Supplementary Material for "Time-to-event Analysis with Unknown Time Origins via Longitudinal Biomarker Registration"

Tianhao Wang, Sarah J. Ratcliffe, and Wensheng Guo*

November 29, 2021

Introduction

In this supplementary material we provide the proofs of model identifiability and the theorems of the paper. We denote $Pf = \int f(x)dP(x)$, $\mathbb{P}_n f = \frac{1}{n}\sum_{i=1}^n f(X_i)$ and $\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n - P)f$. We use the symbol $\lesssim$ to denote that the left hand side is bounded above by a constant times the right hand side, $\gtrsim$ to denote that the left hand side is bounded below by a constant times the right hand side, and $\asymp$ to denote that both $\lesssim$ and $\gtrsim$ apply. We use $|X|$ to denote the absolute value of X if X is a scalar, or the square-root of the largest eigenvalue of $XX^\top$ if X is a vector or matrix.

Let $\mathcal{K} = \{\tau_1, ..., \tau_K\}$ be a set of partition points of $[0,1]$ with $\max_{1<j\le K} |\tau_j - \tau_{j-1}| = O(K^{-1})$. Let $\mathcal{S}(\mathcal{K}, p)$ denote the space of polynomial splines of order $p \ge 1$ with the knots sequence $\mathcal{K}$ as defined in the Definition 4.1 of Schumaker (1981). Let $\mathcal{K}_\mu = \{\tau_1, ..., \tau_{K_\mu}\}$ be a set of partition points of $[0,1]$ with $\max_{1<j\le K_\mu} |\tau_j - \tau_{j-1}| = O(K_\mu^{-1})$; and $\mathcal{K}_\psi = \{t_1, ..., t_{K_\psi}\}$ denote a set of partition points of $[a,b]$ with $\max_{1<j\le K_\psi} |t_j - t_{j-1}| = O(K_\psi^{-1})$. Define $\mathcal{S}(\mathcal{K}_\mu, p_\mu)$ and $\mathcal{S}(\mathcal{K}_\psi, p_\psi)$ similarly as $\mathcal{S}(\mathcal{K}, p)$. For notation consistency, let $\mathcal{K}_\beta = \mathcal{K}$ and $p_\beta = p$. Define $\mathcal{B}^\beta = \mathcal{S}(\mathcal{K}_\beta, p_\beta)$, and the sieve spaces $\mathcal{B}_n^\mu = \mathcal{S}(\mathcal{K}_\mu, p_\mu)$, $\mathcal{G}_n^\psi = \mathcal{S}(\mathcal{K}_\psi, p_\psi)$,

*Tianhao Wang is Assistant Professor, Department of Neurological Sciences, and Faculty Statistician, Rush Alzheimer's Disease Center, Rush University Medical Center, Chicago, IL 60612 (tianhao_wang@rush.edu); Sarah J. Ratcliffe is Professor, Division of Biostatistics, Department of Public Health Sciences, University of Virginia School of Medicine, Charlottesville, VA 22908 (sratcliffe@virginia.edu); and Wensheng Guo is Professor, Department of Biostatistics, Epidemiology and Informatics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA 19104 (wguo@pennmedicine.upenn.edu). This research was supported by National Institutes of Health grants R01–GM104470 and R01–DK117208. The authors are grateful to the Editor, Associate Editor, and three referees for their insightful comments that greatly improved the paper.

and $\Theta_n = \Xi \times \mathcal{B}^\beta \times \mathcal{G}_n^\psi \times \mathcal{B}_n^\mu \times \mathcal{M}^K$. The original parameter space is denoted as $\Theta = \Xi \times \mathcal{B}^\beta \times \mathcal{G}^\psi \times \mathcal{B}^\mu \times \mathcal{M}^K$.

Working assumption for outcome-dependent follow-up

Following the idea of Lipsitz et al. (2002), we present a working assumption under which we can ignore the possible outcome-dependent follow-up.

Let $D_{ik} = t_{ik} - t_{i,k-1}$, $k = 2, ..., m_i$, be the time between follow-up times $k-1$ and k, and

$$\mathbf{D}_i = (D_{i2}, ..., D_{i,m_i}).$$

We assume that the conditional distribution of D_{ik}, given the observed history of longitudinal outcomes $\boldsymbol{y}_i$, time-independent variable $\boldsymbol{z}_i$, and the survival data (r_i, δ_i), depends only on the previously observed longitudinal outcomes $(y_{i1}, ..., y_{i,k-1})$ and the time-independent variable $\boldsymbol{z}_i$, i.e.,

$$\begin{aligned} f(D_{ik}|D_{i2}, ..., D_{i,k-1}, \boldsymbol{y}_i, r_i, \delta_i, \boldsymbol{z}_i) &= f(D_{ik}|D_{i2}, ..., D_{i,k-1}, y_{i1}, ..., y_{i,k-1}, \boldsymbol{z}_i) \\ &= f(D_{ik}|y_{i1}, ..., y_{i,k-1}, \boldsymbol{z}_i), \end{aligned}$$

for $k = 2, ..., m_i$. For example, under this assumption, women with a history of fast dilation on previous measurements $(y_{i1}, ..., y_{i,k-1})$ may be expected to have smaller values of D_{ik}. Although m_i is also a random variable, its value is completely determined by $\mathbf{D}_i$ and r_i, and thus can be neglected in the following discussion. The joint density function of the observed data $(\mathbf{D}_i, \boldsymbol{y}_i, r_i, \delta_i)$ for the ith invidual can be written as

$$f(\mathbf{D}_i, \boldsymbol{y}_i, r_i, \delta_i, \boldsymbol{z}_i) = \left\{ \prod_{k=2}^{m_i} f(D_{ik}|y_{i1}, ..., y_{i,k-1}, \boldsymbol{z}_i) \right\} f(\boldsymbol{y}_i, r_i, \delta_i|\boldsymbol{z}_i) f(\boldsymbol{z}_i).$$

We further assume that $\left\{\prod_{k=2}^{m_i} f(D_{ik}|y_{i1}, ..., y_{i,k-1}, \boldsymbol{z}_i)\right\}$ and $f(\boldsymbol{z}_i)$ are free of the parameters of interest. Therefore, the joint likelihood function reduces to be $f(\boldsymbol{y}_i, r_i, \delta_i|\boldsymbol{z}_i)$ which is the likelihood function considered in Section 3.2.

Identifiability

Proof of identifiability is similar to that of Gervini and Gasser (2004). We need to prove that if there exists (μ, g_i) and (μ^*, g_i^*) satisfying the above assumptions, such that

$$\mu(g_i^{-1}(t_{ij})) = \mu^*(g_i^{*-1}(t_{ij})) \text{ with probability 1,} \tag{1}$$

then $\mu \equiv \mu^*$ and $g_i \equiv g_i^*$.

Suppose

$$\tau_{ij} = g_i^{-1}(t_{ij}), \text{ i.e., } t_{ij} = g_i(\tau_{ij}).$$

Then if the equation (1) holds, we have

$$\mu(\tau_{ij}) = \mu^*(g_i^{*-1}(g_i(\tau_{ij}))) \text{ with probability 1.}$$

By the B-spline model assumption that $g_i(\tau) = g(\tau; \boldsymbol{u}_i)$ with $\boldsymbol{u}_i$ ranging from $-\infty$ to ∞ with positive density, we have $\tau_{ij} = g_i^{-1}(t_{ij})$ has a positive density function in $(0, 1)$. Thus,

$$\mu(\tau) = \mu^*(g_i^{*-1}(g_i(\tau))), \forall \tau \in (0, 1).$$

By condition (C.7) in the main paper, the local extrema of μ are isolated points because μ is piecewise monotone without flat areas. Further, since the left hand side does not depend on i and by condition (C.7), the right hand side should not neither, i.e., there exists a fixed function $h(\tau)$, such that

$$g_i^{*-1}(g_i(\tau)) = h(\tau), \text{ i.e., } g_i(\tau) = g_i^*(h(\tau)).$$

Note that the mean registration function, denoted by $E[g_i(\tau)] = \nu_0(\tau)$, is a strictly increasing function as ensured by $E(\boldsymbol{u}_i) = \mathbf{0}$. We have

$$\nu_0(\tau) = E[g_i(\tau)] = E[g_i^*(h(\tau))] = \nu_0(h(\tau)),$$

i.e., $h(\tau) = \tau$. Therefore, we have proved that $\mu(\tau) = \mu^*(\tau)$ and $g_i(\tau) = g_i^*(\tau)$. $\square$

Proof of Theorem 1

Denote by $\boldsymbol{O}_i := (r_i, \delta_i, Z_i, \boldsymbol{y}_i, \boldsymbol{t}_i)$ the ith observation. By the definition of $\hat{\boldsymbol{\theta}}_n$ and model assumptions about $\boldsymbol{\theta}_0$, we have

$$\begin{aligned} \hat{\boldsymbol{\theta}}_n &= \arg\max_{\boldsymbol{\theta} \in \Theta_n} \mathbb{P}_n \left\{ \ell(\boldsymbol{\theta}; \boldsymbol{O}_i) \right\}, \\ \boldsymbol{\theta}_0 &= \arg\max_{\boldsymbol{\theta} \in \Theta} P \left\{ \ell(\boldsymbol{\theta}; \boldsymbol{O}_i) \right\}. \end{aligned}$$

The convergence rate of $d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0)$ is obtained by applying Theorem 1 of Shen and Wong (1994). We need to verify the following three conditions:

A1. For some constants $A_1 > 0$ and $\alpha_1 > 0$, and for all small $\varepsilon > 0$,

$$\inf_{\{d(\boldsymbol{\theta},\boldsymbol{\theta}_0)\geq\epsilon,\boldsymbol{\theta}\in\Theta_n\}} P(\ell(\boldsymbol{\theta}_0) - \ell(\boldsymbol{\theta})) \geq 2A_1\epsilon^{2\alpha_1}.$$

A2. For some constants $A_2 > 0$ and $\alpha_2 > 0$, and for all small $\varepsilon > 0$,

$$\sup_{\{d(\boldsymbol{\theta},\boldsymbol{\theta}_0)\leq\epsilon,\boldsymbol{\theta}\in\Theta_n\}} \mathrm{var}(\ell(\boldsymbol{\theta}_0) - \ell(\boldsymbol{\theta})) \leq 2A_2\epsilon^{2\alpha_2}.$$

A3. Let $\mathcal{F}_n = \{\ell(\boldsymbol{\theta}) - \ell(\pi_n\boldsymbol{\theta}_0), \boldsymbol{\theta} \in \Theta_n\}$, where $\pi_n\boldsymbol{\theta}_0$ is the projection of $\boldsymbol{\theta}_0$ in Θ_n. For some constants $r_0 < \frac{1}{2}$ and $A_3 > 0$,

$$H(\epsilon, \mathcal{F}_n, \|\cdot\|_\infty) \leq A_3 n^{2r_0}\epsilon^{-r}$$

for all small $\epsilon > 0$, where $H(\epsilon, \mathcal{F}_n, \|\cdot\|_\infty)$ is the L_∞-metric entropy of the space $\mathcal{F}_n$, i.e., $\exp(H(\epsilon, \mathcal{F}_n, \|\cdot\|_\infty))$ is the number of ϵ-balls in the L_∞-metric needed to cover the space $\mathcal{F}_n$.

