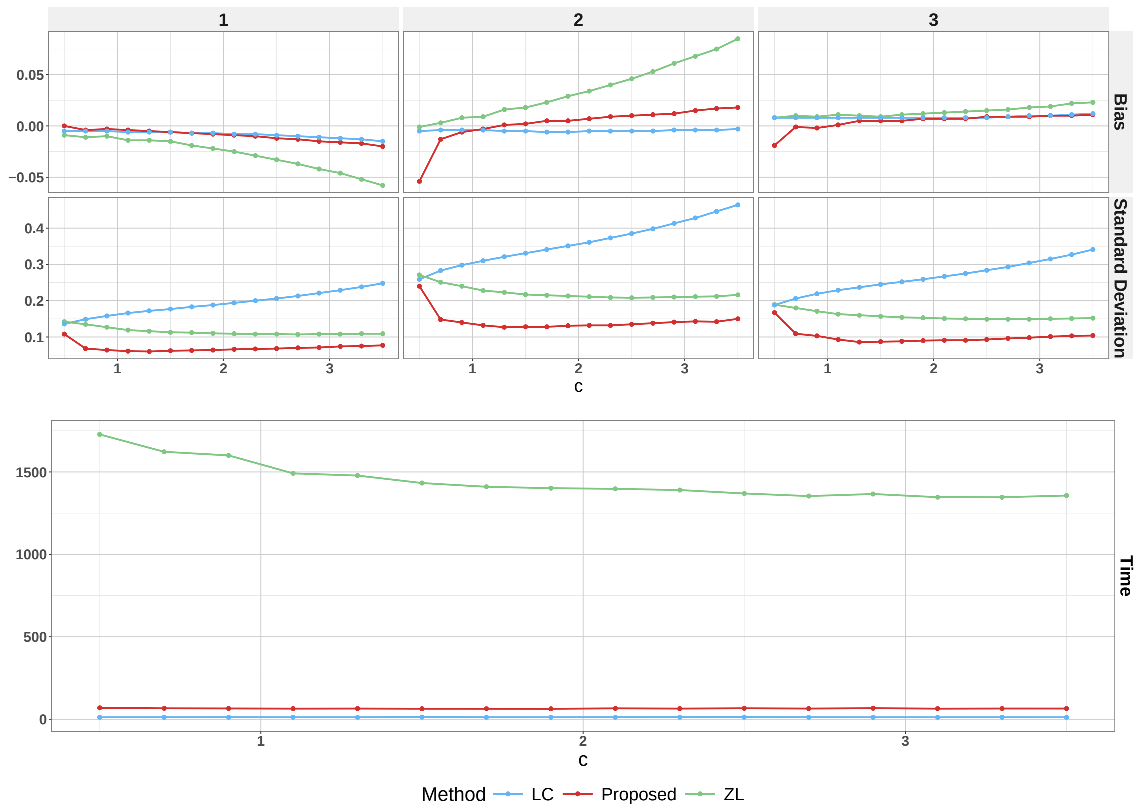

Figure S2: Simulation results when data were generated from Case II, where the error term follows a generalized extreme value distribution. The three plots (from left to right) in the upper panel display empirical biases and standard deviations of the estimators for β_1, β_2, and β_3, respectively, when c varies. The lower panel displays computation times of the three methods.

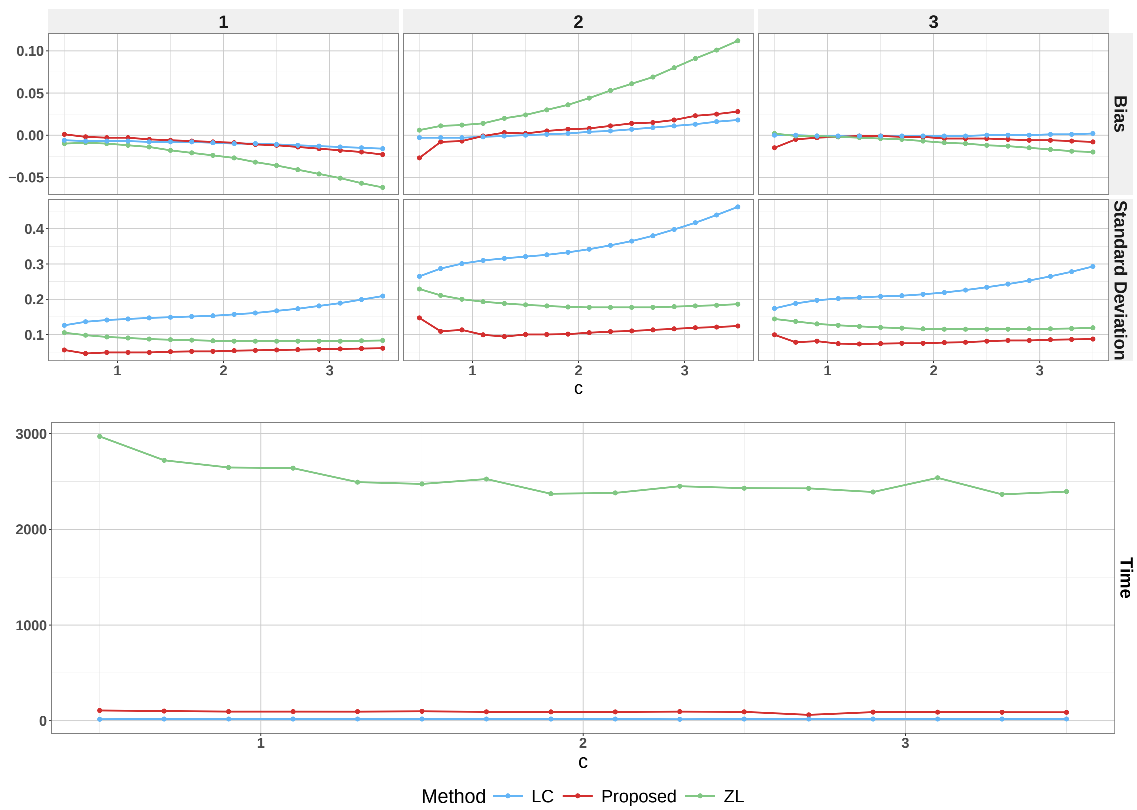

Figure S3: Empirical results when data were generated from Case III, where the error term follows a Weibull distribution. The three plots (from left to right) in the upper panel display empirical biases and standard deviations of the estimators for β_1, β_2, and β_3, respectively, when c varies. The lower panel displays computation times of the three methods.

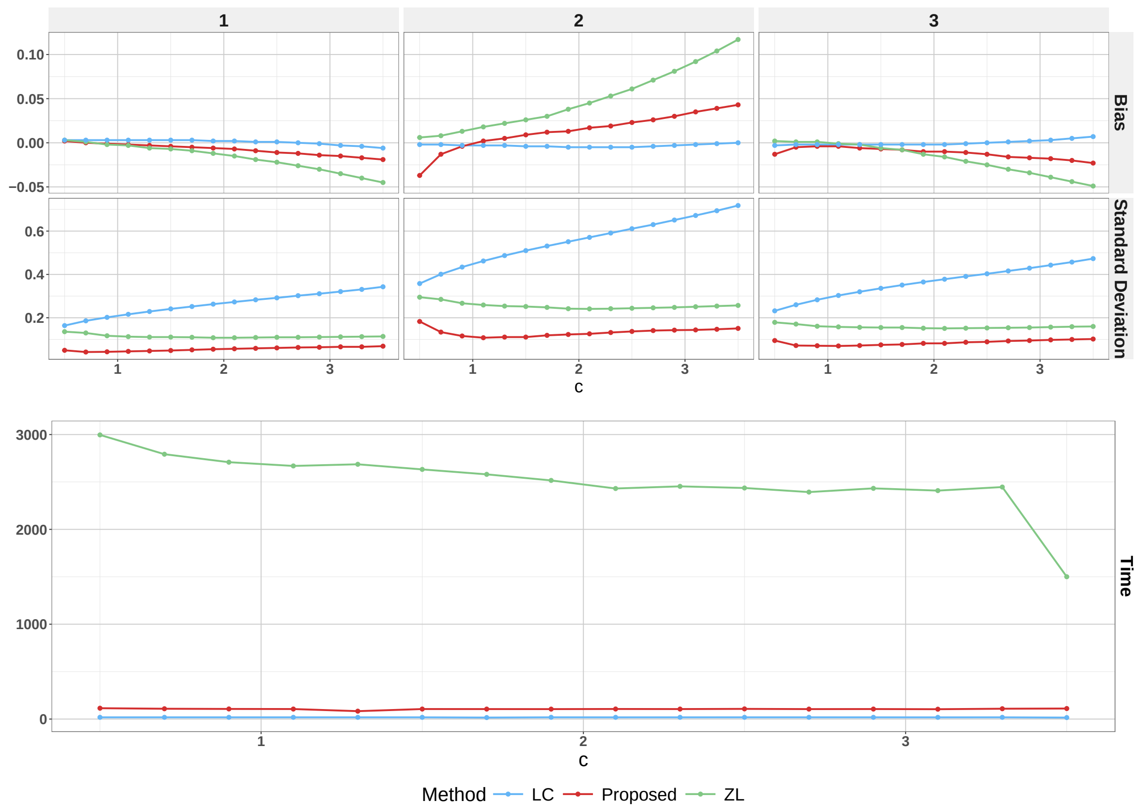

Figure S4: Simulation results when data were generated from Case IV, where the error term follows a log-normal distribution. The three plots (from left to right) in the upper panel display empirical biases and standard deviations of the estimators for β_1, β_2, and β_3, respectively, when c varies. The lower panel displays computation times of the three methods.

