SEMIPARAMETRIC TRANSFORMATION MODELS WITH MULTILEVEL RANDOM EFFECTS FOR CORRELATED DISEASE ONSET IN FAMILIES

Baosheng Liang, Yuanjia Wang and Donglin Zeng

Peking University, Columbia University, and University of North Carolina at Chapel Hill

Supplementary Material

Proofs of Theorem 1 and Theorem 2 and additional simulation studies mimicking the real data are presented in this supplementary material.

S1 Proofs of the Asymptotic Results

We define $Z_{ij} = (G_{ij}, X_{ij}^T)^T$, $\widetilde{\theta} = (\beta, \gamma)$, $\Omega_{1ij}(\widetilde{\theta}, \Lambda) = \big\{1+\alpha\Lambda(Y_{ij})\exp(\widetilde{\theta}^T Z_{ij}+ b_i + r_{ij})\big\}^{-1}$, $\Omega_{2ij}(\widetilde{\theta}, \Lambda) = \exp(\widetilde{\theta}^T Z_{ij} + b_i + r_{ij})\Omega_{1ij}(\widetilde{\theta}, \Lambda)$, and $\Omega_{3ij}(\widetilde{\theta}, \Lambda) = \Lambda(Y_{ij})\Omega_{2ij}(\widetilde{\theta}, \Lambda)$. We further define

$$\Psi(O_i^*; \theta, \Lambda)$$
$$= \sum_{j=1}^{n_i} \left[\Delta_{ij}\{\log \lambda(Y_{ij}) + \widetilde{\theta}^T Z_{ij} + b_i + r_{ij}\} + (\Delta_{ij} + \frac{1}{\alpha}) \log\left\{\Omega_{1ij}(\widetilde{\theta}, \Lambda)\right\}\right]$$

$$- \big\{ \frac{1}{2}\log(2\pi\sigma_b^2) + \frac{b_i^2}{2\sigma_b^2} + \frac{n_i}{2}\log(2\pi\sigma_r^2) + \frac{1}{2}\log|\Sigma_i| + \frac{1}{2\sigma_r^2} r_{i\cdot}^T \Sigma_i^{-1} r_{i\cdot} \big\}$$

$$+ I(G_{i\cdot} = g_{i\cdot}) \log\{p_i(g_{i\cdot})\}$$

if $\alpha > 0$, and

$$\begin{aligned}
&\Psi(O_i^*; \theta, \Lambda) \\
=&\sum_{j=1}^{n_i} \big[\Delta_{ij}\{\log\lambda(Y_{ij}) + \widetilde{\theta}^T Z_{ij} + b_i + r_{ij}\} - \Lambda(Y_{ij}) \exp(\widetilde{\theta}^T Z_{ij} + b_i + r_{ij})\big] \\
&- \big\{ \frac{1}{2}\log(2\pi\sigma_b^2) + \frac{b_i^2}{2\sigma_b^2} + \frac{n_i}{2}\log(2\pi\sigma_r^2) + \frac{1}{2}\log|\Sigma_i| + \frac{1}{2\sigma_r^2} r_{i\cdot}^T \Sigma_i^{-1} r_{i\cdot} \big\} \\
&+ I(G_{i\cdot} = g_{i\cdot}) \log\{p_i(g_{i\cdot})\}
\end{aligned}$$

when $\alpha = 0$. For abbreviation, we use $\Psi_{1i}(\theta, \Lambda) \triangleq \Psi(O_i^*; \theta, \Lambda)$ so the log-likelihood function is

$$l_n(O; \theta, \Lambda) = \sum_{i=1}^{n} \log\Big[\sum_{g_{i\cdot}} \int_{b_i, r_{i\cdot}} \exp\big\{\Psi_{1i}(\theta, \Lambda)\big\} db_i dr_{i\cdot}\Big].$$

In the following, we only prove the theorems for $\alpha > 0$ as the proof is similar when $\alpha = 0$.

Proof of Theorem 1. If we can show the proposed objective function and the proposed estimators satisfy conditions (C1)–(C5) of Theorem 1 in Zeng and Lin (2010), then our Theorem 1 directly follows from their results. Since the conditions (C1) and (C2) naturally hold under our conditions (A.1)–(A.2), we only need to verify (C3), (C4) and (C5) in Zeng and Lin

(2010). To this end, we denote

$$\Psi_{2i}(\theta, \Lambda) = \sum_{g_{i\cdot}} I(G_{i\cdot} = g_{i\cdot})\, p_i(g_{i\cdot}) \int_{b_i, r_{i\cdot}} \prod_{j=1}^{n_i} \Psi_{2ij}(\widetilde{\theta}, \Lambda)\psi(b_i, r_{i\cdot}; \sigma^2) db_i dr_{i\cdot},$$

where $\Psi_{2ij}(\widetilde{\theta}, \Lambda) = \big\{\Omega_{2ij}(\widetilde{\theta}, \Lambda)\big\}^{\Delta_{ij}} \big\{\Omega_{1ij}(\widetilde{\theta}, \Lambda)\big\}^{1/\alpha}$, $\psi(b_i, r_{i\cdot}; \sigma^2) = |\Sigma_i|^{-1/2}$ $\times(2\pi\sigma_b^2)^{-1/2}(2\pi\sigma_r^2)^{-n_i/2} \exp(-\frac{b_i^2}{2\sigma_b^2} - \frac{r_{i\cdot}^T \Sigma_i^{-1} r_{i\cdot}}{2\sigma_r^2})$ and $\sigma^2 = (\sigma_b^2, \sigma_r^2)$.

We first check condition (C3) in Zeng and Lin (2010). According to (A.1) and (A.2), there exist constants m and M such that $0 < m \leq \exp\{-(\beta G_{ij} + \gamma^T X_{ij})\} \leq M < \infty$ almost surely for all $i = 1, ..., n$, $j = 1, ..., n_i$. Then, we have

$$\begin{aligned} &\alpha\Lambda(Y_{ij}) + \exp\{-(\beta G_{ij} + \gamma^T X_{ij} + b_i + r_{ij})\} \geq \{\alpha\Lambda(Y_{ij}) + m\} \exp(-|b_i + r_{ij}|), \\ &\alpha\Lambda(Y_{ij}) + \exp\{-(\beta G_{ij} + \gamma^T X_{ij} + b_i + r_{ij})\} \leq \{\alpha\Lambda(Y_{ij}) + M]\} \exp(|b_i + r_{ij}|). \end{aligned} \tag{S1.1}$$

If $\|\Lambda\|_{V[0,\tau]}$ is bounded, then $\Psi_{2ij}(\widetilde{\theta}, \Lambda) \geq \big[\{\alpha\Lambda(\tau)+M\} \exp(|b_i+r_{ij}|)\big]^{-(\frac{1}{\alpha}+1)}$, so $\Psi_{2i}(\theta, \Lambda) \geq O(1)\big\{\alpha\Lambda(\tau) + M\big\}^{-(\frac{1}{\alpha}+1)} > 0$. Thus, the second half of (C3) holds. Furthermore, we let $N_{ij}(\cdot)$ be a counting process of the jth family member in ith family, by (S1.1) and the characteristic of logarithmic transformation function $H(\cdot)$, there exist constants C_1 such that

$$\begin{aligned} &\Psi_{2ij}(\widetilde{\theta}, \Lambda) \\ \leq &C_1 \prod_{t\leq\tau} \Big[\exp\{-(b_i + r_{ij})\}\big\{1 + \alpha \int_0^t R_{ij}(s) \exp(\widetilde{\theta}^T Z_{ij} + b_i + r_{ij}) d\Lambda(s)\big\}\Big]^{-dN_{ij}(t)} \end{aligned}$$

$$\times \big\{1+\alpha\int_0^\tau R_{ij}(s)\exp(\widetilde{\theta}^T Z_{ij}+b_i+r_{ij})d\Lambda(s)\big\}^{-\frac{1}{\alpha}}.$$