The above three conditions are checked via the Fréchet derivatives of $\ell(\boldsymbol{\theta}; \boldsymbol{O}_i)$, which are calculated by the Gateaux derivatives. For any fixed $\beta \in \mathcal{B}^\beta$, $\mu \in \mathcal{B}^\mu$ and $\psi \in \mathcal{G}^\psi$, let β_Δ, μ_Δ and ψ_Δ respectively be smooth curves in $\mathcal{B}^\beta$, $\mathcal{B}^\mu$ and $\mathcal{G}^\psi$ running through β, μ and ψ at $\Delta = 0$. Define tangent spaces

$$\begin{aligned}
\mathcal{H}_\beta &= \left\{h_\beta :\ h_\beta = \frac{\partial\beta_\Delta}{\partial\Delta}\Big|_{\eta=0},\ \beta_\Delta \in \mathcal{B}^\beta\right\},\\
\mathcal{H}_\mu &= \left\{h_\mu :\ h_\mu = \frac{\partial\mu_\Delta}{\partial\Delta}\Big|_{\Delta=0},\ \mu_\Delta \in \mathcal{B}^\mu\right\},\\
\mathcal{H}_\psi &= \left\{h_\psi :\ h_\psi = \frac{\partial\psi_\Delta}{\partial\Delta}\Big|_{\Delta=0},\ \psi_\Delta \in \mathcal{G}^\psi\right\},
\end{aligned}$$

and $\mathcal{H}_{\boldsymbol{\theta}} = \{\boldsymbol{h}_{\boldsymbol{\theta}} = (\boldsymbol{h}_{\boldsymbol{\xi}}, h_\beta, h_\psi, h_\mu, \boldsymbol{h}_\Sigma) \in \Theta : \|\boldsymbol{h}_{\boldsymbol{\xi}}\|, \|h_\beta\|_{\Sigma_0}, \|h_\psi\|_{\Lambda_0}, \|h_\mu\|_{F_0}, \|\boldsymbol{h}_\Sigma\| \leq M_0 < \infty\}$. The one-dimensional submodel along $\boldsymbol{h}_{\boldsymbol{\theta}} = (\boldsymbol{h}_{\boldsymbol{\xi}}, h_\beta, h_\psi, h_\mu, h_\Sigma)$ can be expressed as $\boldsymbol{\theta}_\Delta = \boldsymbol{\theta} + \Delta\boldsymbol{h}_{\boldsymbol{\theta}} = (\boldsymbol{\xi} + \Delta\boldsymbol{h}_{\boldsymbol{\xi}}, \beta + \Delta h_\beta, \psi + \Delta h_\psi, \mu + \Delta h_\mu, \Sigma + \Delta\boldsymbol{h}_\Sigma)$. The Gateaux derivatives are then calculated by differentiating $\ell(\boldsymbol{\theta}_\Delta; \boldsymbol{O}_i)$ with respect to Δ and letting $\Delta = 0$. Let α and υ represent any two parameters in $(\boldsymbol{\xi}, \Sigma, \beta, \mu, \psi)$. Denote by $\dot{\ell}_\alpha(\boldsymbol{\theta}; \boldsymbol{O}_i)$ and $\ddot{\ell}_{\alpha\upsilon}(\boldsymbol{\theta}; \boldsymbol{O}_i)$ the first and second order derivatives of $\ell(\boldsymbol{\theta}; \boldsymbol{O}_i)$, respectively.

By the definition of $\boldsymbol{\theta}_0$ and model assumptions, we have $P\left\{\dot{\ell}(\boldsymbol{\theta}_0; \boldsymbol{O}_i)\right\} = 0$. For Condition A1, given that all the second derivatives of $\ell(\boldsymbol{\theta}_0; \boldsymbol{O}_i)$ are continuous and uniformly bounded, the Taylor's expansion gives

$$
\begin{aligned}
& 2\left[P\left\{\ell(\boldsymbol{\theta}_0;\boldsymbol{O}_i)\right\}-P\left\{\ell(\boldsymbol{\theta};\boldsymbol{O}_i)\right\}\right] \\
= \; & -P\Big\{(\boldsymbol{\xi}-\boldsymbol{\xi}_0)^\top\ddot{\ell}_{\boldsymbol{\xi}\boldsymbol{\xi}}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)(\boldsymbol{\xi}-\boldsymbol{\xi}_0)+\ddot{\ell}_{\Sigma\Sigma}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\Sigma-\Sigma_0,\Sigma-\Sigma_0] \\
& +2(\boldsymbol{\xi}-\boldsymbol{\xi}_0)^\top\ddot{\ell}_{\boldsymbol{\xi}\Sigma}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\Sigma-\Sigma_0]+2(\boldsymbol{\xi}-\boldsymbol{\xi}_0)^\top\ddot{\ell}_{\boldsymbol{\xi}\mu}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\mu(\cdot)-\mu_0(\cdot)] \\
& +2(\boldsymbol{\xi}-\boldsymbol{\xi}_0)^\top\ddot{\ell}_{\boldsymbol{\xi}\beta}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\beta-\beta_0]+2(\boldsymbol{\xi}-\boldsymbol{\xi}_0)^\top\ddot{\ell}_{\boldsymbol{\xi}\psi}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\psi-\psi_0] \\
& +2\ddot{\ell}_{\Sigma\beta}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\Sigma-\Sigma_0,\beta-\beta_0]+2\ddot{\ell}_{\Sigma\mu}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\Sigma-\Sigma_0,\mu-\mu] \\
& +2\ddot{\ell}_{\Sigma\psi}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\Sigma-\Sigma_0,\psi-\psi_0]+\ddot{\ell}_{\beta\beta}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\beta-\beta_0,\beta-\beta_0] \\
& +2\ddot{\ell}_{\beta\mu}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\beta-\beta_0,\mu-\mu_0]+2\ddot{\ell}_{\beta\psi}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\beta-\beta_0,\psi-\psi_0] \\
& +\ddot{\ell}_{\mu\mu}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\mu-\mu_0,\mu-\mu_0]+2\ddot{\ell}_{\mu\psi}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\mu-\mu_0,\psi-\psi_0]\Big\} \\
& +\ddot{\ell}_{\psi\psi}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\psi-\psi_0,\psi-\psi_0]+o(d^2(\boldsymbol{\theta},\boldsymbol{\theta}_0)).
\end{aligned}
$$

By the fact of zero-mean for a score function, it is straightforward to verify that

$$
\begin{aligned}
P\left\{\ddot{\ell}_{\boldsymbol{\xi}\boldsymbol{\xi}}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)\right\} &= -P\left\{\dot{\ell}_{\boldsymbol{\xi}}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)\dot{\ell}_{\boldsymbol{\xi}}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)^\top\right\}, \\
P\left\{\ddot{\ell}_{\boldsymbol{\xi}\alpha}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[h_\alpha]\right\} &= -P\left\{\dot{\ell}_{\boldsymbol{\xi}}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)\dot{\ell}_{\alpha}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[h_\alpha]\right\}, \\
P\left\{\ddot{\ell}_{\alpha\upsilon}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[h_\alpha,h_\upsilon]\right\} &= -P\left\{\dot{\ell}_{\alpha}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[h_\alpha]\dot{\ell}_{\upsilon}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[h_\upsilon]\right\},
\end{aligned}
$$

where α and υ represent any two parameters in (β,ψ,μ,Σ). Then it follows from direct calculations that

$$
\begin{aligned}
& 2\left[P\left\{\ell(\boldsymbol{\theta}_0;\boldsymbol{O}_i)\right\}-P\left\{\ell(\boldsymbol{\theta};\boldsymbol{O}_i)\right\}\right] \\
= \; & P\Big\{\Big[(\boldsymbol{\xi}-\boldsymbol{\xi}_0)^\top\dot{\ell}_{\xi}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)+\dot{\ell}_{\beta}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\beta-\beta_0]+\dot{\ell}_{\psi}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\psi-\psi_0] \\
& +\dot{\ell}_{\mu}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\mu-\mu_0]+\dot{\ell}_{\Sigma}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\Sigma-\Sigma_0]\Big]^2\Big\}+o(d^2(\boldsymbol{\theta},\boldsymbol{\theta}_0)). \qquad (2)
\end{aligned}
$$

Define

$$
\begin{aligned}
\Psi[\boldsymbol{\theta}-\boldsymbol{\theta}_0] = \; & P\Big\{\Big[(\boldsymbol{\xi}-\boldsymbol{\xi}_0)^\top\dot{\ell}_{\boldsymbol{\xi}}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)+\dot{\ell}_{\beta}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\beta-\beta_0]+\dot{\ell}_{\psi}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\psi-\psi_0] \\
& +\dot{\ell}_{\mu}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\mu-\mu_0]+\dot{\ell}_{\Sigma}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\Sigma-\Sigma_0]\Big]^2\Big\},
\end{aligned}
$$

which is a real-valued functional operator on the tangent space $\mathcal{H}_{\boldsymbol{\theta}}$ endowed with the norm

$$
\|\boldsymbol{h}\|_\Theta=\left\{\|\boldsymbol{h}_{\boldsymbol{\xi}}\|^2+\|h_\beta\|^2_{\Sigma_0}+\|h_\psi\|^2_{\Lambda_0}+\|h_\mu\|^2_{F_0}+\|\boldsymbol{h}_\Sigma\|^2\right\}^{1/2},
$$

for $\boldsymbol{h} = (\boldsymbol{h}_{\boldsymbol{\xi}}, h_\beta, h_\psi, h_\mu, \boldsymbol{h}_\Sigma) \in \mathcal{H}_{\boldsymbol{\theta}}$.