Table S1: Simulation results under the homogeneous scenario

		β_1				β_2				β_3			
Case	Est	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
I	$\hat{\beta}_{\text{os}}$	-5	60	59	93.4	8	105	99	91.7	-3	87	82	92.6
	$\tilde{\beta}$	-3	63	63	93.8	5	128	120	92.5	-3	92	92	94.1
	$\hat{\beta}_{\text{L}}$	-1	60	53	91.1	-1	120	110	92.5	-4	87	75	89.7
	$\hat{\beta}_{\text{G1}}$	-11	60	48	88.8	1	98	83	89.2	14	91	59	79.4
	$\hat{\beta}_{\text{G0}}$	-10	63	51	89.4	9	121	105	89.8	18	99	67	81.2
II	$\hat{\beta}_{\text{os}}$	-5	65	61	92.2	8	109	104	92.5	6	85	85	93.6
	$\tilde{\beta}$	-4	67	66	91.9	5	130	129	93.2	6	92	96	94.6
	$\hat{\beta}_{\text{L}}$	-1	86	74	90.2	-7	175	158	92.1	3	122	106	90.4
	$\hat{\beta}_{\text{G1}}$	-7	65	53	89.0	-12	105	90	89.3	5	89	68	85.7
	$\hat{\beta}_{\text{G0}}$	-7	71	56	89.5	2	130	115	91.6	16	102	77	87.5
III	$\hat{\beta}_{\text{os}}$	-4	48	47	91.9	7	90	86	91.7	0	69	69	93.9
	$\tilde{\beta}$	-3	48	49	92.2	7	96	95	92.2	1	71	74	94.5
	$\hat{\beta}_{\text{L}}$	-3	74	60	88.4	-7	148	129	90.7	2	102	87	89.2
	$\hat{\beta}_{\text{G1}}$	-7	61	43	88.2	-7	107	72	80.8	-2	90	55	82.0
	$\hat{\beta}_{\text{G0}}$	-6	60	45	88.3	3	104	91	92.4	12	85	61	86.6
IV	$\hat{\beta}_{\text{os}}$	-4	45	45	91.5	7	86	95	94.9	-1	75	76	93.4
	$\tilde{\beta}$	-3	45	47	92.2	8	90	102	94.6	3	74	78	94.1
	$\hat{\beta}_{\text{L}}$	0	90	79	93.3	-1	178	168	94.8	-1	133	118	92.3
	$\hat{\beta}_{\text{G1}}$	-7	62	43	86.4	-3	125	74	80.6	-6	109	63	81.5
	$\hat{\beta}_{\text{G0}}$	-9	66	45	88.3	6	109	92	92.0	15	93	66	86.7
V	$\hat{\beta}_{\text{os}}$	-4	39	43	96.6	21	88	75	96.5	9	47	61	98.0
	$\tilde{\beta}$	-3	39	44	96.7	6	95	95	97.4	4	47	67	97.8
	$\hat{\beta}_{\text{L}}$	3	117	126	96.9	-10	231	243	98.0	1	161	172	96.4
	$\hat{\beta}_{\text{G1}}$	-12	117	44	89.6	-29	136	67	86.7	3	161	57	83.7
	$\hat{\beta}_{\text{G0}}$	-22	153	50	91.5	1	158	94	93.9	39	223	71	88.1

Est, estimator; $\hat{\beta}_{\text{os}}$, the proposed one-step estimator incorporating auxiliary information; $\tilde{\beta}$, the maximum conditional likelihood estimator without auxiliary information; $\hat{\beta}_{\text{L}}$, the weighted log-rank estimator with unit weight; $\hat{\beta}_{\text{G1}}$, the generalized method of moments estimator incorporating the auxiliary information (Sheng et al., 2020); $\hat{\beta}_{\text{G0}}$, the generalized method of moments estimator without the auxiliary information (Sheng et al., 2020); Bias, empirical bias ($\times 1000$); SD, empirical standard deviation ($\times 1000$); SE, estimated standard error ($\times 1000$); CP, empirical coverage probability.

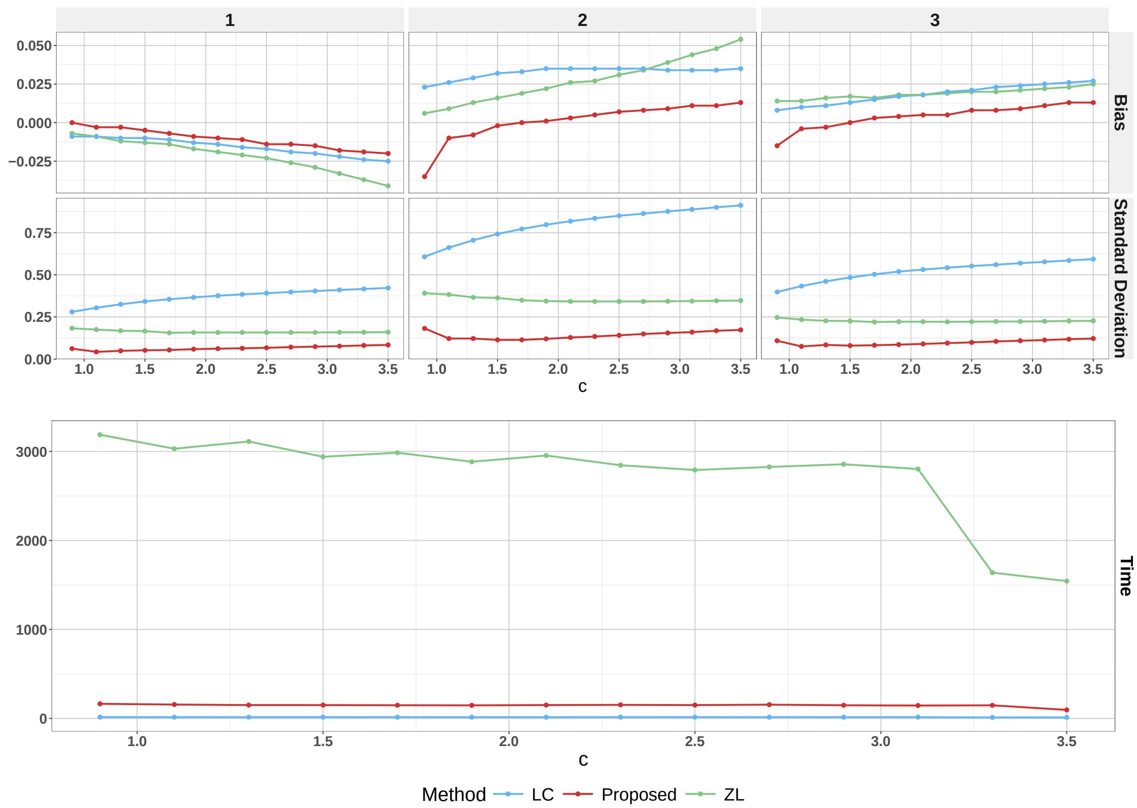

Figure S5: Simulation results when data were generated from Case V, where the error term follows a Chi-squared distribution. The three plots (from left to right) in the upper panel display empirical biases and standard deviations of the estimators for β_1, β_2, and β_3, respectively, when c varies. The lower panel displays computation times of the three methods.