Because $\exp(\widetilde{\theta}^T Z_{ij}+b_i+r_{ij}) \geq \exp\{-O(1+|b_i+r_{ij}|)\}$, we have

$$\begin{aligned} &1+\alpha\int_0^t R_{ij}(s)\exp(\widetilde{\theta}^T Z_{ij}+b_i+r_{ij})d\Lambda(s) \\ \geq\ &\exp\{-O(1+|b_i+r_{ij}|)\}\big\{1+\alpha\int_0^t R_{i\cdot}(s)d\Lambda(s)\big\}. \end{aligned}$$

This gives

$$\begin{aligned} \Psi_{2ij}(\widetilde{\theta},\Lambda) \leq &\exp\{O(1+|b_i+r_{ij}|)\} \\ &\times\prod_{t\leq\tau}\big\{1+\int_0^t R_{i\cdot}(s)d\Lambda(s)\big\}^{-dN_{ij}(t)}\big\{1+\int_0^\tau R_{i\cdot}(s)d\Lambda(s)\big\}^{-\frac{1}{\alpha}}. \end{aligned} \tag{S1.2}$$

Hence, the other half of condition (C3) in Zeng and Lin (2010) holds.

We next verify condition (C4) in Zeng and Lin (2010). From inequation (S1.2), we obtain $|\Psi_{2ij}(\widetilde{\theta},\Lambda)| \leq \exp\{O(1+|b_i+r_{ij}|)\}$. Thus,

$$\begin{aligned} \Big|\frac{\partial}{\partial\beta}\Psi_{2ij}(\widetilde{\theta},\Lambda)\Big| &= \Big|\Psi_{2ij}(\widetilde{\theta},\Lambda)\big\{\Delta_{ij}-(\alpha\Delta_{ij}+1)\Omega_{3ij}(\widetilde{\theta},\Lambda)\big\}G_{ij}\Big| \\ &\leq \exp\{O(1+|b_i+r_{ij}|)\}, \end{aligned}$$

$$\begin{aligned} \Big|\frac{\partial}{\partial\gamma}\Psi_{2ij}(\widetilde{\theta},\Lambda)\Big| &= \Big|\Psi_{2ij}(\widetilde{\theta},\Lambda)\big\{\Delta_{ij}-(\alpha\Delta_{ij}+1)\Omega_{3ij}(\widetilde{\theta},\Lambda)\big\}X_{ij}\Big| \\ &\leq \exp\{O(1+|b_i+r_{ij}|)\}, \\ \Big|\frac{\partial}{\partial\Lambda}\Psi_{2ij}(\widetilde{\theta},\Lambda)[\widetilde{H}]\Big| &= \Big|\Psi_{2ij}(\widetilde{\theta},\Lambda)(\alpha\Delta_{ij}+1)\int_0^{Y_{ij}}\Omega_{2ij}(\widetilde{\theta},\Lambda)d\widetilde{H}(s)\Big| \\ &\leq \exp\{O(1+|b_i+r_{ij}|)\}, \end{aligned}$$

where $\widetilde{H}$ belongs to a set of functions in which $\Lambda + \epsilon\widetilde{H}$ is increasing with bounded total variation. Then by the mean-value theorem, we have

$$\begin{aligned}
&\left|\Psi_{2ij}(\beta^{(1)}, \gamma, \Lambda) - \Psi_{2ij}(\beta^{(2)}, \gamma, \Lambda)\right| \le \exp\{O(1 + |b_i + r_{ij}|)\}|\beta^{(1)} - \beta^{(2)}|, \\
&\left|\Psi_{2ij}(\widetilde{\theta}^{(1)}, \Lambda) - \Psi_{2ij}(\widetilde{\theta}^{(2)}, \Lambda)\right| \le \exp\{O(1 + |b_i + r_{ij}|)\}|\gamma^{(1)} - \gamma^{(2)}|, \\
&\left|\Psi_{2ij}(\widetilde{\theta}, \Lambda^{(1)}) - \Psi_{2ij}(\widetilde{\theta}, \Lambda^{(2)})\right| = \left|\frac{\partial}{\partial\Lambda}\Psi_{2ij}(\widetilde{\theta}, \Lambda^*)[\Lambda^{(1)} - \Lambda^{(2)}]\right| \\
&\qquad\qquad \le \exp\{O(1 + |b_i + r_{ij}|)\}\int_0^\tau |\Lambda^{(1)}(s) - \Lambda^{(2)}(s)|ds.
\end{aligned}$$

Fix $G_{i\cdot} = g_{i\cdot}$, by the form of $\psi(b_i, r_{i\cdot}; \sigma^2)$, it holds

$$\sup_{\sigma^2} E\Big[\int_{b_i, r_{i\cdot}} \exp\{O(1 + |b_i + r_{ij}|)\}\psi(b_i, r_{i\cdot}; \sigma^2) db_i dr_{i\cdot}\Big] < \infty,$$

and there exists $C > 0$ such that

$$\left|\frac{\dot{\psi}(b_i, r_{i\cdot}; \sigma^2)}{\psi(b_i, r_{i\cdot}; \sigma^2)}\right| + \left|\frac{\ddot{\psi}(b_i, r_{i\cdot}; \sigma^2)}{\psi(b_i, r_{i\cdot}; \sigma^2)}\right| + \left|\frac{\dddot{\psi}(b_i, r_{i\cdot}; \sigma^2)}{\psi(b_i, r_{i\cdot}; \sigma^2)}\right| \le C\exp(1 + |b_i + r_{i\cdot}|),$$

where $\dot{\psi}$, $\ddot{\psi}$ and $\dddot{\psi}$ indicate the first, second and third derivative with respect to σ^2. Thus

$$\begin{aligned}
&\left|\Psi_{2i}(\theta^{(1)}, \Lambda^{(1)}) - \Psi_{2i}(\theta^{(2)}, \Lambda^{(2)})\right| \\
\le& O(1)\exp(1 + |b_i + r_{i\cdot}|)\Big\{|\theta^{(1)} - \theta^{(2)}| + \int_0^\tau |\Lambda^{(1)}(s) - \Lambda^{(2)}(s)|ds\Big\}.
\end{aligned}$$

Similarly, we can verify the bounds for the other three terms in (C4) in Zeng and Lin (2010).

Condition (C5) in Zeng and Lin (2010) follows from the identifiability result in Lemma 1 at the end of this appendix. Hence, we have verified all conditions in Theorem 1 of Zeng and Lin (2010) so our Theorem 1 holds. □