We claim that there exist a constant $A_1 > 0$, such that

$$\Psi[\boldsymbol{\theta} - \boldsymbol{\theta}_0] \geq A_1 d^2(\boldsymbol{\theta}, \boldsymbol{\theta}_0),\ \forall \boldsymbol{\theta} \in \Theta \text{ and } d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) > 0,$$

i.e., the Conditon A1 is satisfied with $\alpha_1 = 1$ since $\Theta_n \subset \Theta$. Otherwise, we can find a sequence $\boldsymbol{\theta}^{[k]} \in \Theta,\ k = 1, 2, ...,$ where $d(\boldsymbol{\theta}^{[k]}, \boldsymbol{\theta}_0) > 0$, such that

$$\Psi[\boldsymbol{\theta}^{[k]} - \boldsymbol{\theta}_0] / d^2(\boldsymbol{\theta}^{[k]}, \boldsymbol{\theta}_0) \to 0,\ \text{as } k \to \infty.$$

Define

$$\boldsymbol{h}^{[k]} = (\boldsymbol{\theta}^{[k]} - \boldsymbol{\theta}_0)/d(\boldsymbol{\theta}^{[k]}, \boldsymbol{\theta}_0) = (\boldsymbol{\theta}^{[k]} - \boldsymbol{\theta}_0)/\|\boldsymbol{\theta}^{[k]} - \boldsymbol{\theta}_0\|_\Theta.$$

We have $\|\boldsymbol{h}^{[k]}\|_\Theta \equiv 1$ and $\Psi[\boldsymbol{\theta}^{[k]} - \boldsymbol{\theta}_0]/d^2(\boldsymbol{\theta}^{[k]}, \boldsymbol{\theta}_0) = \Psi[\boldsymbol{h}^{[k]}]$. It is equivalent to say that we can find a sequence $\boldsymbol{h}^{[k]} \in \mathcal{H}_{\boldsymbol{\theta}},\ k = 1, 2, ...,$ such that $\|\boldsymbol{h}^{[k]}\|_\Theta \equiv 1$ and

$$\Psi[\boldsymbol{h}^{[k]}] \to 0.$$

We can reach a contradiction if we can prove the following three statements:

[S1.1] $\{\boldsymbol{h} : \|\boldsymbol{h}\|_\Theta \equiv 1\}$ is a compact subset of $\mathcal{H}_{\boldsymbol{\theta}}$;

[S1.2] $\Psi[\boldsymbol{h}]$ is a compact operator of $\boldsymbol{h}$ with respect to $\|\cdot\|_\Theta$ norm;

[S1.3] $\Psi[\boldsymbol{h}] = 0$ implies $\|\boldsymbol{h}\|_\Theta = 0$.

Statement S1.1 implies that we can find a convergent subsequence of $\boldsymbol{h}^{[k]}$, denoted by $\boldsymbol{h}^{[k_l]}$, such that there exits $\boldsymbol{h}^{[0]} \in \{\boldsymbol{h} : \|\boldsymbol{h}\|_\Theta \equiv 1\}$ and $\lim_{l\to\infty} \|\boldsymbol{h}^{[k_l]} - \boldsymbol{h}^{[0]}\|_\Theta = 0$. Statement S1.2 implies that

$$\Psi[\boldsymbol{h}^{[0]}] = \lim_{l\to\infty} \Psi[\boldsymbol{h}^{[k_l]}] = 0.$$

By statement S1.3, we have $\|\boldsymbol{h}^{[0]}\|_\Theta = 0$ from which we reach the contradiction.

S1.1 is proved by Bolzano-Weierstrass Theorem. S1.2 is proved by noting that $\Psi[\boldsymbol{h}]$ maps $\boldsymbol{h}$ into a finite-dimensional space, namely $\mathbb{R}$, and that $\psi(\cdot)$, $\dot{\psi}(\cdot)$, $\beta(\cdot)$ and $\mu(\cdot)$ are uniformly bounded. Now we verify S1.3. The equation $\Psi[\boldsymbol{h}] = 0$ implies that, with probability 1,

$$\begin{aligned} &\boldsymbol{h}_{\boldsymbol{\xi}}^\top \dot{\ell}_{\boldsymbol{\xi}}(\boldsymbol{\theta}_0; \boldsymbol{O}_i) + \dot{\ell}_\beta(\boldsymbol{\theta}_0; \boldsymbol{O}_i)[h_\beta] + \dot{\ell}_\psi(\boldsymbol{\theta}_0; \boldsymbol{O}_i)[h_\psi] \\ &\quad + \dot{\ell}_\mu(\boldsymbol{\theta}_0; \boldsymbol{O}_i)[h_\mu] + \dot{\ell}_\Sigma(\boldsymbol{\theta}_0; \boldsymbol{O}_i)[\boldsymbol{h}_\Sigma] = 0. \end{aligned} \tag{3}$$

By definition, we have

$$\begin{aligned}
\ell(\boldsymbol{\theta}_0;\boldsymbol{O}_i) &= \log\int_{\boldsymbol{u}_i} f_S(r_i,\delta_i|Z_i,\boldsymbol{u}_i;\boldsymbol{\theta}_S^{[0]})f_y(\boldsymbol{y}_i|\boldsymbol{t}_i,\boldsymbol{u}_i;\boldsymbol{\theta}_y^{[0]})f_u(\boldsymbol{u}_i;\Sigma_0)d\boldsymbol{u}_i \\
&= \log\int_{\boldsymbol{u}_i} \exp\{\ell_S(\boldsymbol{\theta}_S^{[0]};r_i,\delta_i,Z_i,\boldsymbol{u}_i)+\ell_y(\boldsymbol{\theta}_y^{[0]};\boldsymbol{y}_i,\boldsymbol{t}_i,\boldsymbol{u}_i)+\ell_u(\Sigma_0;\boldsymbol{u}_i)\}d\boldsymbol{u}_i, \quad (4)
\end{aligned}$$

where

$$\begin{aligned}
\ell_S(\boldsymbol{\theta}_S^{[0]};r_i,\delta_i,Z_i,\boldsymbol{u}_i) &= \delta_i\log\lambda_0\Big(H(r_i-g(\tau_0;\boldsymbol{u}_i))-\int_0^1 g(\tau;\boldsymbol{u}_i)\beta(\tau)d\tau-\boldsymbol{\xi}^\top Z_i\Big) \\
&\quad -\Lambda_0\Big(H(r_i-g(\tau_0;\boldsymbol{u}_i))-\int_0^1 g(\tau;\boldsymbol{u}_i)\beta(\tau)d\tau-\boldsymbol{\xi}^\top Z_i\Big) \\
&= \delta_i\psi_0\Big(H(r_i-g(\tau_0;\boldsymbol{u}_i))-\int_0^1 g(\tau;\boldsymbol{u}_i)\beta(\tau)d\tau-\boldsymbol{\xi}^\top Z_i\Big) \\
&\quad -\int_a^b I\Big(s\le H(r_i-g(\tau_0;\boldsymbol{u}_i))-\int_0^1 g(\tau;\boldsymbol{u}_i)\beta(\tau)d\tau-\boldsymbol{\xi}^\top Z_i\Big)\exp\{\psi_0(s)\}ds, \\
\ell_y(\boldsymbol{\theta}_y;\boldsymbol{y}_i,\boldsymbol{t}_i,\boldsymbol{u}_i) &= -\frac{1}{2}(\boldsymbol{y}_i-\mu_0\{g^{-1}(\boldsymbol{t}_i;\boldsymbol{u}_i)\})^\top(\boldsymbol{y}_i-\mu_0\{g^{-1}(\boldsymbol{t}_i;\boldsymbol{u}_i)\}), \\
\ell_u(\Sigma_u;\boldsymbol{u}_i) &= -\frac{1}{2}\log|\Sigma_0|-\frac{1}{2}\boldsymbol{u}_i^\top\Sigma_0^{-1}\boldsymbol{u}_i.
\end{aligned}$$

For simplicity of notations, define $\tilde{r}_i=\tilde{r}_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)=H(r_i-g(\tau_0;\boldsymbol{u}_i))-\int_0^1 g(\tau;\boldsymbol{u}_i)\beta(\tau)d\tau-\boldsymbol{\xi}^\top Z_i$, $\ell_S(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)=\ell_S(\boldsymbol{\theta}_S^{[0]};r_i,\delta_i,Z_i,\boldsymbol{u}_i)$ and $\ell_y(\boldsymbol{\theta}_y;\boldsymbol{u}_i)=\ell_y(\boldsymbol{\theta}_y;\boldsymbol{y}_i,\boldsymbol{t}_i,\boldsymbol{u}_i)$. It follows from direct calculations that

$$\dot{\ell}(\boldsymbol{\theta}_0;\boldsymbol{O}_i) = \frac{\int_{\boldsymbol{u}_i}\exp\{\ell_S(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)+\ell_y(\boldsymbol{\theta}_y;\boldsymbol{u}_i)+\ell_u(\Sigma_u;\boldsymbol{u}_i)\}\times\{\dot{\ell}_S+\dot{\ell}_y+\dot{\ell}_u\}d\boldsymbol{u}_i}{\int_{\boldsymbol{u}_i}\exp\{\ell_S(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)+\ell_y(\boldsymbol{\theta}_y;\boldsymbol{u}_i)+\ell_u(\Sigma_u;\boldsymbol{u}_i)\}d\boldsymbol{u}_i}, \quad (5)$$

where

$$\begin{aligned}
\dot{\ell}_{S,\boldsymbol{\xi}}(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i) &= -Z_i\Big[\delta_i\dot{\psi}_0(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i))-\int_a^b I(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\ge s)\exp\{\psi_0(s)\}\dot{\psi}_0(s)ds\Big], \\
\dot{\ell}_{S,\beta}(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)[h_\beta] &= -\int_0^1 h_\beta(\tau)g(\tau;\boldsymbol{u}_i)d\tau\Big[\delta_i\dot{\psi}_0(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)) \\
&\quad -\int_a^b I(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\ge s)\exp\{\psi_0(s)\}\dot{\psi}_0(s)ds\Big], \\
\dot{\ell}_{S,\boldsymbol{\psi}}(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)[h_\psi] &= \delta_i h_\psi(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i))-\int_a^b I(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\ge s)\exp\{\psi_0(s)\}h_\psi(s)ds,
\end{aligned}$$

$$\begin{aligned}
\dot{\ell}_{y,\mu}(\boldsymbol{\theta}_y;\boldsymbol{u}_i)[h_\mu] &= (\boldsymbol{y}_i-\mu_0\{g^{-1}(\boldsymbol{t}_i;\boldsymbol{u}_i)\})^\top h_\mu(g^{-1}(\boldsymbol{t}_i;\boldsymbol{u}_i)), \\
\dot{\ell}_{u,\Sigma}(\Sigma_0;\boldsymbol{u}_i)[\boldsymbol{h}_\Sigma] &= -\frac{1}{2}\mathrm{trace}\{(\Sigma_0^{-1}-\Sigma_0^{-1}\boldsymbol{u}_i\boldsymbol{u}_i^\top\Sigma_0^{-1})\boldsymbol{h}_\Sigma\}.
\end{aligned}$$