S6.3 Bias and variance trade-off of $\hat{\beta}_{os}$

To study the bias and variance trade-off of $\hat{\beta}_{os}$ in the presence/absence of population heterogeneity, we focus on Case I and allow ρ varying from 0.1 to 3.5 with step length 0.1. The empirical biases (Bias) and standard deviations (SD) of $\hat{\beta}_{\text{os,homo}}$ and $\hat{\beta}_{\text{os,hete}}$ for different values of ρ are displayed in Figure S6. We see that when ρ is equal to or near 1, $\hat{\beta}_{\text{os,homo}}$ and $\hat{\beta}_{\text{os,hete}}$ have nearly the same biases and SDs. As ρ goes away from 1, $\hat{\beta}_{\text{os,hete}}$ has an almost unchanged bias, however the bias of $\hat{\beta}_{\text{os,homo}}$ gets larger. In terms of SD, in the estimation of β_2 and β_3, $\hat{\beta}_{\text{os,hete}}$ outperforms $\hat{\beta}_{\text{os,homo}}$ when $\rho < 1$, and $\hat{\beta}_{\text{os,homo}}$ outperforms $\hat{\beta}_{\text{os,hete}}$ when $\rho > 1$. For the estimation of β_1, $\hat{\beta}_{\text{os,hete}}$ outperforms $\hat{\beta}_{\text{os,homo}}$ when $\rho > 1$, and they are comparable when $\rho < 1$. Thus the two estimators $\hat{\beta}_{\text{os,hete}}$ and $\hat{\beta}_{\text{os,homo}}$ have comparable estimation performance. When $\rho \neq 1$, the heterogeneous model is correct, $\hat{\beta}_{\text{os,hete}}$ should be asymptotically unbiased, and $\hat{\beta}_{\text{os,homo}}$ should be asymptotically biased. However, $\hat{\beta}_{\text{os,homo}}$ may have less fluctuation because the homogeneous model involves one less unknown parameter (ρ) than the heterogeneous model.

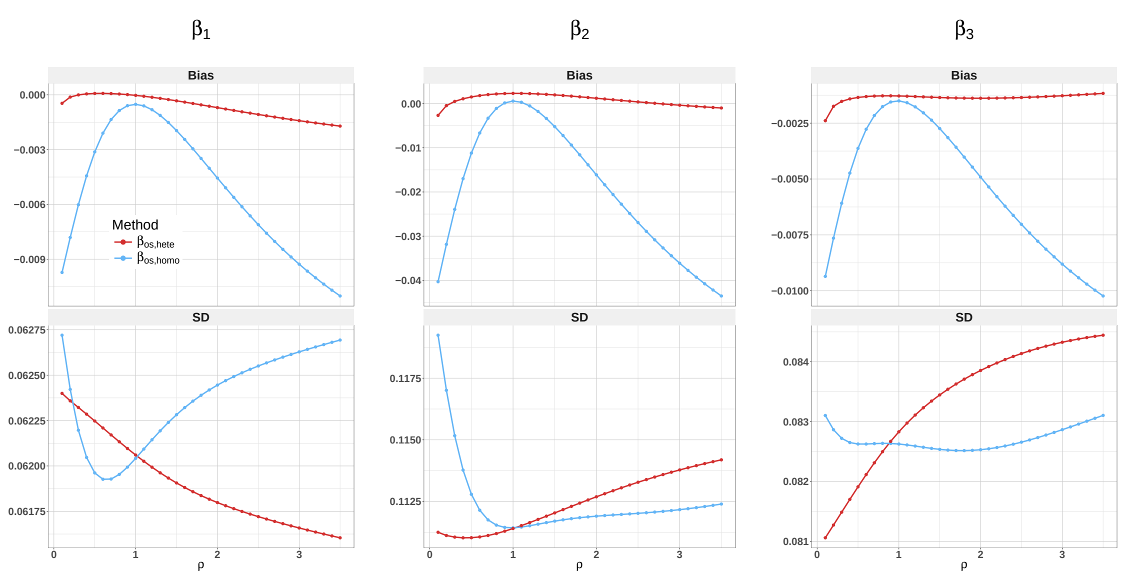

Figure S6: Simulated empirical bias (Bias) and standard deviation (SD) of $\beta_{\text{os,homo}}$ and $\beta_{\text{os,hete}}$ when ρ varies from 0.1 to 3.5.

S6.4 Simulations for different values of ρ

In this subsection, we consider simulation studies for ρ =0.5, 0.8, 1.1, and 1.2. Simulation results for the estimator with auxiliary information ($\hat{\beta}_{\text{os, hete}}$) and that without auxiliary information ($\tilde{\beta}$) under these different scenarios are presented in Table S2. From Table S2, we can see that the proposed estimator $\hat{\beta}_{\text{os, hete}}$ maintains a desirable finite-sample performance, with little variation in Bias, SD, SE, and CP, indicating that our method is robust to different values of ρ. These results are quite similar to those obtained when $\rho = 0.9$.

S6.5 Robustness of our method to the nonconstancy of ρ

To demonstrate the robustness of our method to the non-constancy of ρ, we conduct simulations in the simulation settings where the auxiliary information ζ_1 and ζ_2 are calculated from ρ_1 and ρ_2 respectively ($\rho_1 \neq \rho_2$). The values of ρ_j, ζ_j, $j = 1, 2$, and simulation results under Cases I–V are shown in Table S3, where the sample size is set to $n = 100$ and the experiments are repeated 1000 times. Compared to $\tilde{\beta}$, the initial estimator without auxiliary information, the impact of the varying form of ρ on $\hat{\beta}_{\text{os,hete}}$ is mainly reflected in a slight increase in bias. However, the bias remains small and does not affect the conclusion of the unbiasedness of $\hat{\beta}_{\text{os,hete}}$. The finite sample performance of $\hat{\beta}_{\text{os,hete}}$ in terms of SD, SE, and CP remains as good as that under a constant ρ. Thus, the proposed estimation method incorporating auxiliary information does have certain robustness to the nonconstancy of ρ.

Table S2: Simulation results for β under different heterogeneous scenarios.

		β_1				β_2				β_3			
Est	ρ	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
Case I													
$\tilde{\beta}$		0	63	67	94.5	2	125	129	95.3	-1	84	97	96.4
$\hat{\beta}_{\mathrm{os,hete}}$	0.5	0	63	65	94.0	2	111	108	94.0	-1	82	93	96.2
$\hat{\beta}_{\mathrm{os,hete}}$	0.8	0	62	65	94.1	2	111	109	94.1	-1	83	93	96.1
$\hat{\beta}_{\mathrm{os,hete}}$	1.1	0	62	65	94.1	3	112	109	94.5	-1	83	94	96.0
$\hat{\beta}_{\mathrm{os,hete}}$	1.2	0	62	65	93.9	3	112	109	94.3	-1	83	94	96.1
Case II													
$\tilde{\beta}$		-4	64	72	93.5	4	131	139	94.9	4	90	103	95.9
$\hat{\beta}_{\mathrm{os,hete}}$	0.5	-3	64	68	93.1	11	112	114	94.6	4	90	101	95.6
$\hat{\beta}_{\mathrm{os,hete}}$	0.8	-3	63	68	93.3	8	114	116	94.5	4	90	102	95.7
$\hat{\beta}_{\mathrm{os,hete}}$	1.1	-4	63	67	93.5	5	116	117	94.6	4	90	102	95.8
$\hat{\beta}_{\mathrm{os,hete}}$	1.2	-5	63	67	93.5	4	117	118	94.4	4	90	102	95.9
Case III													
$\tilde{\beta}$		-3	48	52	93.6	4	86	100	96.3	1	67	78	95.4
$\hat{\beta}_{\mathrm{os,hete}}$	0.5	-3	47	51	93.7	6	83	91	95.7	1	67	77	96.0
$\hat{\beta}_{\mathrm{os,hete}}$	0.8	-3	47	51	93.7	6	83	90	95.6	1	67	77	96.0
$\hat{\beta}_{\mathrm{os,hete}}$	1.1	-3	47	51	93.5	6	83	91	95.7	1	67	77	95.6
$\hat{\beta}_{\mathrm{os,hete}}$	1.2	-3	47	51	93.5	6	83	91	95.8	1	67	77	95.6
Case IV													
$\tilde{\beta}$		-1	41	47	92.6	5	77	98	96.7	0	62	75	95.4
$\hat{\beta}_{\mathrm{os,hete}}$	0.5	-2	41	46	92.3	2	76	91	96.0	-1	63	74	95.2
$\hat{\beta}_{\mathrm{os,hete}}$	0.8	-2	41	46	92.3	3	76	90	96.0	-1	63	74	95.2
$\hat{\beta}_{\mathrm{os,hete}}$	1.1	-2	41	46	92.3	4	76	90	96.0	-1	63	74	95.1
$\hat{\beta}_{\mathrm{os,hete}}$	1.2	-2	41	46	92.3	4	76	90	96.0	-1	62	74	95.1
Case V													
$\tilde{\beta}$		-2	38	38	94.1	10	112	82	94.5	4	54	59	95.8
$\hat{\beta}_{\mathrm{os,hete}}$	0.5	-1	39	38	93.7	17	100	70	94.7	5	54	56	95.8
$\hat{\beta}_{\mathrm{os,hete}}$	0.8	-1	38	38	93.7	17	101	70	94.7	4	54	57	95.8
$\hat{\beta}_{\mathrm{os,hete}}$	1.1	-2	38	38	93.7	16	101	70	94.7	4	54	57	95.9
$\hat{\beta}_{\mathrm{os,hete}}$	1.2	-2	38	37	93.6	16	101	70	94.7	4	54	57	95.9

Bias, SD, and SE represents the corresponding one multiplies 1000.