Proof of Theorem 2. Consider a parameter set

$$\mathcal{H} = \{(h_1, h_2) : h_1 \in \mathrm{R}^d, h_2(\cdot) \text{ is a function on } [0,\tau] : \|h_1\| \leq 1, \|h_2\|_{BV} \leq 1\},$$

where d is the dimension of θ_0. For subject i, by taking derivative with respect to Euclidian parameter θ and Λ along the submodel $\theta + \epsilon h_1$ and $d\Lambda_\epsilon = (1+\epsilon h_2)d\Lambda$, we obtain a map $\mathbb{S}_n : (\theta, \Lambda) \mapsto l^\infty(\mathcal{H})$ defined as

$$\begin{aligned}\mathbb{S}_n(\theta,\Lambda)[h_1,h_2] &\triangleq n^{-1}\frac{d}{d\epsilon} l_n\Big(\theta+\epsilon h_1, \Lambda(t) + \epsilon\int_0^t h_2(s)d\Lambda(s)\Big)\Big|_{\epsilon=0} \\ &= \mathbb{P}_n \dot{l}_\theta(\theta,\Lambda)[h_1] + \mathbb{P}_n \dot{l}_\Lambda(\theta,\Lambda)[h_2],\end{aligned}$$

where $\mathbb{P}_n$ denotes the empirical measure based on n independent families,

$$\dot{l}_\theta(\theta,\Lambda)[h_1] = h_1^T E_f\{V_{1i}(\theta,\Lambda)\}, \quad \dot{l}_\Lambda(\theta,\Lambda)[h_2] = E_f\{V_{2i}(\theta,\Lambda)[h_2]\},$$

and $V_{1i}(\theta,\Lambda) = \left(\dot{l}_{\sigma^2}(\theta,\Lambda)^T, \sum_{j=1}^{n_i}\left\{\Delta_{ij} - (\alpha\Delta_{ij}+1)\Omega_{3ij}(\widetilde{\theta},\Lambda)\right\}Z_{ij}\right)^T$,

$$\begin{aligned}&V_{2i}(\theta,\Lambda)[h_2] \\ =&\sum_{j=1}^{n_i}\Big\{\Delta_{ij}h_2(Y_{ij}) - (\alpha\Delta_{ij}+1)\int_0^\tau \Omega_{2ij}(\widetilde{\theta},\Lambda)I(s\leq Y_{ij})h_2(s)d\Lambda(s)\Big\}.\end{aligned}$$

Here, $E_f(Q)$ is defined as

$$\Big[\sum_{g_{i\cdot}}\int \exp\big\{\Psi_{1i}(\theta,\Lambda)\big\}db_i dr_{i\cdot}\Big]^{-1}\sum_{g_{i\cdot}}\int Q\exp\big\{\Psi_{1i}(\theta,\Lambda)\big\}db_i dr_{i\cdot}.$$

We further define map $\mathbb{S}$: $(\theta, \Lambda) \mapsto l^{\infty}(\mathcal{H})$ as the limit of $\mathbb{S}_n$, and

$$\mathbb{S}(\theta, \Lambda)[h_1, h_2] = \mathbb{P}\dot{l}_{\theta}(\theta, \Lambda)[h_1] + \mathbb{P}\dot{l}_{\Lambda}(\theta, \Lambda)[h_2],$$

where $\mathbb{P}$ is the expectation corresponds to the empirical measure $\mathbb{P}_n$, then $\mathbb{S}_n(\widehat{\theta}_n, \widehat{\Lambda}_n) = 0$ and $\mathbb{S}(\theta_0, \Lambda_0) = 0$.

First, because $\dot{l}_{\theta}$ and $\dot{l}_{\Lambda}$ are bounded Lipschitz functions on $\mathcal{H}$, and $\mathcal{H}$ is a Donsker class, by Donsker theorem it is easy to know $\sqrt{n}(\mathbb{S}_n(\theta_0, \Lambda_0) - \mathbb{S}(\theta_0, \Lambda_0))$ converges weakly to a Gaussian process on $\mathcal{H}$. Following the arguments in Theorem 2 of Murphy (1995), we obtain that

$$\begin{aligned}
&\sup_{(h_1,h_2)\in\mathcal{H}} \left|\sqrt{n}(\mathbb{S}_n - \mathbb{S})(\widehat{\theta}_n, \widehat{\Lambda}_n)[h_1, h_2] - \sqrt{n}(\mathbb{S}_n - \mathbb{S})(\theta_0, \Lambda_0)[h_1, h_2]\right| \\
=&o_P\{1 + \sqrt{n}(\|\theta - \theta_0\| + \|\Lambda - \Lambda_0\|)\}.
\end{aligned}$$

Next we show operator $\dot{\mathbb{S}}(\theta_0, \Lambda_0)$, which is the derivative of $\mathbb{S}(\theta, \Lambda)$ at (θ_0, Λ_0), mapping set $\{(\theta-\theta_0, \Lambda-\Lambda_0) : (\theta, \Lambda) \text{ is in the neighborhood of } (\theta_0, \Lambda_0)\}$ to $l^{\infty}(\mathcal{H})$, is continuous and invertible on its range. Note that

$$\begin{aligned}
\ddot{l}_{\theta\theta}(\theta, \Lambda)[h_1, \widetilde{h}_1] =& h_1^T\Big\{E_f V_{1i}(\theta, \Lambda)V_{1i}^T(\theta, \Lambda) + E_f\frac{\partial}{\partial\theta}V_{1i}(\theta, \Lambda) \\
&- E_f V_{1i}(\theta, \Lambda)E_f V_{1i}^T(\theta, \Lambda)\Big\}\widetilde{h}_1, \\
\ddot{l}_{\theta\Lambda}(\theta, \Lambda)[h_1, h_2] =& h_1^T\Big\{E_f V_{1i}(\theta, \Lambda)V_{2i}(\theta, \Lambda)[h_2] + E_f\frac{\partial}{\partial\Lambda}V_{1i}(\theta, \Lambda)[h_2] \\
&- E_f V_{1i}(\theta, \Lambda)E_f V_{2i}(\theta, \Lambda)[h_2]\Big\},
\end{aligned}$$

$$\begin{aligned}\ddot{l}_{\Lambda\theta}(\theta,\Lambda)[h_2,h_1] =& h_1^T\Big\{E_f V_{1i}(\theta,\Lambda)V_{2i}(\theta,\Lambda)[h_2] + E_f\frac{\partial}{\partial\theta}V_{2i}(\theta,\Lambda)[h_2]\\ &- E_f V_{1i}(\theta,\Lambda)E_f V_{2i}(\theta,\Lambda)[h_2]\Big\},\end{aligned}$$

$$\begin{aligned}\ddot{l}_{\Lambda\Lambda}(\theta,\Lambda)[h_2,\widetilde{h}_2] =& E_f V_{2i}(\theta,\Lambda)[h_2]V_{2i}(\theta,\Lambda)[\widetilde{h}_2] + E_f\frac{\partial}{\partial\Lambda}V_{2i}(\theta,\Lambda)[h_2,\widetilde{h}_2]\\ &- E_f V_{2i}(\theta,\Lambda)[h_2]E_f V_{2i}(\theta,\Lambda)[\widetilde{h}_2],\end{aligned}$$

where

$$\frac{\partial}{\partial\theta}V_{1i}(\theta,\Lambda) = \mathrm{diag}\Big(\frac{1}{2\sigma_b^4}-\frac{b_i^2}{\sigma_b^6}, \frac{n_i}{2\sigma_r^4}-\frac{r_{i\cdot}^T\Sigma_i r_{i\cdot}}{\sigma_r^6}, \sum_{j=1}^{n_i}(\alpha\Delta_{ij}+1)\big(\alpha\Omega_{3ij}^2-\Omega_{3ij}\big)Z_{ij}Z_{ij}^T\Big),$$

$$\frac{\partial}{\partial\Lambda}V_{1i}(\theta,\Lambda)[h_2] = \frac{\partial}{\partial\theta}V_{2i}(\theta,\Lambda)[h_2] = (0,0,\sum_{j=1}^{n_i}B_{ij}^*)^T,$$

and

$$B_{ij}^* = (\alpha\Delta_{ij}+1)(\alpha\Omega_{3ij}(\widetilde{\theta},\Lambda)-1)\Omega_{2ij}(\widetilde{\theta},\Lambda)Z_{ij}\int_0^{Y_{ij}} h_2(s)d\Lambda(s),$$

$$\begin{aligned}\frac{\partial}{\partial\Lambda}V_{2i}(\theta,\Lambda)[h_2,\widetilde{h}_2] =& -\sum_{j=1}^{n_i}(\alpha\Delta_{ij}-1)\\ &\times\int_0^{Y_{ij}}\Big\{\Omega_{2ij}(\widetilde{\theta},\Lambda)\widetilde{h}_2(s)-\alpha\Omega_{2ij}^2(\widetilde{\theta},\Lambda)\int_0^{Y_{ij}}\widetilde{h}_2(t)d\Lambda(t)\Big\}h_2(s)d\Lambda(s).\end{aligned}$$