For any function of $\boldsymbol{u}_i$ denoted by $f(\boldsymbol{u}_i)$, define $\mathcal{L}_i\{f\} = \int_{\boldsymbol{u}_i} \exp\{\ell_S(\boldsymbol{\theta}_S^{[0]}; \boldsymbol{u}_i) + \ell_y(\boldsymbol{\theta}_y; \boldsymbol{u}_i) + \ell_u(\Sigma_u; \boldsymbol{u}_i)\} \times f(\boldsymbol{u}_i) d\boldsymbol{u}_i$. Then equation (3) can be written as

$$
\begin{aligned}
& \mathcal{L}_i\Big\{\boldsymbol{h}_{\boldsymbol{\xi}}^\top \dot{\ell}_{S,\boldsymbol{\xi}} + \dot{\ell}_{S,\beta}[\boldsymbol{h}_\beta] + \dot{\ell}_{S,\psi}[h_\psi] + \dot{\ell}_{y,\mu}[h_\mu] + \dot{\ell}_{u,\Sigma}[\boldsymbol{h}_\Sigma]\Big\} \\
= \;& \mathcal{L}_i\Big\{\boldsymbol{h}_{\boldsymbol{\xi}}^\top Z_i\Big[\delta_i \dot{\psi}_0(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]}; \boldsymbol{u}_i)) - \int_a^b I(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]}; \boldsymbol{u}_i) \geq s) \exp\{\psi_0(s)\}\dot{\psi}_0(s) ds\Big]\Big\} \\
+ \;& \mathcal{L}_i\Big\{\int_0^1 h_\beta(\tau) g(\tau; \boldsymbol{u}_i) d\tau \Big[\delta_i \dot{\psi}_0(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]}; \boldsymbol{u}_i)) - \int_a^b I(\tilde{r}_i \geq s) \exp\{\psi_0(s)\}\dot{\psi}_0(s) ds\Big]\Big\} \\
- \;& \mathcal{L}_i\Big\{\delta_i h_\psi(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]}; \boldsymbol{u}_i)) - \int_a^b I(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]}; \boldsymbol{u}_i) \geq s) \exp\{\psi_0(s)\} h_\psi(s) ds\Big\} \\
- \;& \mathcal{L}_i\Big\{\big(\boldsymbol{y}_i - \mu_0\{g^{-1}(\boldsymbol{t}_i; \boldsymbol{u}_i)\}\big) h_\mu(g^{-1}(\boldsymbol{t}_i; \boldsymbol{u}_i))\Big\} \\
+ \;& \mathcal{L}_i\Big\{\frac{1}{2}\mathrm{trace}\{(\Sigma_0^{-1} - \Sigma_0^{-1}\boldsymbol{u}_i\boldsymbol{u}_i^\top \Sigma_0^{-1})\boldsymbol{h}_\Sigma\}\Big\} = 0,
\end{aligned} \tag{6}
$$

where the denominator $\int_{\boldsymbol{u}_i} \exp\{\ell_S(\boldsymbol{\theta}_S^{[0]}; \boldsymbol{u}_i) + \ell_y(\boldsymbol{\theta}_y; \boldsymbol{u}_i) + \ell_u(\Sigma_u; \boldsymbol{u}_i)\} d\boldsymbol{u}_i$ is cancelled out.

In [S1.3] it is assumed that (6) is true with probability 1. We prove $\boldsymbol{h}_{\boldsymbol{\theta}} = 0$ by two-steps. In the first step, we prove $\|\boldsymbol{h}_\Sigma\| = \|h_\mu\|_{F_0} = 0$. Let $\delta_i = 1$. we can integrate over r_i on both sides of (6). By model assumption, we have

$$
\begin{aligned}
& \int_{r_i} \mathcal{L}_i\Big\{\boldsymbol{h}_{\boldsymbol{\xi}}^\top \dot{\ell}_{S,\boldsymbol{\xi}} + \dot{\ell}_{S,\beta}[\boldsymbol{h}_\beta] + \dot{\ell}_{S,\psi}[h_\psi]\Big\} dr_i \\
= \;& \int_{r_i}\int_{\boldsymbol{u}_i} \exp\{\ell_S(\boldsymbol{\theta}_S^{[0]}; \boldsymbol{u}_i) + \ell_y(\boldsymbol{\theta}_y; \boldsymbol{u}_i) + \ell_u(\Sigma_u; \boldsymbol{u}_i)\} \times \{\boldsymbol{h}_{\boldsymbol{\xi}}^\top \dot{\ell}_{S,\boldsymbol{\xi}} + \dot{\ell}_{S,\Sigma}[\boldsymbol{h}_\Sigma] + \dot{\ell}_{S,\psi}[h_\psi]\} d\boldsymbol{u}_i dr_i \\
= \;& \int_{\boldsymbol{u}_i} \Big[\int_{r_i} \exp\{\ell_S(\boldsymbol{\theta}_S^{[0]}; \boldsymbol{u}_i)\} \times \dot{\ell}_{S,\boldsymbol{\theta}_S^{[0]}}[\boldsymbol{h}_{\boldsymbol{\theta}_S^{[0]}}] dr_i\Big] \exp\{\ell_y(\boldsymbol{\theta}_y; \boldsymbol{u}_i) + \ell_u(\Sigma_u; \boldsymbol{u}_i)\} d\boldsymbol{u}_i \\
= \;& \int_{\boldsymbol{u}_i} E_{r_i}\Big[\dot{\ell}_{S,\boldsymbol{\theta}_S^{[0]}}[\boldsymbol{h}_{\boldsymbol{\theta}_S^{[0]}}]\Big| Z_i, \boldsymbol{u}_i\Big] \exp\{\ell_y(\boldsymbol{\theta}_y; \boldsymbol{u}_i) + \ell_u(\Sigma_u; \boldsymbol{u}_i)\} d\boldsymbol{u}_i = 0,
\end{aligned}
$$

and

$$
\begin{aligned}
& \int_{r_i} \mathcal{L}_i\Big\{\dot{\ell}_{y,\mu}[h_\mu] + \dot{\ell}_{u,\Sigma}[\boldsymbol{h}_\Sigma]\Big\} dr_i \\
= \;& \int_{r_i}\int_{\boldsymbol{u}_i} \exp\{\ell_S(\boldsymbol{\theta}_S^{[0]}; \boldsymbol{u}_i) + \ell_y(\boldsymbol{\theta}_y; \boldsymbol{u}_i) + \ell_u(\Sigma_u; \boldsymbol{u}_i)\} \times \{\dot{\ell}_{y,\mu}[h_\mu] + \dot{\ell}_{u,\Sigma}[\boldsymbol{h}_\Sigma]\} d\boldsymbol{u}_i dr_i \\
= \;& \int_{\boldsymbol{u}_i} \Big[\int_{r_i} \exp\{\ell_S(\boldsymbol{\theta}_S^{[0]}; \boldsymbol{u}_i)\} dr_i\Big] \exp\{\ell_y(\boldsymbol{\theta}_y; \boldsymbol{u}_i) + \ell_u(\Sigma_u; \boldsymbol{u}_i)\} \times \{\dot{\ell}_{y,\mu}[h_\mu] + \dot{\ell}_{u,\Sigma}[\boldsymbol{h}_\Sigma]\} d\boldsymbol{u}_i \\
= \;& \int_{\boldsymbol{u}_i} \exp\{\ell_y(\boldsymbol{\theta}_y; \boldsymbol{u}_i) + \ell_u(\Sigma_u; \boldsymbol{u}_i)\} \times \{\dot{\ell}_{y,\mu}[h_\mu] + \dot{\ell}_{u,\Sigma}[\boldsymbol{h}_\Sigma]\} d\boldsymbol{u}_i.
\end{aligned}
$$

Note that

$$
\begin{aligned}
0 &= \int_{\boldsymbol{u}_i} \exp\{\ell_y(\boldsymbol{\theta}_y;\boldsymbol{u}_i)+\ell_u(\Sigma_u;\boldsymbol{u}_i)\}\times\{\dot{\ell}_{y,\mu}[h_\mu]+\dot{\ell}_{u,\Sigma}[\boldsymbol{h}_\Sigma]\}d\boldsymbol{u}_i\\
&= \lim_{\Delta\to 0}\int_{\boldsymbol{u}_i}\exp\{\ell_y(\boldsymbol{\theta}_y+\Delta\boldsymbol{h}_{\boldsymbol{\theta}_y};\boldsymbol{u}_i)+\ell_u(\Sigma_0+\Delta\boldsymbol{h}_\Sigma;\boldsymbol{u}_i)\}d\boldsymbol{u}_i\\
&= \lim_{\Delta\to 0}\frac{1}{|\Sigma_0|^{1/2}}\int_{\boldsymbol{u}_i}\exp\Big\{-\frac{1}{2}\|\boldsymbol{y}_i-[\mu_0+\Delta h_\mu]\{g^{-1}(\boldsymbol{t}_i;\boldsymbol{u}_i)\}\}\|^2-\frac{1}{2}\boldsymbol{u}_i^\top(\Sigma_0+\Delta\boldsymbol{h}_\Sigma)^{-1}\boldsymbol{u}_i\Big\}d\boldsymbol{u}_i\\
&= \lim_{\Delta\to 0}\int_{\tilde{\boldsymbol{u}}_i}\exp\Big\{-\frac{1}{2}\|\boldsymbol{y}_i-[\mu_0+\Delta h_\mu]\{g^{-1}(\boldsymbol{t}_i;(\Sigma_0+\Delta\boldsymbol{h}_\Sigma)^{1/2}\tilde{\boldsymbol{u}}_i)\}\}\|^2-\frac{1}{2}\tilde{\boldsymbol{u}}_i^\top\tilde{\boldsymbol{u}}_i\Big\}d\tilde{\boldsymbol{u}}_i\\
&= \int_{\tilde{\boldsymbol{u}}_i}\exp\Big\{-\frac{1}{2}\|\boldsymbol{y}_i-\mu_0\{g^{-1}(\boldsymbol{t}_i;\Sigma_0^{1/2}\tilde{\boldsymbol{u}}_i)\}\}\|^2-\frac{1}{2}\tilde{\boldsymbol{u}}_i^\top\tilde{\boldsymbol{u}}_i\Big\}\\
&\quad\times\big(\boldsymbol{y}_i-\mu_0\{g^{-1}\big(\boldsymbol{t}_i;\Sigma_0^{1/2}\tilde{\boldsymbol{u}}_i)\}\big)^\top\Big[h_\mu(g^{-1}(\boldsymbol{t}_i;\Sigma_0^{1/2}\tilde{\boldsymbol{u}}_i))\\
&\qquad+\nabla\mu_0\{g^{-1}(\boldsymbol{t}_i;\Sigma_0^{1/2}\tilde{\boldsymbol{u}}_i)\}\frac{\partial(\Sigma_0+\Delta\boldsymbol{h}_\Sigma)^{1/2}\tilde{\boldsymbol{u}}_i}{\partial\Delta}\Big|_{\Delta=0}\Big]d\tilde{\boldsymbol{u}}_i,
\end{aligned}
$$