Table S3: Simulation results for β under heterogeneous scenarios with varying ρ.

				β_1				β_2				β_3			
Case	(ρ_1, ρ_2)	(ζ_1, ζ_2)	Est	Bias	SD	SE	CP	Bias	SD	SE	CP	Bias	SD	SE	CP
I	(1.51,1.31)	(0.58,0.78)	$\tilde{\beta}$	0	64	65	93.7	2	127	125	93.6	-1	86	94	95.7
			$\hat{\beta}_{\text{os,hete}}$	3	64	63	92.9	15	114	107	93.2	0	85	92	94.6
II	(1.55,1.38)	(0.4,0.4)	$\tilde{\beta}$	-4	64	72	93.5	4	131	139	94.9	4	90	103	95.9
			$\hat{\beta}_{\text{os,hete}}$	-1	63	67	93.6	15	118	120	94.8	3	90	102	95.9
III	(1.78,1.52)	(0.55,0.55)	$\tilde{\beta}$	-3	48	52	93.6	4	86	100	96.3	1	67	78	95.4
			$\hat{\beta}_{\text{os,hete}}$	-1	48	51	93.0	14	83	91	95.8	1	68	77	95.4
IV	(1.10,1.39)	(0.65,0.45)	$\tilde{\beta}$	-1	41	47	92.6	5	77	98	96.7	0	62	75	95.4
			$\hat{\beta}_{\text{os,hete}}$	-6	40	46	92.3	-11	77	90	96.0	0	63	74	94.9
V	(1.14,1.79)	(0.6,0.6)	$\tilde{\beta}$	-2	38	38	94.1	10	112	82	94.5	4	54	59	95.8
			$\hat{\beta}_{\text{os,hete}}$	-7	37	37	94.2	-16	105	71	94.2	-2	55	57	95.7

Bias, SD, and SE represents the corresponding one multiplies 1000.

S7 Heterogeneity in covariate distribution and uncertainty in auxiliary information

In the main paper, we assume that the internal and external data may have different error distributions, and assume that the auxiliary information comes from a large database where the sample size of the external study, denoted by m, is much larger than that of the internal study. In this case, the variability in the external aggregated information can be ignored when compared with the variability of $\hat{\beta}_{\text{au}}$. In practice, the internal and external data may have different covariate distributions, and the auxiliary information may be derived from the external study whose sample size is of the same order as n (Sheng et al., 2021). Hence, it is desired to account for heterogeneity in covariate distribution and uncertainty in the auxiliary information under the proposed estimation procedure.

Let $d^*(z)$ and $d(z)$ denote the density functions of Z in the external and internal studies, respectively. To account for heterogeneity in covariate distributions, we postulated a semiparametric density ratio model on the density functions of covariate Z in external and internal studies

$$d^*(z) = \exp(\alpha_1 + \alpha_2^\top w) d(z), \tag{S7.27}$$

where w is a q-dimensional sub-vector of z and accounts for the predictor distributional heterogeneity. The parameter vector (α_1, α_2) characterizes the degree of heterogeneity and $\alpha_2 = 0$ indicates that the homogeneity assumption holds. We study the impact of uncertainty in the auxiliary information on the efficiency gain, and write the function of auxiliary information depending on ζ. We investigate the large-sample properties of our estimators as $n \to \infty$ and n/m converges to a positive constant κ. Denote by $\hat{\zeta}$ the root-m consistent estimator of the population parameter ζ_0 using the external data and assume that $\hat{\zeta}$ is asymptotically normal. Suppose that as $m \to \infty$, $\sqrt{m}(\hat{\zeta} - \zeta_0)$ converges in distribution to a mean zero normal distribution with covariance matrix Σ_ζ.

Let $\alpha = (\alpha_1, \alpha_2^\top)^\top$. All constraints can be summarized as

$$E\{\Psi(\beta, \alpha, \zeta, g; Z)\} = 0,$$

where $\Psi(\beta, \alpha, \zeta, g; Z) = (\exp(\alpha_1 + \alpha_2^\top W) - 1, \exp(\alpha_1 + \alpha_2^\top W)\Psi(\beta, \zeta, g; Z)^\top)^\top$. The first constraint $E\{\exp(\alpha_1 + \alpha_2^\top w)\} = 1$ reflects the fact that $d^*(z)$ and $d(z)$ are proper density functions. Define $D_\zeta = E\{\partial\Psi(\beta_0, \alpha_0, \zeta, g_0; Z)/\partial\zeta|_{\zeta=\zeta_0}\}$.

After replacing ζ with its estimate $\hat{\zeta}$, one can estimate (β, α, g) by

$$(\hat{\beta}_{\alpha\zeta}, \hat{\alpha}_{\zeta}, \hat{\nu}_{\alpha\zeta}, \hat{g}_{\alpha\zeta}) = \arg\max_{\beta,\alpha,g} \min_{\nu} \Big\{ \ell_n(\beta, g) - \sum_{i=1}^{n} \log(1 + \nu^{\top}\Psi(\beta, \alpha, \hat{\zeta}, g; Z_i)) \Big\}.$$

In Theorem S2, we establish the asymptotic normality of $\hat{\beta}_{\alpha\zeta}$ and derive a one-step estimator of (β, α) that has the same limiting distribution as $(\hat{\beta}_{\alpha\zeta}, \hat{\alpha}_{\zeta})$.

An initial estimator $\tilde{\alpha}_{\zeta}$ for α is needed in the implementation of this method. We choose $\tilde{\alpha}_{\zeta}$ to be the solution to the estimating equations $\mathbb{P}_n\{\exp(\alpha_1 + \alpha_2^{\top} W) - 1\} = 0$ and $\mathbb{P}_n\{W 1_{1\times J} \exp(\alpha_1 + \alpha_2^{\top} W)\Psi(\tilde{\beta}, \hat{\zeta}, \tilde{g}; Z)\} = 0$, or equivalently $\mathbb{P}_n\{F\Psi(\tilde{\beta}, \alpha, \hat{\zeta}, \tilde{g}; Z_i)\} = 0$, where $F = \mathrm{diag}(1, W 1_{1\times J})$.

Theorem S2. *Assume that the regularity conditions (C1)–(C8) are satisfied and that model* (S7.27) *is correctly specified. Let Q_α, B_α and A_α be those defined in* (S7.39)-(S7.41)*, respectively. Also suppose that the matrices $E\{\Psi(\beta_0, \alpha_0, \zeta_0, g_0; Z)^{\otimes 2}\}$ and Σ are positive definite, the matrix $C = Q_\alpha + \kappa D_\zeta \Sigma_\zeta D_\zeta^{\top}$ is nonsingular, and the vector A_α is nonzero. Then as $n \to \infty$, we have (i) $n^{1/2}(\hat{\beta}_{\alpha\zeta} - \beta_0)$ converges in distribution to a mean zero normal distribution with covariance $\Gamma_{\alpha\zeta} = \{\Sigma + B_\alpha C^{-1} B_\alpha^{\top} - B_\alpha C^{-1} A_\alpha (A_\alpha^{\top} C^{-1} A_\alpha)^{-1} A_\alpha^{\top} C^{-1} B_\alpha^{\top}\}^{-1}$. (ii) The following estimator*

$$\begin{pmatrix} \tilde{\beta} \\ \tilde{\alpha}_{\zeta} \end{pmatrix} + \begin{pmatrix} 0_{p\times p} & -\hat{\Sigma}^{-1}\hat{B}_\alpha & 0_{p\times(q+1)} \\ 0_{1\times p} & -(F\hat{A}_\alpha)^{-1} F(\hat{B}_\alpha^{\top}\hat{\Sigma}_\alpha^{-1}\hat{B}_\alpha + \hat{Q}_\alpha + \kappa \hat{D}_\zeta \hat{\Sigma}_\zeta \hat{D}_\zeta^{\top}) & 0_{1\times(q+1)} \end{pmatrix}$$

$$\times \begin{pmatrix} \hat{\Sigma} & \hat{B}_\alpha & 0 \\ -\hat{B}_\alpha^\top & \hat{Q}_\alpha + \kappa \hat{D}_\zeta \hat{\Sigma}_\zeta \hat{D}_\zeta^\top & \hat{A}_\alpha \\ 0 & -\hat{A}_\alpha^\top & 0 \end{pmatrix}^{-1} \times \begin{pmatrix} \mathbb{P}_n\{\iota(\tilde{\beta}, \tilde{g}; O)\} \\ \mathbb{P}_n\{\chi(\tilde{\beta}, \tilde{\alpha}, \hat{\zeta}, \tilde{g}; O)\} \\ 0 \end{pmatrix}$$

with $\chi(\tilde{\beta}, \tilde{\alpha}, \hat{\zeta}, \tilde{g}; O) = \Psi(\tilde{\beta}, \tilde{\alpha}, \hat{\zeta}, \tilde{g}; Z) - \dot{l}_g(\tilde{\beta}, \tilde{g}; O)[\tilde{h}_2^*]$ *is a one-step estimator of* (β, α) *that has the same limiting distribution as* $(\hat{\beta}_{\alpha\zeta}, \hat{\alpha}_\zeta)$.