Hence,

$$\dot{\mathbb{S}}(\theta_0,\Lambda_0)(\theta-\theta_0,\Lambda-\Lambda_0)[h_1,h_2] = (\theta-\theta_0)^T A_1(h_1,h_2)+\int_0^\tau A_2(h_1,h_2)d(\Lambda-\Lambda_0),$$

where $A_1(h_1,h_2) = B_1h_1 + B_2h_2(s) + \int_0^\tau h_2(s)D_1(s)ds$,

$$
\begin{aligned}
D_1(s) =& E_f\Big[V_{1i}(\theta,\Lambda)\sum_{j=1}^{n_i}\big\{(\alpha\Delta_{ij}+1)\Omega_{2ij}(\widetilde{\theta},\Lambda)I(s\le Y_{ij})\big\}\Big] \\
&+ E_f\Big(0,0,\sum_{j=1}^{n_i}(\alpha\Delta_{ij}+1)\big\{\alpha\Omega_{3ij}(\widetilde{\theta},\Lambda)-1\big\}\Omega_{2ij}(\widetilde{\theta},\Lambda)Z_{ij}^T I(s\le Y_{ij})\Big)^T \\
&- E_f V_{1i}(\theta,\Lambda)E_f\sum_{j=1}^{n_i}\big\{(\alpha\Delta_{ij}+1)\Omega_{2ij}(\widetilde{\theta},\Lambda)I(s\le Y_{ij})\big\},
\end{aligned}
$$

$$
A_2(h_1,h_2) = B_3(s)h_1 + C_1(s)h_2(s) + \int_0^\tau h_2(t)D_2(s,t)dt,
$$

$$
\begin{aligned}
C_1(s) = E_f\Big[&\sum_{j=1}^{n_i}\big\{\Delta_{ij}\frac{I(s=Y_{ij})}{\lambda(s)} - (\alpha\Delta_{ij}+1)\Omega_{2ij}(\widetilde{\theta},\Lambda)I(s\le Y_{ij})\big\}\sum_{j=1}^{n_i}\Delta_{ij}I(s=Y_{ij})\Big] \\
&- E_f\Big\{\sum_{j=1}^{n_i}(\alpha\Delta_{ij}-1)\Omega_{2ij}(\widetilde{\theta},\Lambda)I(s\le Y_{ij})\Big\} \\
&- E_f\Big\{\sum_{j=1}^{n_i}\Delta_{ij}\frac{I(s=Y_{ij})}{\lambda(s)} - (\alpha\Delta_{ij}+1)\Omega_{2ij}(\widetilde{\theta},\Lambda)I(s\le Y_{ij})\Big\} \\
&\quad \times E_f\Big\{\sum_{j=1}^{n_i}\Delta_{ij}I(s=Y_{ij})\Big\},
\end{aligned}
$$

and

$$
\begin{aligned}
D_2(s,t) =& E_f\Big\{\sum_{j=1}^{n}\frac{\Delta_{ij}I(s=Y_{ij})}{\lambda(s)} - (\alpha\Delta_{ij}+1)\Omega_{2ij}(\widetilde{\theta},\Lambda)I(s\le Y_{ij})\Big\} \\
&\times E_f\Big\{\sum_{j=1}^{n_i}(\alpha\Delta_{ij}+1)\Omega_{2ij}(\widetilde{\theta},\Lambda)I(t\le Y_{ij})\Big\} \\
&- E_f\Big[\sum_{j=1}^{n}\big\{\frac{\Delta_{ij}I(s=Y_{ij})}{\lambda(s)} - (\alpha\Delta_{ij}+1)\Omega_{2ij}(\widetilde{\theta},\Lambda)I(s\le Y_{ij})\big\} \\
&\quad \times \sum_{j=1}^{n}\big\{(\alpha\Delta_{ij}+1)\Omega_{2ij}(\widetilde{\theta},\Lambda)I(t\le Y_{ij})\big\}\Big]
\end{aligned}
$$

$$- E_f\Big\{\sum_{j=1}^{n}(\alpha\Delta_{ij}-1)\alpha\Omega_{2ij}^2(\widetilde{\theta},\Lambda)I(t\le Y_{ij}, s\le Y_{ij})\Big\},$$

and B_1, B_2 and B_3 are in turn

$$B_1 = E_f\frac{\partial}{\partial\theta}V_{1i}(\theta,\Lambda) + E_fV_{1i}(\theta,\Lambda)V_{1i}(\theta,\Lambda)^T - E_fV_{1i}(\theta,\Lambda)E_fV_{1i}(\theta,\Lambda)^T,$$
$$B_2 = E_f\Big\{V_{1i}(\theta,\Lambda)\sum_{j=1}^{n_i}\Delta_{ij}I(s=Y_{ij})\Big\} - E_f\{V_{1i}(\theta,\Lambda)\}E_f\Big\{\sum_{j=1}^{n_i}\Delta_{ij}I(s=Y_{ij})\Big\},$$

$$B_3(s) = E_f\Big[V_{1i}(\theta,\Lambda)\sum_{j=1}^{n_i}\big\{\Delta_{ij}\frac{I(s=Y_{ij})}{\lambda(s)} - (\alpha\Delta_{ij}+1)\Omega_{2ij}(\widetilde{\theta},\Lambda)I(s\le Y_{ij})\big\}\Big]$$
$$+ E_f\Big(0,0,\sum_{j=1}^{n_i}(\alpha\Delta_{ij}+1)\big\{\alpha\Omega_{3ij}(\widetilde{\theta},\Lambda)-1\big\}\Omega_{2ij}(\widetilde{\theta},\Lambda)Z_{ij}^T I(s\le Y_{ij})\Big)^T$$
$$- E_fV_{1i}(\theta,\Lambda)E_f\sum_{j=1}^{n_i}\Big\{\Delta_{ij}\frac{I(s=Y_{ij})}{\lambda(s)} - (\alpha\Delta_{ij}+1)\Omega_{2ij}(\widetilde{\theta},\Lambda)I(s\le Y_{ij})\Big\}.$$

It is easy to show that $\dot{\mathbb{S}}(\theta_0,\Lambda_0)$ is a continuous Fredholm operator. From Lemma 2 at the end of this appendix, we show that if $\dot{l}_\theta[h_1]+\dot{l}_\Lambda[h_2]=0$ almost surely, then $h_1\equiv 0$ and $h_2\equiv 0$. Thus, $\dot{\mathbb{S}}(\theta_0,\Lambda_0)$ is one-to-one so is invertible.