where

$$
\begin{aligned}
\nabla\mu_0\{g^{-1}(\boldsymbol{t}_i;\Sigma_0^{1/2}\tilde{\boldsymbol{u}}_i)\} &= \nabla_{\boldsymbol{u}_i}\mu_0\{g^{-1}(\boldsymbol{t}_i;\boldsymbol{u}_i)\}\big|_{\boldsymbol{u}_i=\Sigma_0^{1/2}\tilde{\boldsymbol{u}}_i},\\
\frac{\partial(\Sigma_0+\Delta\boldsymbol{h}_\Sigma)^{1/2}\tilde{\boldsymbol{u}}_i}{\partial\Delta}\Big|_{\Delta=0} &= (\tilde{\boldsymbol{u}}_i\otimes I_K)\Big[\Sigma_0^{1/2}\otimes I_K+I_K\otimes\Sigma_0^{1/2}\Big]\mathrm{vec}(\boldsymbol{h}_\Sigma)\\
&= \big(\Sigma_0^{1/2}\boldsymbol{h}_\Sigma+\boldsymbol{h}_\Sigma\Sigma_0^{1/2}\big)\tilde{\boldsymbol{u}}_i.
\end{aligned}
$$

Let $\boldsymbol{\eta}_i=\mu_0\{g^{-1}(\boldsymbol{t}_i;\Sigma_0^{1/2}\tilde{\boldsymbol{u}}_i)\}$. Define $\mu_0^+(\boldsymbol{\eta}_i;\boldsymbol{t}_i)=\arg\min_{\{\boldsymbol{\tau}_i:\mu_0(\boldsymbol{\tau}_i)=\boldsymbol{\eta}_i\}}\{\|\boldsymbol{\tau}_i-\boldsymbol{t}_i\|^2\}$, which is a generalized inverse of the function μ_0. It is guaranteed to exist and be unique by condition (C.7) that μ_0 is piecewise monotonic without flat areas. We can write $\tilde{\boldsymbol{u}}_i$ as a function of $\boldsymbol{t}_i$ and $\boldsymbol{\eta}_i$ denoted by $U(\boldsymbol{t}_i,\mu_0^+(\boldsymbol{\eta}_i;\boldsymbol{t}_i))$. Then we have

$$
\begin{aligned}
0 &= \int_{\boldsymbol{\eta}_i}\exp\{-\frac{1}{2}\|\boldsymbol{y}_i-\boldsymbol{\eta}_i\|^2-\frac{1}{2}\|U(\boldsymbol{t}_i,\mu_0^+(\boldsymbol{\eta}_i;\boldsymbol{t}_i))\|^2\}(\boldsymbol{y}_i-\boldsymbol{\eta}_i)^\top\times\Big[h_\mu(\mu_0^{-1}(\boldsymbol{\eta}_i))\\
&\quad+\nabla\mu_0\{g^{-1}(\boldsymbol{t}_i;\Sigma_0^{1/2}U(\boldsymbol{t}_i,\mu_0^+(\boldsymbol{\eta}_i;\boldsymbol{t}_i)))\}\big(\Sigma_0^{1/2}\boldsymbol{h}_\Sigma+\boldsymbol{h}_\Sigma\Sigma_0^{1/2}\big)U(\boldsymbol{t}_i,\mu_0^+(\boldsymbol{\eta}_i;\boldsymbol{t}_i))\Big]\\
&\quad\times|\nabla_{\boldsymbol{\eta}_i}U(\boldsymbol{t}_i,\mu_0^+(\boldsymbol{\eta}_i;\boldsymbol{t}_i))|d\boldsymbol{\eta}_i.
\end{aligned}
$$

Fixing $\boldsymbol{t}_i$, the above equation is a Fredholm Integral Equation of the first kind w. r. t. $\boldsymbol{\eta}_i$. It follows that

$$
\begin{aligned}
&\exp\{-\frac{1}{2}\|U(\boldsymbol{t}_i,\mu_0^+(\boldsymbol{\eta}_i;\boldsymbol{t}_i))\|^2\}|\nabla_{\boldsymbol{\eta}_i}U(\boldsymbol{t}_i,\mu_0^+(\boldsymbol{\eta}_i;\boldsymbol{t}_i))|\Big[h_\mu(\mu_0^{-1}(\boldsymbol{\eta}_i))\\
&\quad+\nabla\mu\{g^{-1}(\boldsymbol{t}_i;\Sigma_0^{1/2}U(\boldsymbol{t}_i,\mu_0^+(\boldsymbol{\eta}_i;\boldsymbol{t}_i)))\}\big(\Sigma_0^{1/2}\boldsymbol{h}_\Sigma+\boldsymbol{h}_\Sigma\Sigma_0^{1/2}\big)U(\boldsymbol{t}_i,\mu_0^+(\boldsymbol{\eta}_i;\boldsymbol{t}_i))\Big]\\
&= 0,\ \text{a.s.},
\end{aligned}
$$

i.e.

$$
\begin{aligned}
& -h_\mu(\mu_0^{-1}(\boldsymbol{\eta}_i)) \\
= \quad & \nabla\mu\{g^{-1}(\boldsymbol{t}_i;\Sigma_0^{1/2}U(\boldsymbol{t}_i,\mu_0^{+}(\boldsymbol{\eta}_i;\boldsymbol{t}_i)))\}\big(\Sigma_0^{1/2}\boldsymbol{h}_\Sigma+\boldsymbol{h}_\Sigma\Sigma_0^{1/2}\big)U(\boldsymbol{t}_i,\mu_0^{+}(\boldsymbol{\eta}_i;\boldsymbol{t}_i)),
\end{aligned}
$$

almost surely for $\boldsymbol{\eta}_i$ and $\boldsymbol{t}_i$. The LHS of the above equation is independent of $\boldsymbol{t}_i$. The matrix $\nabla\mu_0\{g^{-1}(\boldsymbol{t}_i;\Sigma_0^{1/2}U(\boldsymbol{t}_i,\mu_0^{+}(\boldsymbol{\eta}_i;\boldsymbol{t}_i)))\}$ is non-degenerate due to the model identifiability. It is straightforward to verify that the RHS of the above equation is a constant function of $\boldsymbol{t}_i$ if and only if $\boldsymbol{h}_\Sigma=0$, i.e., $\|\boldsymbol{b}_\Sigma\|=0$. Then we also have $h_\mu(\mu_0^{-1}(\boldsymbol{\eta}_i))=0$ a.s., i.e., $\|h_\mu\|_{F_0}=0$.

In the second step, we prove $\|\boldsymbol{h}_{\boldsymbol{\xi}}\|=\|h_\beta\|_{\Sigma_0}=\|h_\psi\|_{\Lambda_0}=0$. By a similar argument as in the first step, we can prove that

$$
\begin{aligned}
& \boldsymbol{h}_{\boldsymbol{\xi}}^\top Z_i\Big[\delta_i\dot\psi_0(\tilde r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i))-\int_a^b I(\tilde r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\ge s)\exp\{\psi_0(s)\}\dot\psi_0(s)ds\Big] \\
+ \quad & \int_0^1 h_\beta(\tau)g(\tau;\boldsymbol{u}_i)d\tau\Big[\delta_i\dot\psi_0(\tilde r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i))-\int_a^b I(\tilde r_i\ge s)\exp\{\psi_0(s)\}\dot\psi_0(s)ds\Big] \\
- \quad & \delta_i h_\psi(\tilde r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i))-\int_a^b I(\tilde r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\ge s)\exp\{\psi_0(s)\}h_\psi(s)ds\Big\}=0
\end{aligned}
$$

almost surely for $\boldsymbol{u}_i$, r_i and Z_i. For any $Z_1\neq Z_2$, we can find $\boldsymbol{u}_1$, r_1 and $\boldsymbol{u}_2$, r_2 such that $\tilde r_1(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_1)=\tilde r_2(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_2)$. Since $E(Z_iZ_i^\top)$ is non-singular by condition (C.1), we have

$$
\boldsymbol{h}_{\boldsymbol{\xi}}^\top\Big[\delta_i\dot\psi_0(\tilde r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i))-\int_a^b I(\tilde r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\ge s)\exp\{\psi_0(s)\}\dot\psi_0(s)ds\Big]=0
$$

almost surely. By model assumption, $\Big[\delta_i\dot\psi_0(\tilde r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i))-\int_a^b I(\tilde r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\ge s)\exp\{\psi_0(s)\}\dot\psi_0(s)ds\Big]$ is not constant zero. Thus $\boldsymbol{h}_{\boldsymbol{\xi}}=\boldsymbol{0}$. Now we have

$$
\begin{aligned}
& \int_0^1 h_\beta(\tau)g(\tau;\boldsymbol{u}_i)d\tau\Big[\delta_i\dot\psi_0(\tilde r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i))-\int_a^b I(\tilde r_i\ge s)\exp\{\psi_0(s)\}\dot\psi_0(s)ds\Big] \\
- \quad & \delta_i h_\psi(\tilde r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i))-\int_a^b I(\tilde r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\ge s)\exp\{\psi_0(s)\}h_\psi(s)ds\Big\}=0,
\end{aligned}
$$

i.e.,

$$
\int_0^1 h_\beta(\tau)g(\tau;\boldsymbol{u}_i)d\tau \quad = \quad \frac{\delta_i h_\psi(\tilde r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i))-\int_a^b I(\tilde r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\ge s)\exp\{\psi_0(s)\}h_\psi(s)ds}{\delta_i\dot\psi_0(\tilde r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i))-\int_a^b I(\tilde r_i\ge s)\exp\{\psi_0(s)\}\dot\psi_0(s)ds},
$$

almost surely. The LHS of the the above equation does not depend on r_i while the RHS is a constant function of r_i if and only if $h_\psi(\tilde r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i))=0$ almost surely, i.e., $\|h_\psi\|_{\Lambda_0}=0$. It follows that $\int_0^1 h_\beta(\tau)g(\tau;\boldsymbol{u}_i)d\tau=0$ almost surely, i.e., $\|h_\beta\|_{\Sigma_0}=0$. This completes the proof for statement [S1.3].