Theorem S2 indicates that the asymptotic variance of $\hat{\beta}_{\alpha\zeta}$ is $\Gamma_{\alpha\zeta} = \{\Sigma + B_\alpha C^{-1} B_\alpha^\top - B_\alpha C^{-1} A_\alpha (A_\alpha^\top C^{-1} A_\alpha)^{-1} A_\alpha^\top C^{-1} B_\alpha^\top\}^{-1}$. We have shown in Theorem 1 in the main paper that Σ^{-1} is the asymptotic variance of the maximum conditional likelihood estimator $\tilde{\beta}$ that does not utilize the auxiliary information. Because

$$\Gamma_{\alpha\zeta}^{-1} - \Sigma = B_\alpha C^{-1}\{C - A_\alpha (A_\alpha^\top C^{-1} A_\alpha)^{-1} A_\alpha^\top\} C^{-1} B_\alpha^\top$$

and it can be verified that $C - A_\alpha (A_\alpha^\top C^{-1} A_\alpha) A_\alpha$ is nonnegative definite, we conclude that $\hat{\beta}_{\alpha\zeta}$ is asymptotically more efficient than $\tilde{\beta}$. Let $\hat{\beta}_\alpha$ denote $\hat{\beta}_{\alpha\zeta}$ in the case of $\kappa = 0$. Thus, $\hat{\beta}_\alpha$ is an estimator of β when the heterogeneity in covariate distribution exists and the uncertainty in auxiliary information can be ignored. In this situation, $C = Q_\alpha$ and the asymptotic variance of $\hat{\beta}_\alpha$ is

$$\Gamma_{\alpha\zeta,1} = \{\Sigma + B_\alpha Q_\alpha^{-1} B_\alpha^\top - B_\alpha Q_\alpha^{-1} A_\alpha (A_\alpha^\top Q_\alpha^{-1} A_\alpha)^{-1} A_\alpha^\top Q_\alpha^{-1} B_\alpha^\top\}^{-1}.$$

It can be verified that

$$\Gamma_{\alpha\zeta} = \left\{ \Sigma + \begin{pmatrix} B_\alpha \\ 0 \end{pmatrix} \begin{pmatrix} C & A_\alpha \\ A_\alpha^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} B_\alpha \\ 0 \end{pmatrix}^\top \right\}^{-1},$$

and

$$\Gamma_{\alpha\zeta,1} = \left\{ \Sigma + \begin{pmatrix} B_\alpha \\ 0 \end{pmatrix} \begin{pmatrix} Q_\alpha & A_\alpha \\ A_\alpha^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} B_\alpha \\ 0 \end{pmatrix}^\top \right\}^{-1}.$$

Because $C - Q_\alpha = \kappa D_\zeta \Sigma_\zeta D_\zeta^\top \geq 0$ (meaning that $C - Q_\alpha$ is nonnegative definite), it follows that

$$\begin{pmatrix} C & A_\alpha \\ A_\alpha^\top & 0 \end{pmatrix} - \begin{pmatrix} Q_\alpha & A_\alpha \\ A_\alpha^\top & 0 \end{pmatrix} \geq 0 \Longrightarrow \begin{pmatrix} Q_\alpha & A_\alpha \\ A_\alpha^\top & 0 \end{pmatrix}^{-1} - \begin{pmatrix} C & A_\alpha \\ A_\alpha^\top & 0 \end{pmatrix}^{-1} \geq 0.$$

Consequently, we have

$$\Gamma_{\alpha\zeta,1} \leq \Gamma_{\alpha\zeta},$$

which implies that $\hat{\beta}_{\alpha\zeta}$ is asymptotically less efficient than $\hat{\beta}_\alpha$.

Proof of Theorem S2: We prove result (i) first. For any $h \in L_2(P)$, the estimator $(\hat{\beta}_{\alpha\zeta}, \hat{\nu}_{\alpha\zeta}, \hat{\alpha}_\zeta, \hat{g}_{\alpha\zeta})$ satisfies

$$\mathbb{P}_n \left\{ \dot{l}_\beta(\hat{\beta}_{\alpha\zeta}, \hat{g}_{\alpha\zeta}; O) - \frac{\hat{\nu}_{\alpha\zeta}^\top \dot{\Psi}_\beta(\hat{\beta}_{\alpha\zeta}, \hat{\alpha}_\zeta, \hat{\zeta}, \hat{g}_{\alpha\zeta}; Z)}{1 + \hat{\nu}_{\alpha\zeta}^\top \Psi(\hat{\beta}_{\alpha\zeta}, \hat{\alpha}_\zeta, \hat{\zeta}, \hat{g}_{\alpha\zeta}; Z)} \right\} = 0,$$

$$\mathbb{P}_n\Big\{\dot{l}_g(\hat{\beta}_{\alpha\zeta},\hat{g}_{\alpha\zeta};O)[h]-\frac{\hat{\nu}_{\alpha\zeta}^\top\dot{\Psi}_g(\hat{\beta}_{\alpha\zeta},\hat{\alpha}_\zeta,\hat{\zeta},\hat{g}_{\alpha\zeta};Z)[h]}{1+\hat{\nu}_{\alpha\zeta}^\top\Psi(\hat{\beta}_{\alpha\zeta},\hat{\alpha}_\zeta,\hat{\zeta},\hat{g}_{\alpha\zeta};Z)}\Big\}=0,$$
$$\mathbb{P}_n\Big\{\frac{\Psi(\hat{\beta}_{\alpha\zeta},\hat{\alpha}_\zeta,\hat{\zeta},\hat{g}_{\alpha\zeta};Z)}{1+\hat{\nu}_{\alpha\zeta}^\top\Psi(\hat{\beta}_{\alpha\zeta},\hat{\alpha}_\zeta,\hat{\zeta},\hat{g}_{\alpha\zeta};Z)}\Big\}=0,$$
$$\mathbb{P}_n\Big\{\frac{\hat{\nu}_{\alpha\zeta}^\top\dot{\Psi}_\alpha(\hat{\beta}_{\alpha\zeta},\hat{\alpha}_\zeta,\hat{\zeta},\hat{g}_{\alpha\zeta};Z)}{1+\hat{\nu}_{\alpha\zeta}^\top\Psi(\hat{\beta}_{\alpha\zeta},\hat{\alpha}_\zeta,\hat{\zeta},\hat{g}_{\alpha\zeta};Z)}\Big\}=0.$$

By Lemmas 1 and 2 and the consistency of $(\hat{\beta}_{\alpha\zeta},\hat{\alpha}_\zeta,\hat{\zeta},\hat{g}_{\alpha\zeta})$, we can show that $\hat{\nu}_{\alpha\zeta}=o_p(1)$. Further, using first-order Taylor series approximations, we can rewrite the above equations as

$$\begin{aligned}-\mathbb{P}_n\big\{\dot{l}_\beta(\beta_0,g_0;O)\big\}=&P\big\{\ddot{l}_{\beta\beta}(\beta_0,g_0;O)(\hat{\beta}_{\alpha\zeta}-\beta_0)+\ddot{l}_{\beta g}(\beta_0,g_0;O)[\hat{g}_{\alpha\zeta}-g_0]\big\}\\&-P\big\{\dot{\Psi}_\beta(\beta_0,\alpha_0,\zeta_0,g_0;O)\big\}\hat{\nu}_{\alpha\zeta}+o_p(n^{-1/2}),\end{aligned}\tag{S7.28}$$

$$\begin{aligned}-\mathbb{P}_n\{\dot{l}_g(\beta_0,g_0;O)[h]\}=&P\big\{\ddot{l}_{g\beta}(\beta_0,g_0;O)[h]\big\}(\hat{\beta}_{\alpha\zeta}-\beta_0)+P\big\{\ddot{l}_{gg}(\beta_0,g_0;O)[h,\hat{g}_{\alpha\zeta}-g_0]\big\}\\&-P\big\{\dot{\Psi}_g(\beta_0,\alpha_0,\zeta_0,g_0;Z)[h]\big\}^\top\hat{\nu}_{\alpha\zeta}+o_p(n^{-1/2}),\end{aligned}\tag{S7.29}$$

$$\begin{aligned}-\mathbb{P}_n\{\Psi(\beta_0,\alpha_0,\hat{\zeta},g_0;Z)\}=&P\big\{\dot{\Psi}_\beta(\beta_0,\alpha_0,\zeta_0,g_0;Z)\big\}^\top(\hat{\beta}_{\alpha\zeta}-\beta_0)\\&+P\{\dot{\Psi}_g(\beta_0,\alpha_0,\zeta_0,g_0;Z)[\hat{g}_{\alpha\zeta}-g_0]\}\\&+P\big\{\dot{\Psi}_\alpha(\beta_0,\alpha_0,\zeta_0,g_0;Z)\big\}(\hat{\alpha}_\zeta-\alpha_0)\\&-P\big\{\Psi(\beta_0,\alpha_0,\hat{\zeta},g_0;Z)^{\otimes 2}\big\}\hat{\nu}_{\alpha\zeta}+o_p(n^{-1/2}),\end{aligned}\tag{S7.30}$$

$$0=P\big\{\dot{\Psi}_\alpha(\beta_0,\alpha_0,\zeta_0,g_0;Z)\big\}^\top\hat{\nu}_{\alpha\zeta}+o_p(n^{-1/2}).\tag{S7.31}$$