Finally, according to Theorem 3.3.1 of van der Varrt and Wellner (1996), we conclude that $\sqrt{n}(\widehat{\theta}_n-\theta_0,\widehat{\Lambda}_n-\Lambda_0)$ converges weakly to a mean zero Gaussian process in the metric space $\mathbb{B}^d\times l^\infty(\mathcal{L})$. In addition, since

$$\sqrt{n}\dot{\mathbb{S}}(\theta_0,\Lambda_0)(\widehat{\theta}_n-\theta_0,\widehat{\Lambda}_n-\Lambda_0)[h_1,h_2] = \sqrt{n}(\mathbb{S}_n-\mathbb{S})(\theta_0,\Lambda_0)[h_1,h_2]+o_p(1),$$

we can choose a h_2^* such that $\dot{\mathbb{S}}(\theta_0,\Lambda_0)[h_1,h_2^*]=h_1$. Thus,

$$\sqrt{n}(\widehat{\theta}_n-\theta_0)^T h_1 = \sqrt{n}(\mathbb{S}_n-\mathbb{S})(\theta_0,\Lambda_0)[h_1,h_2^*]+o_p(1).$$

We conclude that $\widehat{\theta}_n$ is an asymptotically linear estimator for θ_0 and that its influence function is on the space spanned by the score functions. Thus $\widehat{\theta}_n$ is an asymptotically efficient estimator. □

Lemma 1. *Suppose that conditions (A.1)-(A.5) hold. If*

$$\begin{aligned}
&\sum_{g_{i\cdot}} \Big[I(G_{i\cdot} = g_{i\cdot})\, p_i(g_{i\cdot}) \\
&\quad \times \int \prod_{j=1}^{n_i} \big\{ \lambda(Y_{ij}) \Omega_{2ij}(\beta, \gamma, \Lambda) \big\}^{\Delta_{ij}} \big\{ \Omega_{1ij}(\beta, \gamma, \Lambda) \big\}^{\frac{1}{\alpha}} \psi(b_i, r_{i\cdot}; \sigma^2) db_i dr_{i\cdot} \Big] \\
=& \sum_{g_{i\cdot}} \Big[I(G_{i\cdot} = g_{i\cdot})\, p_i(g_{i\cdot}) \\
&\quad \times \int \prod_{j=1}^{n_i} \big\{ \lambda_0(Y_{ij}) \Omega_{2ij}(\beta_0, \gamma_0, \Lambda_0) \big\}^{\Delta_{ij}} \big\{ \Omega_{1ij}(\beta_0, \gamma_0, \Lambda_0) \big\}^{\frac{1}{\alpha}} \psi(b_i, r_{i\cdot}; \sigma_0^2) db_i dr_{i\cdot} \Big],
\end{aligned} \tag{S1.3}$$

then $\theta = \theta_0$ *and* $\Lambda = \Lambda_0$.

Proof. First, by condition (A.5), it holds

$$\begin{aligned}
&\int \prod_{j=1}^{n_i} \big\{ \lambda(Y_{ij}) \Omega_{2ij}(\beta, \gamma, \Lambda) \big\}^{\Delta_{ij}} \big\{ \Omega_{1ij}(\beta, \gamma, \Lambda) \big\}^{\frac{1}{\alpha}} \psi(b_i, r_{i\cdot}; \sigma^2) db_i dr_{i\cdot} \\
=& \int \prod_{j=1}^{n_i} \big\{ \lambda_0(Y_{ij}) \Omega_{2ij}(\beta_0, \gamma_0, \Lambda_0) \big\}^{\Delta_{ij}} \big\{ \Omega_{1ij}(\beta_0, \gamma_0, \Lambda_0) \big\}^{\frac{1}{\alpha}} \psi(b_i, r_{i\cdot}; \sigma_0^2) db_i dr_{i\cdot}
\end{aligned} \tag{S1.4}$$

then it suffices to show that (S1.4) implies $\theta = \theta_0$ and $\Lambda = \Lambda_0$.

Let $Y_{ij} = 0$ for $j = 1, ..., n_i$, $\Delta_{i1} = 1$ and $\Delta_{ij'} = 0$ for $j' \neq 1$, then (S1.4) becomes

$$\begin{aligned}&\lambda(0)\exp(\beta g_{i1}+\gamma^T X_{i1})\int\exp(b_i+r_{i1})\psi(b_i,r_{i\cdot};\sigma^2)db_i dr_{i\cdot}\\=&\lambda_0(0)\exp(\beta_0 g_{i1}+\gamma_0^T X_{i1})\int\exp(b_i+r_{i1})\psi(b_i,r_{i\cdot};\sigma_0^2)db_i dr_{i\cdot},\end{aligned}$$

which could be written as

$$\log\lambda(0)+\beta g_{ij}+\gamma^T X_{ij}+C_{n_i}(\sigma^2)=\log\lambda_0(0)+\beta_0 g_{ij}+\gamma_0^T X_{ij}+C_{n_i}(\sigma_0^2).$$

According to condition (A.4), we first conclude that $\beta=\beta_0$ and $\gamma=\gamma_0$.

Fix any k such that $1\le k\le n_i$, let $\Delta_{ij}=1$, $Y_{ij}=0$ in (S1.4) for $j=1,...,k$; for those j such that $j>k$, if $\Delta_{ij}=0$, then replace $Y_{ij}=\tau$, otherwise integrate Y_{ij} from 0 to τ. Then we obtain

$$\begin{aligned}&\int\prod_{j=1}^{k}\lambda(0)\exp(b_i+r_{ij})\prod_{j=k+1}^{n_i}\left\{1+\alpha\Lambda(\tau)\exp(\beta g_{ij}+\gamma^T X_{ij}+b_i+r_{ij})\right\}^{-\frac{1-\Delta_{ij}}{\alpha}}\\&\times\left[\frac{1}{\alpha}-\frac{1}{\alpha}\left\{1+\alpha\Lambda(\tau)\exp(\beta g_{ij}+\gamma^T X_{ij}+b_i+r_{ij})\right\}^{-\frac{1}{\alpha}}\right]^{\Delta_{ij}}\psi(b_i,r_{i\cdot};\sigma^2)db_i dr_{i\cdot}\\&=\int\prod_{j=1}^{k}\lambda_0(0)\exp(b_i+r_{ij})\prod_{j=k+1}^{n_i}\left\{1+\alpha\Lambda_0(\tau)\exp(\beta_0 g_{ij}+\gamma_0^T X_{ij}+b_i+r_{ij})\right\}^{-\frac{1-\Delta_{ij}}{\alpha}}\\&\times\left[\frac{1}{\alpha}-\frac{1}{\alpha}\left\{1+\alpha\Lambda_0(\tau)\exp(\beta_0 g_{ij}+\gamma_0^T X_{ij}+b_i+r_{ij})\right\}^{-\frac{1}{\alpha}}\right]^{\Delta_{ij}}\psi(b_i,r_{i\cdot};\sigma_0^2)db_i dr_{i\cdot}.\end{aligned}\tag{S1.5}$$

Since $\{\Delta_{ij}:j=k+1,...,n_i\}$ are arbitrary, by summing over Δ_{ij}, $j=$

$k+1,...,n_i$, on both sides of (S1.5), we obtain

$$\int \lambda(0)^k \prod_{j=1}^{k} \exp\{b_i + r_{ij}\}\psi(b_i, r_{i\cdot}; \sigma^2) db_i dr_{i\cdot}$$
$$= \int \lambda_0(0)^k \prod_{j=1}^{k} \exp\{b_i + r_{ij}\}\psi(b_i, r_{i\cdot}; \sigma_0^2) db_i dr_{i\cdot}. \tag{S1.6}$$

Let $\Sigma^* = \begin{pmatrix} \sigma_b^2 & 0 \\ 0 & \sigma_r^2 \Sigma_i \end{pmatrix}$, $\Sigma_0^* = \begin{pmatrix} \sigma_{b0}^2 & 0 \\ 0 & \sigma_{r0}^2 \Sigma_i \end{pmatrix}$ where Σ_i is the kinship matrix, then by moment generating function, (S1.6) turns to be

$$\lambda(0)^k \exp\Big\{\big(\sum_{j=1}^{k} e_{ij}\big)^T \Sigma^* \big(\sum_{j=1}^{k} e_{ij}\big)/2\Big\} = \lambda_0(0)^k \exp\Big\{\big(\sum_{j=1}^{k} e_{ij}\big)^T \Sigma_0^* \big(\sum_{j=1}^{k} e_{ij}\big)/2\Big\}. \tag{S1.7}$$

Condition (A.1) implies that $\lambda(0) > 0$. Note that the index set $\{1,...,k\}$ in (S1.7) can be replaced by any subset of $\{1,...,n_i\}$. Thus, $e_{ij}^T \Sigma^* e_{ij'} = e_{ij}^T \Sigma_0^* e_{ij'}$, $j \neq j'$, $j, j' = 1,...,n_i$. By (A.4) we obtain $\sigma_b^2 = \sigma_{b0}^2$, $\sigma_r^2 = \sigma_{r0}^2$, and $\lambda(0) = \lambda_0(0)$.