Next, we verify the Condition A2. Let $\tilde{\ell}(\boldsymbol{\theta};\boldsymbol{u}_i)=\ell_S(\boldsymbol{\theta}_S;\boldsymbol{u}_i)+\ell_y(\boldsymbol{\theta}_y;\boldsymbol{u}_i)+\ell_u(\Sigma_u;\boldsymbol{u}_i)$. It follows from direct calculations that

$$\begin{aligned}
[\ell(\boldsymbol{\theta};\boldsymbol{O}_i)-\ell(\boldsymbol{\theta}_0;\boldsymbol{O}_i)]^2 &= \left[\log\left(\frac{\int_{\boldsymbol{u}_i}\exp\{\tilde{\ell}(\boldsymbol{\theta};\boldsymbol{u}_i)\}d\boldsymbol{u}_i}{\int_{\boldsymbol{u}_i}\exp\{\tilde{\ell}(\boldsymbol{\theta}_0;\boldsymbol{u}_i)\}d\boldsymbol{u}_i}\right)\right]^2\\
&= \left[\log\left(1+\frac{\int_{\boldsymbol{u}_i}(\exp\{\tilde{\ell}(\boldsymbol{\theta};\boldsymbol{u}_i)\}-\exp\{\tilde{\ell}(\boldsymbol{\theta}_0;\boldsymbol{u}_i)\})d\boldsymbol{u}_i}{\int_{\boldsymbol{u}_i}\exp\{\tilde{\ell}(\boldsymbol{\theta}_0;\boldsymbol{u}_i)\}d\boldsymbol{u}_i}\right)\right]^2\\
&= \left[\frac{\int_{\boldsymbol{u}_i}(\exp\{\tilde{\ell}(\boldsymbol{\theta};\boldsymbol{u}_i)\}-\exp\{\tilde{\ell}(\boldsymbol{\theta}_0;\boldsymbol{u}_i)\})d\boldsymbol{u}_i}{\int_{\boldsymbol{u}_i}\exp\{\tilde{\ell}(\boldsymbol{\theta}_0;\boldsymbol{u}_i)\}d\boldsymbol{u}_i}\right]^2(1+o(1))\\
&= \left[\frac{\int_{\boldsymbol{u}_i}\exp\{\tilde{\ell}(\boldsymbol{\theta}_0;\boldsymbol{u}_i)\}\{\tilde{\ell}(\boldsymbol{\theta};\boldsymbol{u}_i)-\tilde{\ell}(\boldsymbol{\theta}_0;\boldsymbol{u}_i)\}d\boldsymbol{u}_i}{\int_{\boldsymbol{u}_i}\exp\{\tilde{\ell}(\boldsymbol{\theta}_0;\boldsymbol{u}_i)\}d\boldsymbol{u}_i}\right]^2(1+o(1))\\
&\le \frac{\int_{\boldsymbol{u}_i}\exp\{\tilde{\ell}(\boldsymbol{\theta}_0;\boldsymbol{u}_i)\}\{\tilde{\ell}(\boldsymbol{\theta};\boldsymbol{u}_i)-\tilde{\ell}(\boldsymbol{\theta}_0;\boldsymbol{u}_i)\}^2d\boldsymbol{u}_i}{\int_{\boldsymbol{u}_i}\exp\{\tilde{\ell}(\boldsymbol{\theta}_0;\boldsymbol{u}_i)\}d\boldsymbol{u}_i}(1+o(1)).
\end{aligned}$$

Let $\boldsymbol{O}_i^S=(r_i,\delta_i,Z_i)$ and $\boldsymbol{O}_i^y=(\boldsymbol{y}_i,\boldsymbol{t}_i)$. Then it follows that

$$\begin{aligned}
& P[\ell(\boldsymbol{\theta};\boldsymbol{O}_i)-\ell(\boldsymbol{\theta}_0;\boldsymbol{O}_i)]^2\\
\lesssim\ & \int_{\boldsymbol{O}_i}\int_{\boldsymbol{u}_i}\exp\{\tilde{\ell}(\boldsymbol{\theta}_0;\boldsymbol{u}_i)\}\{\tilde{\ell}(\boldsymbol{\theta};\boldsymbol{u}_i)-\tilde{\ell}(\boldsymbol{\theta}_0;\boldsymbol{u}_i)\}^2d\boldsymbol{u}_id\boldsymbol{O}_i\\
\lesssim\ & \int_{\boldsymbol{O}_i}\int_{\boldsymbol{u}_i}\exp\{\tilde{\ell}(\boldsymbol{\theta}_0;\boldsymbol{u}_i)\}\Big[\{\ell_S(\boldsymbol{\theta}_S;\boldsymbol{u}_i)-\ell_S(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\}^2\\
&\quad+\{\ell_y(\boldsymbol{\theta}_S;\boldsymbol{u}_i)-\ell_y(\boldsymbol{\theta}_y^{[0]};\boldsymbol{u}_i)\}^2+\{\ell_u(\Sigma;\boldsymbol{u}_i)-\ell_u(\Sigma_0;\boldsymbol{u}_i)\}^2\Big]d\boldsymbol{u}_id\boldsymbol{O}_i\\
=\ & \int_{\boldsymbol{O}_i^S}\int_{\boldsymbol{u}_i}\exp\{\ell_S(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)+\ell_u(\Sigma_0;\boldsymbol{u}_i)\}\{\ell_S(\boldsymbol{\theta}_S;\boldsymbol{u}_i)-\ell_S(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\}^2d\boldsymbol{u}_id\boldsymbol{O}_i^S\\
&+\int_{\boldsymbol{O}_i^y}\int_{\boldsymbol{u}_i}\exp\{\ell_y(\boldsymbol{\theta}_y^{[0]};\boldsymbol{u}_i)+\ell_u(\Sigma_0;\boldsymbol{u}_i)\}\{\ell_y(\boldsymbol{\theta}_y;\boldsymbol{u}_i)-\ell_y(\boldsymbol{\theta}_y^{[0]};\boldsymbol{u}_i)\}^2d\boldsymbol{u}_id\boldsymbol{O}_i^y\\
&+\int_{\boldsymbol{u}_i}\exp\{\ell_u(\Sigma_0;\boldsymbol{u}_i)\}\{\ell_u(\Sigma;\boldsymbol{u}_i)-\ell_u(\Sigma_0;\boldsymbol{u}_i)\}^2d\boldsymbol{u}_i\\
:=\ & I_r+I_y+I_u.
\end{aligned}$$

For I_r, by definition, we have

$$
\begin{aligned}
& \{\ell_S(\boldsymbol{\theta}_S;\boldsymbol{u}_i) - \ell_S(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\}^2 \\
= & \left|\delta_i\psi\left(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\right) - \int_a^b I\left(\tilde{\tilde{r}}_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i) \geq s\right)\exp\{\psi(s)\}ds\right. \\
& \left. - \delta_i\psi_0\left(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\right) + \int_a^b I\left(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i) \geq s\right)\exp\{\psi_0(s)\}ds\right|^2 \\
\leq & \delta_i\left|\psi\left(\tilde{r}_i(\boldsymbol{\theta}_S;\boldsymbol{u}_i)\right) - \psi\left(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\right)\right|^2 + \delta_i\left|\psi\left(\tilde{r}_i(\boldsymbol{\theta}_S;\boldsymbol{u}_i)\right) - \psi_0\left(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\right)\right|^2 \\
& + \left|\int_{\tilde{r}_i(\boldsymbol{\theta}_S;\boldsymbol{u}_i)}^{r_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)}\exp\{\psi_0(s)\}ds\right|^2 \\
\leq & \left[\left|\dot{\psi}^2\left(\tilde{r}_i(\boldsymbol{\theta}_S^{[1]};\boldsymbol{u}_i)\right)\right|^2 + \exp\{2\psi_0(\tilde{r}_i(\boldsymbol{\theta}_S^{[2]};\boldsymbol{u}_i))\}\right]\left|\tilde{r}_i(\boldsymbol{\theta}_S;\boldsymbol{u}_i) - \tilde{r}_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\right|^2 \\
& + \left|\psi\left(\tilde{r}_i(\boldsymbol{\theta}_S;\boldsymbol{u}_i)\right) - \psi_0\left(\tilde{r}_i(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i)\right)\right|^2 \\
:= & I_{r1} + I_{r2}
\end{aligned}
$$

where $\boldsymbol{\theta}_S^{[j]} = \kappa_j\boldsymbol{\theta}_S + (1-\kappa_j)\boldsymbol{\theta}_S^{[0]}$ for $j = 1, 2$ and $0 < \kappa_j < 1$. Note that $\dot{\psi}(\cdot)$ and $\psi_0(\cdot)$ are both uniformly bounded. It is straight forward to verify that

$$
\begin{aligned}
I_r &= \int_{\boldsymbol{O}_i^S}\int_{\boldsymbol{u}_i}\exp\{\ell_S(\boldsymbol{\theta}_S^{[0]};\boldsymbol{u}_i) + \ell_u(\Sigma_0;\boldsymbol{u}_i)\}(I_{r1} + I_{r2})d\boldsymbol{u}_i d\boldsymbol{O}_i^S \\
&\lesssim \|\boldsymbol{\xi} - \boldsymbol{\xi}_0\|^2 + \|\beta - \beta_0\|_{\Sigma_0}^2 + \|\psi - \psi_0\|_{\Lambda_0}^2.
\end{aligned}
$$

Similarly, we can show that $I_y \lesssim \|\mu - \mu_0\|_{F_0}^2$ and $I_u \lesssim \|\Sigma - \Sigma_0\|^2$, i.e.,

$$
\begin{aligned}
P[\ell(\boldsymbol{\theta};\boldsymbol{O}_i) - \ell(\boldsymbol{\theta}_0;\boldsymbol{O}_i)]^2 &\lesssim \|\boldsymbol{\xi} - \boldsymbol{\xi}_0\|^2 + \|\beta - \beta_0\|_{\Sigma_0}^2 + \|\psi - \psi_0\|_{\Lambda_0}^2 + \|\mu - \mu_0\|_{F_0}^2 + \|\Sigma - \Sigma_0\|^2 \\
&= \|\boldsymbol{\theta} - \boldsymbol{\theta}_0\|_\Theta^2.
\end{aligned}
$$

Thus, Condition A2 is satisfied with $\alpha_2 = 1$.