The least favorable direction h_1^* is defined as before, but h_2^* is defined as

$$P\big\{\ddot{l}_{gg}(\beta_0, g_0; O)[h_2^*, h]\big\} = P\big\{\dot{\Psi}_g(\beta_0, \alpha_0, \zeta_0, g_0; Z)[h]\big\} \qquad \text{(S7.32)}$$

for all $h \in \mathcal{H}$, where $h_{2j}^* \in \mathcal{H}, j = 1, \ldots, J$. When the aggregate information is the subgroup survival probabilities, one obvious choice of h_2^* is $h_2^*(t, \beta) = (h_{21}^*(t, \beta), \ldots, h_{2J}^*(t, \beta))^\top$ with

$$h_{2j}^*(t, \beta_0) = \frac{\mathbb{E}[\exp(\alpha_1 + \alpha_2^\top W) I(Z \in \Omega_j) I(t \le \log t_j^* - Z^\top \beta_0)\{1 - F(\log t_j^* - Z^\top \beta_0)\}]}{\mathbb{E}\{Y(t, \beta_0)\}},$$

$j = 1, \ldots, J$.

It follows from (S3.6) and (S7.32) that

$$P\big\{\ddot{l}_{\beta g}(\beta_0, g_0; O)[\hat{g}_{\alpha\zeta} - g_0] - \ddot{l}_{gg}(\beta_0, g_0; O)[h_1^*, \hat{g}_{\alpha\zeta} - g_0]\big\} = 0, \qquad \text{(S7.33)}$$

$$P\big\{\ddot{l}_{gg}(\beta_0, g_0; O)[h_2^*, \hat{g}_{\alpha\zeta} - g_0]\big\} = P\big\{\dot{\Psi}_g(\beta_0, \alpha_0, \zeta_0, g_0; Z)[\hat{g}_{\alpha\zeta} - g_0]\big\}. \qquad \text{(S7.34)}$$

And it follows from (S7.29) that

$$\begin{aligned} -\mathbb{P}_n\{\dot{l}_g(\beta_0, g_0; O)[h_1^*]\} =& P\big\{\ddot{l}_{g\beta}(\beta_0, g_0; O)[h_1^*]\big\}(\hat{\beta}_{\alpha\zeta} - \beta_0) + P\big\{\ddot{l}_{gg}(\beta_0, g_0; O)[h_1^*, \hat{g}_{\alpha\zeta} - g_0]\big\} \\ &- P\big\{\dot{\Psi}_g(\beta_0, \alpha_0, \zeta_0, g_0; Z)[h_1^*]\big\}^\top \hat{\nu}_{\alpha\zeta} + o_p(n^{-1/2}), \qquad \text{(S7.35)} \\ -\mathbb{P}_n\{\dot{l}_g(\beta_0, g_0; O)[h_2^*]\} =& P\big\{\ddot{l}_{g\beta}(\beta_0, g_0; O)[h_2^*]\big\}(\hat{\beta}_{\alpha\zeta} - \beta_0) + P\big\{\ddot{l}_{gg}(\beta_0, g_0; O)[h_2^*, \hat{g}_{\alpha\zeta} - g_0]\big\} \\ &- P\big\{\dot{\Psi}_g(\beta_0, \alpha_0, \zeta_0, g_0; Z)[h_2^*]\big\}^\top \hat{\nu}_{\alpha\zeta} + o_p(n^{-1/2}). \qquad \text{(S7.36)} \end{aligned}$$

Subtracting (S7.28) from (S7.35) and using (S7.33), we have

$$\begin{aligned}\mathbb{P}_n\big\{\iota(\beta_0,g_0;O)\big\} =&P\big\{-\ddot{l}_{\beta\beta}(\beta_0,g_0;O)+\ddot{l}_{g\beta}(\beta_0,g_0;O)[h_1^*]\big\}(\hat{\beta}_{\alpha\zeta}-\beta_0)\\ &+P\big\{\dot{\Psi}_\beta(\beta_0,\alpha_0,\zeta_0,g_0;Z)-\{\dot{\Psi}_g(\beta_0,\alpha_0,\zeta_0,g_0;Z)[h_1^*]\}^\top\big\}\hat{\nu}_{\alpha\zeta}+o_p(n^{-1/2}),\end{aligned}\tag{S7.37}$$

where $\iota(\beta_0, g_0; O)$ is defined in (S2.1). Subtracting (S7.30) from (S7.36) and using (S7.34), we have

$$\begin{aligned}\mathbb{P}_n\big\{\chi(\beta_0,\alpha_0,\hat{\zeta},g_0;O)\big\} =&P\big\{\ddot{l}_{g\beta}(\beta_0,g_0;O)[h_2^*]-\{\dot{\Psi}_\beta(\beta_0,\alpha_0,\zeta_0,g_0;Z)\}^\top\big\}(\hat{\beta}_{\alpha\zeta}-\beta_0)\\ &+P\big\{\Psi(\beta_0,\alpha_0,\hat{\zeta},g_0;Z)^{\otimes 2}-\{\dot{\Psi}_g(\beta_0,\alpha_0,\zeta_0,g_0;Z)[h_2^*]\}^\top\big\}\hat{\nu}_{\alpha\zeta}\\ &-P\big\{\dot{\Psi}_\alpha(\beta_0,\alpha_0,\zeta_0,g_0;Z)\big\}(\hat{\alpha}_\zeta-\alpha_0)+o_p(n^{-1/2}),\end{aligned}\tag{S7.38}$$

where $\chi(\beta_0, \alpha_0, \hat{\zeta}, g_0; O) = \Psi(\beta_0, \alpha_0, \hat{\zeta}, g_0; Z) - \dot{l}_g(\beta_0, g_0; O)[h_2^*]$.

Define

$$Q_\alpha \equiv P\big\{\Psi(\beta_0,\alpha_0,\zeta_0,g_0;Z)^{\otimes 2}-\{\dot{\Psi}_g(\beta_0,\alpha_0,\zeta_0,g_0;Z)[h_2^*]\}^\top\big\}, \tag{S7.39}$$

$$-B_\alpha^\top = P\big\{\ddot{l}_{g\beta}(\beta_0,g_0;O)[h_2^*]-\{\dot{\Psi}_\beta(\beta_0,\alpha_0,\zeta_0,g_0;O)\}^\top\big\}, \tag{S7.40}$$

$$A_\alpha = -P\big\{\dot{\Psi}_\alpha(\beta_0,\alpha_0,\zeta_0,g_0;Z)\big\}. \tag{S7.41}$$

Note that

$$P\big\{\Psi(\beta_0,\alpha_0,\hat{\zeta},g_0;Z)^{\otimes 2}-\{\dot{\Psi}_g(\beta_0,\alpha_0,\zeta_0,g_0;Z)[h_2^*]\}^\top\big\}$$

$$
\begin{aligned}
=&P\Big\{\Psi(\beta_0,\alpha_0,\zeta_0,g_0;Z)^{\otimes 2}-\{\dot{\Psi}_g(\beta_0,\alpha_0,\zeta_0,g_0;Z)[h_2^*]\}^{\top}\Big\}\\
&+P\Big\{\{\Psi(\beta_0,\alpha_0,\zeta_0,g_0;Z)+\Psi(\beta_0,\alpha_0,\hat{\zeta},g_0;Z)\\
&-\Psi(\beta_0,\alpha_0,\zeta_0,g_0;Z)\}^{\otimes 2}-\Psi(\beta_0,\alpha_0,\zeta_0,g_0;Z)^{\otimes 2}\Big\}\\
=&Q_\alpha+\kappa D_\zeta\Sigma_\zeta D_\zeta^{\top}+o_p(1).
\end{aligned}
$$