Let $\Delta_{i1} = 1$ in (S1.4) and integrate Y_{i1} from 0 to y, $y \in [0, \tau]$. For $j = 2,...,n_i$, we use the same argument to yield

$$\int \big\{1 + \alpha\Lambda(y)\exp(\beta_0 g_{i1} + \gamma_0^T X_{i1} + b_i + r_{i1})\big\}^{-\frac{1}{\alpha}} \psi(b_i, r_{i\cdot}; \sigma_0^2) db_i dr_{i\cdot}$$
$$= \int \big\{1 + \alpha\Lambda_0(y)\exp(\beta_0 g_{i1} + \gamma_0^T X_{i1} + b_i + r_{i1})\big\}^{-\frac{1}{\alpha}} \psi(b_i, r_{i\cdot}; \sigma_0^2) db_i dr_{i\cdot}. \tag{S1.8}$$

Let $t = \Lambda_0(y) \geq 0$ where $y \in [0, \tau]$, and $g(\cdot) \triangleq \Lambda \circ \Lambda_0^{-1}(\cdot) \geq 0$. Then (S1.8) can be written as

$$\int \left\{1 + \alpha g(t) \exp(\beta_0 g_{i1} + \gamma_0^T X_{i1} + b_i + r_{i1})\right\}^{-\frac{1}{\alpha}} \psi(b_i, r_{i\cdot}; \sigma_0^2) db_i dr_{i\cdot}$$

$$= \int \left\{1 + \alpha t \exp(\beta_0 g_{i1} + \gamma_0^T X_{i1} + b_i + r_{i1})\right\}^{-\frac{1}{\alpha}} \psi(b_i, r_{i\cdot}; \sigma_0^2) db_i dr_{i\cdot}. \quad \text{(S1.9)}$$

By definition, we know $g(0) = 0$, and g is a continuous and differentiable function. By taking first derivative with respect to t on both sides of (S1.9), then let $t = 0$, we obtain $g'(0) = 1$. Similarly, by doing the second derivative, third derivative and so on, we obtain $g^{(2)}(0) = g^{(3)}(0) = \cdots = 0$. Hence, we conclude that $g(t) = t$, $g'(t) = g^{(2)}(t) = \cdots = 0$, $t \geq 0$. That is, $\Lambda(\cdot) = \Lambda_0(\cdot)$. □

Lemma 2. *Suppose conditions (A.1)-(A.5) hold, for family i, the equation*

$$\dot{l}_\theta(\theta_0, \Lambda_0)[h_1] + \dot{l}_\Lambda(\theta_0, \Lambda_0)[h_2] = 0$$

almost surely implies $h_1 = 0$ and $h_2 \equiv 0$.

Proof. From condition (A.5), $\dot{l}_\theta(\theta_0, \Lambda_0)[h_1] + \dot{l}_\Lambda(\theta_0, \Lambda_0)[h_2] = 0$ implies that for each $G_{i\cdot} = g_{i\cdot}$,

$$\dot{l}_\theta(\theta_0, \Lambda_0)\big|_{G_{i\cdot}=g_{i\cdot}}[h_1] + \dot{l}_\Lambda(\theta_0, \Lambda_0)\big|_{G_{i\cdot}=g_{i\cdot}}[h_2] = 0. \quad \text{(S1.10)}$$

Thus,

$$
\begin{aligned}
0 = &\int h_{11}^T \dot{l}_{\sigma_0^2}(\theta_0, \Lambda_0) R_{1i}(\theta_0, \Lambda_0, b_i, r_{ij}) \psi(b_i, r_{i\cdot}; \sigma_0^2) db_i dr_{i\cdot} \\
&+ \int \sum_{j=1}^{n_i} \left\{ \Delta_{ij} - (1 + \alpha \Delta_{ij}) \Omega_{3ij}(\widetilde{\theta}_0, \Lambda_0) \right\} h_{12}^T Z_{ij} \\
&\qquad \times R_{1i}(\theta_0, \Lambda_0, b_i, r_{ij}) \psi(b_i, r_{i\cdot}; \sigma_0^2) db_i dr_{i\cdot} \\
&+ \int \sum_{j=1}^{n_i} \left\{ \Delta_{ij} h_2(Y_{ij}) - (1 + \alpha \Delta_{ij}) \Omega_{2ij}(\widetilde{\theta}_0, \Lambda_0) \int_0^{Y_{ij}} h_2(s) d\Lambda_0(s) \right\} \\
&\qquad \times R_{1i}(\theta_0, \Lambda_0, b_i, r_{ij}) \psi(b_i, r_{i\cdot}; \sigma_0^2) db_i dr_{i\cdot}, \qquad \text{(S1.11)}
\end{aligned}
$$

where $h_1 = (h_{11}^T, h_{12}^T)^T$ and

$$
R_{1i}(\theta_0, \Lambda_0, b_i, r_{ij}) = \prod_{j=1}^{n_i} \left\{ \lambda_0(Y_{ij}) \Omega_{2ij}(\widetilde{\theta}_0, \Lambda_0) \right\}^{\Delta_{ij}} \left\{ \Omega_{1ij}(\widetilde{\theta}_0, \Lambda_0) \right\}^{\alpha^{-1}}.
$$

We next show that (S1.11) establishes $h_1 = 0$ and $h_2 \equiv 0$. The argument is similar with that in the proof of Lemma 1.

First, we let $G_{ij} = g_{ij}$ and X_{ij} be fixed. Then for a fixed k, where $1 \le k \le n_i$, we define a sequence measures μ_1, ..., μ_{n_i} on $\{0,1\} \times [0,\tau]$ as follow. Let A be any Borel set from the Borel σ-field on $[0,\tau]$, then

$$
\begin{cases}
\mu_m(\{0\} \times A) = 0, \ \mu_m(\{1\} \times A) = I(0 \in A), & m \le k, \\
\mu_m(\{0\} \times A) = I(\tau \in A), \ \mu_m(\{1\} \times A) = \int I(x \in A) dx, & m > k.
\end{cases}
$$

We integrate both sides of (S1.11) over $\{(\Delta_{i,1}, Y_{i,1}), ..., (\Delta_{i,n_i}, Y_{i,n_i})\}$ with respect to $\prod_{m=1}^{n_i} \mu_m$. The integration we made here is actually equivalent to let $\Delta_{im} = 1$ and $Y_{im} = 0$ for all $m \leq k$, and let $Y_{im} = \tau$ if $\Delta_{im} = 0$ and integrate Y_{im} from 0 to τ if $\Delta_{im} = 1$ for $k \leq m \leq n_i$. The resulting integration with $k \leq m \leq n_i$ is actually 0. We sum all the equalities of (S1.11) for all possible combinations of $\{\Delta_{i,k+1}, ..., \Delta_{i,n_i}\}$.