Finally, we verify the Condition A3. Define $\boldsymbol{\theta}_{n,0} = (\boldsymbol{\xi}_0, \beta_0, \psi_{n,0}, \mu_{n,0}, \Sigma_0) \in \Theta_n$. Similar to Condition A2, it can be proved that

$$
\begin{aligned}
|\ell(\boldsymbol{\theta};\boldsymbol{O}_i) - \ell(\boldsymbol{\theta}_{n,0};\boldsymbol{O}_i)| &\lesssim \|\boldsymbol{\xi} - \boldsymbol{\xi}_0\| + \|\beta - \beta_{n,0}\|_{\Sigma_0} + \|\psi - \psi_{n,0}\|_{\Lambda_0} \\
&\quad + \|\mu - \mu_{n,0}\|_{F_0} + \|\Sigma - \Sigma_0\| \\
&\leq \|\boldsymbol{\xi} - \boldsymbol{\xi}_0\| + \|\beta - \beta_{n,0}\|_{\Sigma_0} + \|\psi - \psi_{n,0}\|_\infty \\
&\quad + \|\mu - \mu_{n,0}\|_{F_0} + \|\Sigma - \Sigma_0\|,
\end{aligned}
$$

where $\boldsymbol{\xi} \in \Xi$, $\beta \in \mathcal{B}^\beta$, $\psi \in \mathcal{G}_n^\psi$, $\mu \in \mathcal{B}_n^\mu$ and $\Sigma \in \boldsymbol{\Sigma}$. By the calculations similar to Shen and Wong (1994), page 597, the ϵ-bracketing number of $\mathcal{G}_n^\psi$ and $\mathcal{B}_n^\mu$, with respect to the L_∞-norm, are bounded respectively by $(1/\epsilon)^{c_1(K_\mu+p_\mu+1)}$ and $(1/\epsilon)^{c_2(K_\psi+p_\psi+1)}$. It is easy to see that the ϵ-bracketing number of $\Xi \times \boldsymbol{\Sigma} \subset \mathbb{R}^{q+K}$ is bounded by $(1/\epsilon)^{q+K}$. Therefore, the ϵ-bracketing number of $\mathcal{F}_n = \{\ell(\boldsymbol{\theta}) - \ell(\pi_n\boldsymbol{\theta}_0), \boldsymbol{\theta} \in \Theta_n\}$ with respect to the L_∞-norm, denoted by $N(\epsilon, \mathcal{F}_n, \|\cdot\|_\infty)$, is bounded by $(1/\epsilon)^{c(K_\mu,K_\psi)}$, where $c(K_\mu, K_\psi) = q+K+c_1(K_\beta+p_\mu+1)+c_2(K_\psi+p_\psi+1)$. Note that $K_\mu = O(n^{v_\mu})$ and $K_\psi = O(n^{v_\psi})$. We have $c(K_\mu, K_\psi) \asymp n^{\max\{v_\psi,v_\mu\}}$. It follows that

$$\begin{aligned} H(\epsilon, \mathcal{F}_n, \|\cdot\|_\infty) &= \log N(\epsilon, \mathcal{F}_n, \|\cdot\|_\infty) \\ &\lesssim c(K_\mu, K_\psi)\log(1/\epsilon) \asymp n^{\max\{v_\psi,v_\mu\}}\log(1/\epsilon). \end{aligned}$$

Then the Condition A3 holds with constants $2r_0 = \max\{v_\psi, v_\mu\}$ and $r = 0^+$, where $\epsilon^{-0^+} := \log(1/\epsilon)$.

By Theorem 1 of Shen and Wong (1994), we have

$$d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_P(\max\{n^{-\tau}, d(\boldsymbol{\theta}_{n,0}, \boldsymbol{\theta}_0), D_{KL}(\boldsymbol{\theta}_{n,0}, \boldsymbol{\theta}_0)^{1/2}\}),$$

where $D_{KL}(\boldsymbol{\theta}_{n,0}, \boldsymbol{\theta}_0) = P(\ell(\boldsymbol{\theta}_{n,0}; \boldsymbol{O}) - \ell(\boldsymbol{\theta}_0; \boldsymbol{O}))$ and

$$\tau = \frac{1-2r_0}{2} - \frac{\log\log n}{2\log n}.$$

Since $\frac{\log\log n}{2\log n} \to 0$ as $n \to \infty$, we can choose a $\tilde{r}_0$ such that $\frac{1-2\tilde{r}_0}{2} \le \frac{1-2r_0}{2} - \frac{\log\log n}{2\log n}$ for n sufficiently large. We still write $\tilde{r}_0$ as r_0 and $\tau = \frac{1-2r_0}{2}$. By Corollary 6.21 in Schumaker (1981), we have

$$d(\boldsymbol{\theta}_{n,0}, \boldsymbol{\theta}_0) = O(n^{-\min\{p_\psi v_\psi, p_\mu v_\mu\}}).$$

Given $d(\boldsymbol{\theta}_{n,0}, \boldsymbol{\theta}_0)$, the Kullback-Leibler distance $D_{KL}(\boldsymbol{\theta}_{n,0}, \boldsymbol{\theta}_0)$ can be handled similarly as the proofs for Condition A2, which is bounded by $O(d^2(\boldsymbol{\theta}_{n,0}, \boldsymbol{\theta}_0))$, i.e.,

$$K^{1/2}(\boldsymbol{\theta}_{n,0}, \boldsymbol{\theta}_0) \lesssim d(\boldsymbol{\theta}_{n,0}, \boldsymbol{\theta}_0) = O(n^{-\min\{p_\psi v_\psi, p_\mu v_\mu\}}).$$

Thus, we obtain the convergence rate for $\hat{\boldsymbol{\theta}}_n$ as follows

$$d(\hat{\boldsymbol{\theta}}_n, \boldsymbol{\theta}_0) = O_P(n^{-\min\{p_\psi v_\psi, p_\mu v_\mu, (1-\max\{v_\psi,v_\mu\})/2\}}). \quad \square$$

Proof of Theorem 2:

Theorem 2 is proved by referring to Theorem 2.1 in Ding and Nan (2011). Define

$$\begin{aligned} \mathcal{H}_\mu &= \{h_\mu : \|h_\mu\|_\infty \le 1, h_\mu \in \mathcal{B}^\mu\}, \\ \mathcal{H}_\psi &= \{h_\psi : \|h_\psi\|_{\Lambda_0} \le 1, h_\psi \in \mathcal{G}^\psi\}. \end{aligned}$$

Since $\mathcal{B}_n^\mu \subset \mathcal{B}^\mu$, $\mathcal{G}_n^\psi \subset \mathcal{G}^\psi$, and by the consistency theorem, $\mathcal{H}_\mu^{p_\mu}$ and $\mathcal{H}_\psi^{p_\psi}$ are wide enough to cover the tangent sets induced by the sieve M-estimator. For simplicity of notations, we ignore the parameter Σ in the following discussions. Since $\mathcal{B}^\beta = \mathcal{S}(\mathcal{K}_\beta, p_\beta)$ is isomorphic to $\boldsymbol{B} \subset \mathbb{R}^K$ which is the set of B-spline coefficients for $\mathcal{B}^\beta$, they are used interchangeably when the context is clear. Let $\boldsymbol{\zeta} = (\boldsymbol{\xi}, \boldsymbol{b})$. We need to verify the following six conditions:

B1. (Rate of convergence) For an estimator $\hat{\boldsymbol{\theta}}_n = (\hat{\boldsymbol{\zeta}}_n, \hat{\mu}_n, \hat{\psi}_n) \in \Theta_n$ and the true parameter $\boldsymbol{\theta}_0 = (\boldsymbol{\zeta}_0, \mu_0, \psi_0) \in \Theta$, $d(\hat{\theta}_n, \theta_0) = O_P(n^{-\kappa})$ for some $\kappa > 0$.

B2. $P\dot{\ell}_{\boldsymbol{\theta}}(\boldsymbol{\theta}_0; \boldsymbol{O})[\boldsymbol{h}_{\boldsymbol{\theta}}] = 0$ for all $\boldsymbol{h}_{\boldsymbol{\theta}} \in \mathcal{H}_{\boldsymbol{\theta}} = \Xi \times \boldsymbol{B} \times \mathcal{H}_\mu \times \mathcal{H}_\psi$.