Therefore, it follows from (S7.37), (S7.38), and (S7.31) that

$$
\begin{pmatrix}
\Sigma & B_\alpha & 0\\
-B_\alpha^{\top} & Q_\alpha+\kappa D_\zeta\Sigma_\zeta D_\zeta^{\top} & A_\alpha\\
0 & -A_\alpha^{\top} & 0
\end{pmatrix}
\begin{pmatrix}
n^{1/2}(\hat{\beta}_{\alpha\zeta}-\beta_0)\\
n^{1/2}\hat{\nu}_{\alpha\zeta}\\
n^{1/2}(\hat{\alpha}_\zeta-\alpha_0)
\end{pmatrix}
$$

$$
=\begin{pmatrix}
n^{1/2}\mathbb{P}_n\{\iota(\beta_0,g_0;O)\}\\
n^{1/2}\mathbb{P}_n\{\chi(\beta_0,\alpha_0,\hat{\zeta},g_0;O)\}\\
0
\end{pmatrix}+o_p(1),
$$

which implies that

$$
\begin{pmatrix}
\hat{\beta}_{\alpha\zeta}-\beta_0\\
\hat{\alpha}_\zeta-\alpha_0
\end{pmatrix}
=\begin{pmatrix}
I_{p\times p} & 0_{p\times J} & 0_{p\times(q+1)}\\
0_{1\times p} & 0_{1\times J} & 1_{1\times(q+1)}
\end{pmatrix}
\begin{pmatrix}
\hat{\beta}_{\alpha\zeta}-\beta_0\\
\hat{\nu}_{\alpha\zeta}\\
\hat{\alpha}_\zeta-\alpha_0
\end{pmatrix}
$$

$$
=\begin{pmatrix}
I_{p\times p} & 0_{p\times J} & 0_{p\times(q+1)}\\
0_{1\times p} & 0_{1\times J} & 1_{1\times(q+1)}
\end{pmatrix}
\begin{pmatrix}
\Sigma & B_\alpha & 0\\
-B_\alpha^{\top} & Q_\alpha+\kappa D_\zeta\Sigma_\zeta D_\zeta^{\top} & A_\alpha\\
0 & -A_\alpha^{\top} & 0
\end{pmatrix}^{-1}
$$

$$\cdot \begin{pmatrix} \mathbb{P}_n\{\iota(\beta_0, g_0; O)\} \\ \mathbb{P}_n\{\chi(\beta_0, \alpha_0, \hat{\zeta}, g_0; O)\} \\ 0 \end{pmatrix} + o_p(n^{-1/2}). \tag{S7.42}$$

Because the internal individual data and the external aggregate data are independent of each other, the asymptotic variance of

$$\sqrt{n}\mathbb{P}_n\{\chi(\beta_0, \alpha_0, \hat{\zeta}, g_0; O)\} = \sqrt{n}\mathbb{P}_n\big\{\Psi(\beta_0, \alpha_0, \hat{\zeta}, g_0; Z_i) - \dot{l}_g(\beta_0, g_0; O)[h_2^*]\big\}$$

is $Q_\alpha + \kappa D_\zeta \Sigma_\zeta D_\zeta^\top$. Then, the asymptotic distribution of $n^{1/2}(\hat{\beta}_{\alpha\zeta} - \beta_0)$ can be proved by a similar proof to that of $n^{1/2}(\hat{\beta}_{\text{au}} - \beta_0)$ in the proof of Theorem S1.

Next, we prove result (ii). We need to derive an asymptotic representation of $(\tilde{\beta}, \tilde{\alpha}_\zeta)$. It follows from Lemma 1 that

$$\sqrt{n}(\mathbb{P}_n - P)\{F\Psi(\tilde{\beta}, \tilde{\alpha}_\zeta, \hat{\zeta}, \tilde{g}; Z) - F\Psi(\beta_0, \alpha_0, \hat{\zeta}, g_0; Z)\} = o_p(1).$$

By Lemma 2 and the consistency of $(\tilde{\beta}, \tilde{g}, \tilde{\alpha}_\zeta)$, we have

$$\begin{aligned} -\mathbb{P}_n\{F\Psi(\beta_0, \alpha_0, \hat{\zeta}, g_0; Z)\} =& P\{F\dot{\Psi}_\beta(\beta_0, \alpha_0, \zeta_0, g_0; Z)\}(\tilde{\beta} - \beta_0) \\ &+ P\{F\dot{\Psi}_g(\beta_0, \alpha_0, \zeta_0, g_0; Z)[\tilde{g} - g_0]\} \\ &+ P\{F\dot{\Psi}_\alpha(\beta_0, \alpha_0, \zeta_0, g_0; Z)\}(\tilde{\alpha}_\zeta - \alpha_0) + o_p(n^{-1/2}). \end{aligned}$$

Similar to the proofs in Section S4, we can show that

$$\begin{pmatrix} \Sigma & 0_{p\times(q+1)} \\ -FB_\alpha^\top & FA_\alpha \end{pmatrix} \begin{pmatrix} \tilde{\beta}-\beta_0 \\ \tilde{\alpha}_\zeta - \alpha_0 \end{pmatrix} = \mathbb{P}_n \begin{pmatrix} \iota(\beta_0, g_0; O) \\ F\chi(\beta_0, \alpha_0, \hat{\zeta}, g_0; O) \end{pmatrix} + o_p(n^{-1/2}).$$

This implies that

$$\begin{aligned}
&\begin{pmatrix} \tilde{\beta}-\beta_0 \\ \tilde{\alpha}_\zeta - \alpha_0 \end{pmatrix} \\
&= \begin{pmatrix} \Sigma & 0_{p\times(q+1)} \\ -FB_\alpha^\top & FA_\alpha \end{pmatrix}^{-1} \begin{pmatrix} I_{p\times p} & 0_{p\times J} & 0_{p\times(q+1)} \\ 0_{1\times p} & F & 0 \end{pmatrix} \begin{pmatrix} \mathbb{P}_n\{\iota(\beta_0, g_0; O)\} \\ \mathbb{P}_n\{\chi(\beta_0, \alpha_0, \hat{\zeta}, g_0; O)\} \\ 0 \end{pmatrix} \\
&\quad + o_p(n^{-1/2}) \\
&= \begin{pmatrix} \Sigma & 0_{p\times(q+1)} \\ -FB_\alpha^\top & FA_\alpha \end{pmatrix}^{-1} \begin{pmatrix} I_{p\times p} & 0_{p\times J} & 0_{p\times(q+1)} \\ 0_{1\times p} & F & 0 \end{pmatrix} \begin{pmatrix} \Sigma & B_\alpha & 0 \\ -B_\alpha^\top & Q_\alpha + \kappa D_\zeta \Sigma_\zeta D_\zeta^\top & A_\alpha \\ 0 & -A_\alpha^\top & 0 \end{pmatrix} \\
&\quad \cdot \begin{pmatrix} \Sigma & B_\alpha & 0 \\ -B_\alpha^\top & Q_\alpha + \kappa D_\zeta \Sigma_\zeta D_\zeta^\top & A_\alpha \\ 0 & -A_\alpha^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{P}_n\{\iota(\beta_0, g_0; O)\} \\ \mathbb{P}_n\{\chi(\beta_0, \alpha_0, \hat{\zeta}, g_0; O)\} \\ 0 \end{pmatrix} + o_p(n^{-1/2}) \\
&= \begin{pmatrix} I_{p\times p} & \Sigma^{-1}B_\alpha & 0_{p\times(q+1)} \\ 0_{1\times p} & (FA_\alpha)^{-1}F(B_\alpha^\top\Sigma^{-1}B_\alpha + Q_\alpha + \kappa D_\zeta\Sigma_\zeta D_\zeta^\top) & 1_{1\times(q+1)} \end{pmatrix}
\end{aligned}$$

$$\cdot \begin{pmatrix} \Sigma & B_\alpha & 0 \\ -B_\alpha^\top & Q_\alpha + \kappa D_\zeta \Sigma_\zeta D_\zeta^\top & A_\alpha \\ 0 & -A_\alpha^\top & 0 \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{P}_n\{\iota(\beta_0, g_0; O)\} \\ \mathbb{P}_n\{\chi(\beta_0, \alpha_0, \hat{\zeta}, g_0; O)\} \\ 0 \end{pmatrix} + o_p(n^{-1/2}). \tag{S7.43}$$

By comparing the approximations in (S7.43) and (S7.42), we found that the following estimator

$$\begin{pmatrix} \tilde{\beta} \\ \tilde{\alpha}_\zeta \end{pmatrix} + \begin{pmatrix} 0_{p\times p} & -\hat{\Sigma}^{-1}\hat{B}_\alpha & 0_{p\times(q+1)} \\ 0_{1\times p} & -(F\hat{A}_\alpha)^{-1}F(\hat{B}_\alpha^\top \hat{\Sigma}_\alpha^{-1}\hat{B}_\alpha + \hat{Q}_\alpha + \kappa \hat{D}_\zeta \hat{\Sigma}_\zeta \hat{D}_\zeta^\top) & 0_{1\times(q+1)} \end{pmatrix}$$

$$\times \begin{pmatrix} \hat{\Sigma} & \hat{B}_\alpha & 0 \\ -\hat{B}_\alpha^\top & \hat{Q}_\alpha + \kappa \hat{D}_\zeta \hat{\Sigma}_\zeta \hat{D}_\zeta^\top & \hat{A}_\alpha \\ 0 & -\hat{A}_\alpha^\top & 0 \end{pmatrix}^{-1} \times \begin{pmatrix} \mathbb{P}_n\{\iota(\tilde{\beta}, \tilde{g}; O)\} \\ \mathbb{P}_n\{\chi(\tilde{\beta}, \tilde{\alpha}_\zeta, \hat{\zeta}, \tilde{g}; O)\} \\ 0 \end{pmatrix}$$

is a one-step estimator of (β, α) that has the same limiting distribution as $(\hat{\beta}_{\alpha\zeta}, \hat{\alpha}_\zeta)$. This finished the proof.