For the first term on the left side of (S1.11), for any (b_i, r_{ij}), by the integration with respect to $d(\prod_{m=1}^{n_i} \mu_m)$, if $j \leq k$, then we have

$$\begin{aligned} &\int h_{11}^T \dot{l}_{\sigma_0^2}(\theta_0, \Lambda_0) R_{1i}(\theta_0, \Lambda_0, b_i, r_{ij}) d(\prod_{m=1}^{n_i} \mu_m) \\ = \;& h_{11}^T \dot{l}_{\sigma_0^2}(\theta_0, \Lambda_0) \prod_{m \leq k} \lambda_0(0) \exp(\beta_0 g_{ij} + \gamma_0^T X_{ij} + b_i + r_{ij}), \end{aligned}$$

if $j > k$, then

$$\begin{aligned} 0 &= \int h_{11}^T \dot{l}_{\sigma_0^2}(\theta_0, \Lambda_0) R_{1i}(\theta_0, \Lambda_0, b_i, r_{ij}) d(\prod_{m=1}^{n_i} \mu_m) \\ &= \prod_{m \leq k} \lambda_0(0) \exp(\beta_0 g_{ij} + \gamma_0^T X_{ij} + b_i + r_{ij}) \\ &\qquad \times \sum_{\Delta_{ij} \in \{0,1\}} \Big\{ (1 - \Delta_{ij}) h_{11}^T \dot{l}_{\sigma_0^2}(\theta_0, \Lambda_0) R_{1ij}(\theta_0, \Lambda_0(\tau), b_i, r_{ij}) \\ &\qquad\qquad + \Delta_{ij} \int_0^\tau h_{11}^T \dot{l}_{\sigma_0^2}(\theta_0, \Lambda_0) R_{1ij}(\theta_0, \Lambda_0(t), b_i, r_{ij}) dt \Big\}, \end{aligned}$$

where

$$R_{1ij}(\theta_0, \Lambda_0(t), b_i, r_{ij}) = \{\lambda_0(t) \Omega_{2ij}(\widetilde{\theta}_0, \Lambda_0)\}^{\Delta_{ij}} \big\{\Omega_{1ij}(\widetilde{\theta}_0, \Lambda_0)\big\}^{1/\alpha} I(Y_{ij} \geq t).$$

Hence,

$$\int\int h_{11}^T \dot{l}_{\sigma_0^2}(\theta_0,\Lambda_0) R_{1i}(\theta_0,\Lambda_0,b_i,r_{ij})\psi(b_i,r_{i\cdot};\sigma_0^2)db_i dr_{i\cdot} d(\prod_{m=1}^{n_i}\mu_m)$$
$$= \int h_{11}^T \dot{l}_{\sigma_0^2}(\theta_0,\Lambda_0)\prod_{m\le k}\lambda_0(0)\exp(\beta_0 g_{im}+\gamma_0^T X_{im}+b_i+r_{im})\psi(b_i,r_{i\cdot};\sigma_0^2)db_i dr_{i\cdot}. \tag{S1.12}$$

Let $\widetilde{Z}_{ij}=(g_{ij},X_{ij}^T)^T$, then using a similar argument, we have

$$\int\int \sum_{j=1}^{n_i}\{\Delta_{ij}-(1+\alpha\Delta_{ij})\Omega_{3ij}(\widetilde{\theta}_0,\Lambda_0)\}h_{12}^T\widetilde{Z}_{ij}$$
$$\times R_{1i}(\theta_0,\Lambda_0,b_i,r_{ij})\psi(b_i,r_{i\cdot};\sigma_0^2)db_i dr_{i\cdot} d(\prod_{m=1}^{n_i}\mu_m)$$
$$= \int h_{12}^T \sum_{j=1}^{k}\widetilde{Z}_{ij}\prod_{m\le k}\lambda_0(0)\exp(\beta_0 g_{im}+\gamma_0^T X_{im}+b_i+r_{im})\psi(b_i,r_{i\cdot};\sigma_0^2)db_i dr_{i\cdot}, \tag{S1.13}$$

and

$$\int\int \sum_{j=1}^{n_i}\left\{\Delta_{ij}h_2(Y_{ij})-\int_0^{Y_{ij}}(1+\alpha\Delta_{ij})\Omega_{2ij}(\widetilde{\theta}_0,\Lambda_0)h_2(s)d\Lambda_0(s)\right\}$$
$$\times R_{1i}(\theta_0,\Lambda_0,b_i,r_{ij})\psi(b_i,r_{i\cdot};\sigma_0^2)db_i dr_{i\cdot} d(\prod_{m=1}^{n_i}\mu_m)$$
$$= \sum_{j\le k}h_2(0)\int\prod_{m\le k}\lambda_0(0)\exp(\beta_0 g_{im}+\gamma_0^T X_{im}+b_i+r_{im})\psi(b_i,r_{i\cdot};\sigma_0^2)db_i dr_{i\cdot}. \tag{S1.14}$$

Combine (S1.12), (S1.13) and (S1.14), and after integrating over b_i and r_{ij}, we obtain that

$$\frac{1}{2}\Big(\sum_{j=1}^{k} e_{ij}\Big)^T \Sigma_{h_{11}} \Big(\sum_{j=1}^{k} e_{ij}\Big) + \sum_{j=1}^{k} h_{12}^T \widetilde{Z}_{ij} + k h_2(0) = 0,$$

where $\Sigma_{h_{11}} = \begin{pmatrix} h_{111} & 0_{1\times n_i} \\ 0_{n_i\times 1} & h_{112}\Sigma_i \end{pmatrix}$, $h_{11} = (h_{111}, h_{112})^T$. Since the subscript of j is arbitrary, for any $1 \le k_1 < k_2 \le n_i$, we have

$$\frac{1}{2}\Big(\sum_{j=k_1}^{k_2} e_{ij}\Big)^T \Sigma_{h_{11}} \Big(\sum_{j=k_1}^{k_2} e_{ij}\Big) + \sum_{j=k_1}^{k_2} h_{12}^T \widetilde{Z}_{ij} + (k_2 - k_1) h_2(0) = 0,$$

hence it leads to $\frac{1}{2} e_{ij}^T \Sigma_{h_{11}} e_{ij} + h_{12}^T \widetilde{Z}_{ij} + h_2(0) = 0$ and $e_{ij}^T \Sigma_{h_{11}} e_{ij'} = 0$, $j \neq j'$. Again by (A.4), we obtain that $h_{11} = 0$, $h_{12} = 0$, hence $h_1 = 0$.

In equation (S1.11), we let $Y_{ij} = 0$, $j = 2, ..., n_i$ and $\Delta_{ij} = 1$, $j = 1, ..., n_i$, then we have

$$\begin{aligned} h_2(Y_{i1}) = & \int_0^{Y_{i1}} h_2(y) d\Lambda_0(y) \\ & \times \int \frac{(1+\alpha)\exp(\beta_0 G_{i1} + \gamma_0^T X_{i1} + b_i + r_{i1})}{\big\{1 + \alpha\Lambda_0(Y_{i1})\exp(\beta_0 G_{i1} + \gamma_0^T X_{i1} + b_i + r_{i1})\big\}^{\alpha^{-1}+2}} \\ & \quad \times \exp\Big(\sum_{j=2}^{n_i} \beta_0 G_{ij} + \gamma_0^T X_{ij} + b_i + r_{ij}\Big)\psi(b_i, r_{i\cdot}; \sigma_0^2) db_i dr_{i\cdot} \\ & \times \Big[\int \frac{\exp(\sum_{j=1}^{n_i} \beta_0 G_{ij} + \gamma_0^T X_{ij} + b_i + r_{ij})\psi(b_i, r_{i\cdot}; \sigma_0^2)}{\big\{1 + \alpha\Lambda_0(Y_{i1})\exp(\beta_0 G_{i1} + \gamma_0^T X_{i1} + b_i + r_{i1})\big\}^{\alpha^{-1}+1}} db_i dr_{i\cdot}\Big]^{-1}. \end{aligned} \tag{S1.15}$$