S7.1 Heterogeneity in covariate distribution

In this subsection, we consider the special case where the internal and external data may have different covariate distributions, but the uncertainty of auxiliary information can be ignored. The conclusions in Theorem S2 remain valid upon substituting κ with 0 and substituting $\hat{\zeta}$ with ζ_0. Since ζ is always a constant ζ_0 in this subsection, the auxiliary information function $\Psi(\beta, \alpha, \zeta, g; Z)$ and the proposed estimator

$\hat{\beta}_{\alpha\zeta}$ are irrelevant to ζ, we write them as $\Psi(\beta, \alpha, g; Z)$ and $\hat{\beta}_\alpha$ instead. We have the following corollary.

Corollary 1. *Assume that the regularity conditions (C1)–(C8) are satisfied and that model* (S7.27) *is correctly specified. Let* Q_α, B_α *and* A_α *be those defined in* (S7.39)-(S7.41)*, respectively. Also suppose that the matrices* $E\{\Psi(\beta_0, \alpha_0, \zeta_0, g_0; Z)^{\otimes 2}\}$ *and* Σ *are positive definite, the matrix* Q_α *is nonsingular, and the vector* A_α *is nonzero. Then as* $n \to \infty$*, we have (i)* $n^{1/2}(\hat{\beta}_\alpha - \beta_0)$ *converges to a mean zero normal distribution with covariance* $\Gamma_\alpha = (\Sigma + B_\alpha Q_\alpha^{-1} B_\alpha^\top - B_\alpha Q_\alpha^{-1} A_\alpha (A_\alpha^\top Q_\alpha^{-1} A_\alpha)^{-1} A_\alpha^\top Q_\alpha^{-1} B_\alpha^\top)^{-1}$*. (ii) The following estimator*

$$\begin{pmatrix} \tilde{\beta} \\ \tilde{\alpha} \end{pmatrix} + \begin{pmatrix} 0_{p\times p} & -\hat{\Sigma}^{-1}\hat{B}_\alpha & 0_{p\times(q+1)} \\ 0_{1\times p} & -(F\hat{A}_\alpha)^{-1}F(\hat{B}_\alpha^\top \hat{\Sigma}_\alpha^{-1}\hat{B}_\alpha + \hat{Q}_\alpha) & 0_{1\times(q+1)} \end{pmatrix} \\ \times \begin{pmatrix} \hat{\Sigma} & \hat{B}_\alpha & 0 \\ -\hat{B}_\alpha^\top & \hat{Q}_\alpha & \hat{A}_\alpha \\ 0 & -\hat{A}_\alpha^\top & 0 \end{pmatrix}^{-1} \times \begin{pmatrix} \mathbb{P}_n\{\iota(\tilde{\beta}, \tilde{g}; O)\} \\ \mathbb{P}_n\{\chi(\tilde{\beta}, \tilde{\alpha}, \tilde{g}; O)\} \\ 0 \end{pmatrix}$$

is a one-step estimator of (β, α)*, that has the same limiting distribution as* $(\hat{\beta}_\alpha, \hat{\alpha})$*.*

Corollary 1 implies that $\hat{\beta}_\alpha$ is asymptotically more efficient than the maximum conditional likelihood estimator $\tilde{\beta}$ that does not utilize the auxiliary information, and that $\hat{\beta}_\alpha$ is less efficient than $\hat{\beta}_{\alpha_0}$, which is the maximum likelihood estimator of β when the true parameter value α_0 is known.

S7.2 Uncertainty of auxiliary information

In this subsection, we consider the special case where the uncertainty of auxiliary information can not be ignored, but the internal and external data have the same covariate distributions. The conclusions in Theorem S2 remain valid upon substituting $\hat{\alpha}$ with $\alpha_0 = 0$. Since α is always a constant $\alpha_0 = 0$ in this subsection, the auxiliary information function $\Psi(\beta, \alpha, \zeta, g; Z)$ and the proposed estimator $\hat{\beta}_{\alpha\zeta}$ are irrelevant to α, we write them as $\Psi(\beta, \zeta, g; Z)$ and $\hat{\beta}_\zeta$ instead. We have the following corollary.

Corollary 2. *Suppose the conditions specified in Theorem 2 hold. As $n \to \infty$ and $n/m \to \kappa$, (i) $\sqrt{n}(\hat{\beta}_\zeta - \beta_0)$ converges to a mean zero normal distribution with covariance $\{\Sigma + B(Q + \kappa D_\zeta \Sigma_\zeta D_\zeta^\top)^{-1} B^\top\}^{-1}$. (ii) The following estimator*

$$\tilde{\beta} + (I \quad -\hat{\Sigma}^{-1}\hat{B}) \begin{pmatrix} \hat{\Sigma} & \hat{B} \\ -\hat{B}^\top & \hat{Q} + \kappa \hat{D}_\zeta \hat{\Sigma}_\zeta \hat{D}_\zeta^\top \end{pmatrix}^{-1} \begin{pmatrix} \mathbb{P}_n\{\iota(\tilde{\beta}, \tilde{g}; O)\} \\ \mathbb{P}_n\{\Psi(\tilde{\beta}, \hat{\zeta}, \tilde{g}; Z_i) - \dot{l}_g(\tilde{\beta}, \tilde{g}; O)[\hat{h}_2^*]\} \end{pmatrix}$$

is a one-step estimator of β that has the same first-order limiting distribution as $\hat{\beta}_\zeta$.

Corollary 2 shows that $\hat{\beta}_\zeta$ is asymptotically more efficient than the maximum conditional likelihood estimator $\tilde{\beta}$ without incorporating the auxiliary information. The efficiency gain decreases with κ. When κ is a very small constant that close to 0, that is, $m \gg n$, $B(Q + \kappa D_\zeta \Sigma_\zeta D_\zeta^\top)^{-1} B^\top$ is close to $BQ^{-1}B^\top$, and thus $\hat{\beta}_\zeta$ enjoys substantial efficiency gain. When κ is very large, that is, $n \gg m$, the efficiency gain is negligible.

Bibliography

Huang, C.-Y., J. Ning, and J. Qin (2015). Semiparametric likelihood inference for left-truncated and right-censored data. *Biostatistics 16*(4), 785–798.

Kosorok, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference.* New York: Springer-Verlag.

Lin, Y. and K. Chen (2013). Efficient estimation of the censored linear regression model. *Biometrika 100*(2), 525–530.

Pollard, D. (1990). Empirical processes: theory and applications. Hayward, CA: Institute of Mathematical Statistics.

Qin, J. and J. Lawless (1994). Empirical likelihood and general estimating equations. *The Annals of Statistics 22*(1), 300–325.

Qin, J., J. Ning, H. Liu, and Y. Shen (2011). Maximum likelihood estimations and EM algorithms with length-biased data. *Journal of the American Statistical Association 106*(496), 1434–1449.

Sheng, Y., Y. Sun, D. Deng, and C.-Y. Huang (2020). Censored linear regression in the presence or absence of auxiliary survival information. *Biometrics 76*(3), 734–745.

Sheng, Y., Y. Sun, C.-Y. Huang, and M.-O. Kim (2021). Synthesizing external aggregated information in the penalized Cox regression under population heterogeneity. *Statistics in Medicine 40*(23), 4915–4930.

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

Vardi, Y. (1989). Multiplicative censoring, renewal processes, deconvolution and decreasing density: nonparametric estimation. *Biometrika 76*(4), 751–761.

Zeidler, E. (1995). *Applied Functional Analysis: Main Principles and Their Applications*. New York: Springer-Verlag.

Zeng, D. and D. Lin (2007). Efficient estimation for the accelerated failure time model. *Journal of the American Statistical Association 102*(480), 1387–1396.