Let $g^*(t) = \int_0^t h_2(s) d\Lambda_0(s)$, then (S1.15) can be written as a homogeneous ordinary differential equation

$$\frac{1}{\lambda_0(t)}\frac{dg^*(t)}{dt}$$
$$= g^*(t)\int \frac{(1+\alpha)\exp(\sum_{j=1}^{n_i}\beta_0 G_{ij} + \gamma_0^T X_{ij} + b_i + r_{ij})\psi(b_i, r_{i\cdot}; \sigma_0^2)}{\{1+\alpha\Lambda_0(Y_{i1})\exp(\beta_0 G_{i1} + \gamma_0^T X_{i1} + b_i + r_{i1})\}^{\alpha^{-1}+2}} db_i dr_{i\cdot}$$
$$\times \Big[\int \frac{\exp(\sum_{j=1}^{n_i}\beta_0 G_{ij} + \gamma_0^T X_{ij} + b_i + r_{ij})\psi(b_i, r_{i\cdot}; \sigma_0^2)}{\{1+\alpha\Lambda_0(Y_{i1})\exp(\beta_0 G_{i1} + \gamma_0^T X_{i1} + b_i + r_{i1})\}^{\alpha^{-1}+1}} db_i dr_{i\cdot}\Big]^{-1}$$

with boundary condition $g^*(0) = 0$. By solving this equation, we obtain $g^*(t) \equiv 0$, $t \in [0, \tau]$. Hence, $h_2(\cdot) \equiv 0$ by (S1.15). The proof is completed.

□

S2 Additional Simulation I

In the additional simulation study, we mimic the real application setting, where we consider only one confounder 'Education'. We generate the observations of education by randomly sampling the standardized values of education from the real data. We simulate genotype with the same frequency as APOE-ϵ4. The coefficient for the genotype is 0.930 and for the education is -0.687. The transformation model corresponds to $\alpha = 2$ and the baseline hazard rate is set to be a constant so that the survival probability at year 20 (age 70 is year 0) is 20% (i.e., $\Lambda_0(t) = t(0.2^{-\alpha} - 1)/(20\alpha)$). We generate censoring time from the uniform distribution so that the censoring rate is around 80%. We generate 300 and 500 families respectively with heterogeneous pedigree structures randomly sampled from 1705 families in real data, where 1013 families contain only proband and the remaining 692 families contain only parents and/or parents with different number of siblings. We set $\sigma_b^2 = 0.132$ and $\sigma_r^2 = 0.224$. The simulation is conducted with 500 replicates.

We summarize the estimated coefficients and hazard rates in Table S.1 comparing our method and the regular transformation method using proband data only. The results show that the proposed method still performs reasonably well using proband and relatives data even under heavy censoring scenario. While the results with proband data only are less accurate and efficient, especially the estimate of genetic effect. In addition, we also report the estimated α based on maximizing the profile likelihood on a grid-search-point set $\{1.0, 1.2, 1.5, 1.8, 2.0, 2.2, 2.5, 3.0\}$ in Table S.2. The results reveal that the proposed profile likelihood method for α works well using both proband and relatives data. The semiparmetric transformation model for only proband data is a misspecified model but still flexible. However, the estimated α using proband data only is no longer around the true value.

Table S.1: Simulation results of the additional simulation I

			Proband with relatives				Proband only			
n	Par	True	Bias	SD	SE	CP%	Bias	SD	SE	CP%
300	β	0.930	−0.017	0.344	0.336	94.6	−0.078	0.366	0.245	82.5
	γ	−0.687	−0.010	0.153	0.153	95.3	0.055	0.158	0.118	84.3
	$\Lambda(\tau/4)$	0.120	0.003	0.025	0.028	96.1	0.006	0.029	0.025	91.3
	$\Lambda(\tau/2)$	0.240	0.002	0.045	0.048	96.0	0.001	0.055	0.042	88.2
	σ_b^2	0.132	−0.040	0.343	0.389	89.6	——	——	——	——
	σ_r^2	0.224	0.030	0.141	0.232	96.8	——	——	——	——
500	β	0.930	0.004	0.264	0.260	94.7	−0.098	0.282	0.188	79.6
	γ	−0.687	−0.009	0.120	0.120	94.8	0.039	0.128	0.091	81.6
	$\Lambda(\tau/4)$	0.120	−0.004	0.019	0.023	95.6	0.012	0.023	0.020	88.8
	$\Lambda(\tau/2)$	0.240	−0.005	0.036	0.039	95.4	0.012	0.051	0.033	84.3
	σ_b^2	0.132	−0.027	0.193	0.212	93.7	——	——	——	——
	σ_r^2	0.224	0.019	0.125	0.140	95.7	——	——	——	——

S3 Additional Simulation II

We perform more simulation studies for the special case of $\sigma_r^2 = 0$, and compare the results with that of the approach without considering the polygenic heterogeneity. Note that when $\sigma_r^2 = 0$ the proposed transformation model reduces to the frailty model. We consider the setup of Case I scenario with $\alpha = 0$. Other setup is similar to the simulation in Section 4.

Table S.2: Counts of the estimated αs in additional simulation I

		α							
n	Proband	1.0	1.2	1.5	1.8	2.0	2.2	2.5	3.0
300	with relatives	67	16	23	29	276	22	46	21
	proband only	240	39	38	31	21	22	29	80
500	with relatives	38	15	17	33	282	36	56	23
	proband only	185	40	46	35	29	24	31	110

The simulation results are summarized in Table S.3. Overall, the proposed method performs equally well with the frailty method. Despite that the biases by the proposed method are slightly bigger than that by the frailty method and the SDs are slightly smaller than that by the frailty method, but the values of Bias and SD by the two methods are approximately around the same level. Particularly, because the 95% confidence interval (CI) of σ_r^2 is constructed using the Satterthwaite approximation, the lower bounds of CIs are always positive. Therefor, the empirical CP corresponding to σ_r^2 always equals to constant 0. To address this problem, we adjust the 95% CI by forcing those lower bounds below some small number, say 0.05, to 0. In specific, when $n = 300$ the empirical mean, SD, SE and CP of the estimated σ_r^2 by adjustment are 0.101, 0.093, 0.127 and 85.5% respectively, and when $n = 500$ the corresponding results are 0.099,

Table S.3: Simulation results of the additional simulation II with Case I scenario

			Proposed Model				Frailty Model			
n	Par	True	Bias	SD	SE	CP%	Bias	SD	SE	CP%
300	σ_b^2	0.25	−0.031	0.097	0.097	96.6	0.000	0.090	0.085	94.2
	β	0.50	−0.028	0.285	0.277	94.8	−0.010	0.284	0.269	94.6
	γ	−0.50	−0.035	0.096	0.098	95.2	−0.016	0.093	0.093	95.2
	$\Lambda(\tau/4)$	0.75	0.030	0.172	0.167	94.6	0.027	0.169	0.166	93.8
	$\Lambda(\tau/2)$	1.50	0.089	0.336	0.334	96.4	0.053	0.321	0.318	93.4
500	σ_b^2	0.25	−0.029	0.070	0.074	95.8	−0.002	0.063	0.065	95.6
	β	0.50	−0.009	0.225	0.211	94.5	−0.002	0.222	0.208	94.0
	γ	−0.50	−0.031	0.074	0.076	95.3	−0.011	0.071	0.072	95.2
	$\Lambda(\tau/4)$	0.75	0.007	0.135	0.128	94.3	0.012	0.133	0.126	94.0
	$\Lambda(\tau/2)$	1.50	0.064	0.267	0.253	95.5	0.027	0.256	0.241	93.6

0.068, 0.100 and 92.5%.

References

Murphy, S. A. (1995). Asymptotic Theory for the Frailty Model. *Ann. Statist.* **23**, 182-198.

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

Zeng, D. and Lin, D. Y. (2010). A general asymptotic theory for maximum likelihood estimation in semiparametric regression models with censored data. *Statist. Sinica* **20**, 871-910.