B3. (Positive information) There exists $\boldsymbol{h}_\mu^* = (h_{\mu,1}^*, ..., h_{\mu,d}^*)^\top$ and $\boldsymbol{h}_\psi^* = (h_{\psi,1}^*, ..., h_{\psi,d}^*)^\top$, where $h_{\mu,j}^* \in \mathcal{H}_\mu^{p_\mu}$ and $h_{\psi,j}^* \in \mathcal{H}_\psi^{p_\psi}$, for $j = 1, ..., q$, such that

$$\begin{aligned} P\left\{\ddot{\ell}_{\mu\boldsymbol{\zeta}}(\boldsymbol{\theta}_0; \boldsymbol{O})[h_\mu] - \ddot{\ell}_{\mu\mu}(\boldsymbol{\theta}_0; \boldsymbol{O})[\boldsymbol{h}_\mu^*, h_\mu]\right\} &= 0, \\ P\left\{\ddot{\ell}_{\psi\boldsymbol{\zeta}}(\boldsymbol{\theta}_0; \boldsymbol{O})[h_\psi] - \ddot{\ell}_{\psi\psi}(\boldsymbol{\theta}_0; \boldsymbol{O})[\boldsymbol{h}_\psi^*, h_\psi]\right\} &= 0, \end{aligned}$$

and for all $h_\mu \in \mathcal{H}_\mu^{p_\mu}$ and $h_\psi \in \mathcal{H}_\psi^{p_\psi}$. Furthermore, the matrix

$$\Omega_{\boldsymbol{\zeta}} = -P\Big\{\ddot{\ell}_{\boldsymbol{\zeta}\boldsymbol{\zeta}}(\boldsymbol{\theta}_0; \boldsymbol{O}) - \Big(\ddot{\ell}_{\mu\boldsymbol{\zeta}}(\boldsymbol{\theta}_0; \boldsymbol{O})[\boldsymbol{h}_\mu^*] + \ddot{\ell}_{\psi\boldsymbol{\zeta}}(\boldsymbol{\theta}_0; \boldsymbol{O})[\boldsymbol{h}_\psi^*]\Big)\Big\}$$

is nonsingular.

B4. The estimator $\hat{\boldsymbol{\theta}}_n$ satisfies $\mathbb{P}_n\{\dot{\ell}_{\boldsymbol{\zeta}}(\hat{\boldsymbol{\theta}}_n; \boldsymbol{O})\} = o_P(n^{-1/2})$ and

$$\begin{aligned} \mathbb{P}_n\{\dot{\ell}_\mu(\hat{\boldsymbol{\theta}}_n; \boldsymbol{O})[\boldsymbol{h}_\mu^*]\} &= o_P(n^{-1/2}), \\ \mathbb{P}_n\{\dot{\ell}_\psi(\hat{\boldsymbol{\theta}}_n; \boldsymbol{O})[\boldsymbol{h}_\psi^*]\} &= o_P(n^{-1/2}). \end{aligned}$$

B5. (Stochastic equi-continuity) For some $C > 0$,

$$\sup_{d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \le Cn^{-\kappa}, \boldsymbol{\theta} \in \Theta_n} \Big|\mathbb{G}_n\{\dot{\ell}_{\boldsymbol{\zeta}}(\boldsymbol{\theta}; \boldsymbol{O})\} - \mathbb{G}_n\{\dot{\ell}_{\boldsymbol{\zeta}}(\boldsymbol{\theta}_0; \boldsymbol{O})\}\Big| = o_P(1)$$

and

$$\begin{aligned} \sup_{d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \le Cn^{-\kappa}, \boldsymbol{\theta} \in \Theta_n} &\Big|\mathbb{G}_n\{\dot{\ell}_\mu(\boldsymbol{\theta}; \boldsymbol{O})[\boldsymbol{h}_\mu^*]\} - \mathbb{G}_n\{\dot{\ell}_\mu(\boldsymbol{\theta}_0; \boldsymbol{O})[\boldsymbol{h}_\mu^*]\}\Big| = o_P(1), \\ \sup_{d(\boldsymbol{\theta}, \boldsymbol{\theta}_0) \le Cn^{-\kappa}, \boldsymbol{\theta} \in \Theta_n} &\Big|\mathbb{G}_n\{\dot{\ell}_\psi(\boldsymbol{\theta}; \boldsymbol{O})[\boldsymbol{h}_\psi^*]\} - \mathbb{G}_n\{\dot{\ell}_\psi(\boldsymbol{\theta}_0; \boldsymbol{O})[\boldsymbol{h}_\psi^*]\}\Big| = o_P(1). \end{aligned}$$

B6. (Smoothness of the model) For some $a > 1$ satisfying $a\kappa > 1/2$, and for $\boldsymbol{\theta}$ in a neighborhood of $\boldsymbol{\theta}_0$: $\{\boldsymbol{\theta} : d(\boldsymbol{\theta},\boldsymbol{\theta}_0) \le Cn^{-\kappa}, \boldsymbol{\theta} \in \Theta_n\}$,

$$\Big| P\Big\{ \dot{\ell}_{\boldsymbol{\zeta}}(\boldsymbol{\theta};\boldsymbol{O}) - \dot{\ell}_{\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O}) - \ddot{\ell}_{\boldsymbol{\zeta}\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O})(\boldsymbol{\zeta}-\boldsymbol{\zeta}_0) - \ddot{\ell}_{\boldsymbol{\zeta}\mu}(\boldsymbol{\theta}_0;\boldsymbol{O})[\mu-\mu_0] - \ddot{\ell}_{\boldsymbol{\zeta}\psi}(\boldsymbol{\theta}_0;\boldsymbol{O})[\psi-\psi_0] \Big\} \Big| \lesssim d^a(\boldsymbol{\theta},\boldsymbol{\theta}_0),$$

$$\Big| P\{ \dot{\ell}_{\mu}(\boldsymbol{\theta};\boldsymbol{O})[\boldsymbol{h}_\mu^*] - \dot{\ell}_{\mu}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\mu^*] - \ddot{\ell}_{\mu\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\mu^*](\boldsymbol{\zeta}-\boldsymbol{\zeta}_0) - \ddot{\ell}_{\mu\mu}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\mu^*,\mu-\mu_0] - \ddot{\ell}_{\mu\psi}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\mu^*,\psi-\psi_0] \} \Big| \lesssim d^a(\boldsymbol{\theta},\boldsymbol{\theta}_0),$$

and

$$\Big| P\{ \dot{\ell}_{\psi}(\boldsymbol{\theta};\boldsymbol{O})[\boldsymbol{h}_\psi^*] - \dot{\ell}_{\psi}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\psi^*] - \ddot{\ell}_{\psi\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\psi^*](\boldsymbol{\zeta}-\boldsymbol{\zeta}_0) - \ddot{\ell}_{\psi\mu}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\psi^*,\mu-\mu_0] - \ddot{\ell}_{\psi\psi}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\psi^*,\psi-\psi_0] \} \Big| \lesssim d^a(\boldsymbol{\theta},\boldsymbol{\theta}_0).$$

By Theorem 1, B1 holds with $\kappa = \min\{p_\psi v_\psi, p_\mu v_\mu, (1-\max\{v_\psi, v_\mu\})/2\}$. Based on model assumptions, B2 holds automatically.

For B3, since $\ell(\boldsymbol{\theta};\boldsymbol{O})$ is the log-likelihood function, it follows that $\boldsymbol{h}_\psi^*$ and $\boldsymbol{h}_\mu^*$ are the least favorable directions. It is easy to see the directions $\boldsymbol{h}_\psi^*$ and $\boldsymbol{h}_\mu^*$ are respectively the minimizers of the following functional operators

$$\begin{aligned} \Omega_\psi[\boldsymbol{h}_\psi] &= P\{\|\dot{\ell}_{\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O}) - \dot{\ell}_\psi(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\psi]\|^2\}, \\ \Omega_\mu[\boldsymbol{h}_\mu] &= P\{\|\dot{\ell}_{\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O}) - \dot{\ell}_\mu(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\mu]\|^2\}, \end{aligned}$$

which respectively map $\mathcal{H}_\psi$ and $\mathcal{H}_\mu$ to $\mathbb{R}^+$. Then the existence of $\boldsymbol{h}_\psi^*$ and $\boldsymbol{h}_\mu^*$ in B3 is equivalent to the existence of minimizers of $\Omega_\psi[\boldsymbol{h}_\psi]$ and $\Omega_\mu[\boldsymbol{h}_\mu]$ on $\mathcal{H}_\psi$ and $\mathcal{H}_\mu$. Note that $\Omega_\psi[\boldsymbol{h}_\psi]$ and $\Omega_\mu[\boldsymbol{h}_\mu]$ are uniformly bounded below by 0, which implies that $\inf_{\boldsymbol{h}_\psi\in\mathcal{H}_\psi}\Omega_\psi[\boldsymbol{h}_\psi]$ and $\inf_{\boldsymbol{h}_\mu\in\mathcal{H}_\mu}\Omega_\mu[\boldsymbol{h}_\mu]$ exist and are bounded below by 0. Therefore, the existence of $\boldsymbol{h}_\psi^*$ and $\boldsymbol{h}_\mu^*$ can be proved by showing that $\mathcal{H}_\psi$ and $\mathcal{H}_\mu$ are compact sets, and that $\Omega_\psi[\boldsymbol{h}_\psi]$ and $\Omega_\mu[\boldsymbol{h}_\mu]$ are compact operators of $\boldsymbol{h}_\psi$ and $\boldsymbol{h}_\mu$, which can be proved similarly as the Condition A1 proved in Theorem 1. It follows from direct calculations that

$$\begin{aligned} \Omega_{\boldsymbol{\zeta}} &= -P\Big\{\ddot{\ell}_{\boldsymbol{\zeta}\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O}) - \Big(\ddot{\ell}_{g\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\psi^*] + \ddot{\ell}_{\mu\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\mu^*]\Big)\Big\} \\ &= -P\Big\{\ddot{\ell}_{\boldsymbol{\zeta}\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O}) + \ddot{\ell}_{\psi\psi}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\psi^*,\boldsymbol{h}_\psi^*] + \ddot{\ell}_{\mu\mu}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\mu^*,\boldsymbol{h}_\mu^*] \\ &\qquad + \ddot{\ell}_{\psi\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\psi^*] + \ddot{\ell}_{\mu\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\mu^*]\Big)\Big\} \\ &= P\Big\{\Big(\dot{\ell}_{\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O}) - \dot{\ell}_\psi(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\psi^*] - \dot{\ell}_\mu(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}_\mu^*]\Big)^{\otimes 2}\Big\}. \end{aligned}$$

By a similar argument that we used to prove the Condition A1 in Theorem 1, we can prove that $\Omega_{\boldsymbol{\zeta}}$ is non-singular, which is the information matrix for $\boldsymbol{\zeta}_0$.

For B4, the first equation $\mathbb{P}_n\{\dot{\ell}_{\boldsymbol{\zeta}}(\hat{\boldsymbol{\theta}}_n;\boldsymbol{O}_i\} = o_P(n^{-1/2})$ holds automatically by the definition of $\hat{\boldsymbol{\theta}}_n$ which maximizes $\mathbb{P}_n\{\ell(\hat{\boldsymbol{\theta}}_n;\boldsymbol{O}_i)\}$. Namely, the score equation for $\hat{\boldsymbol{\zeta}}_n$ is

$$\mathbb{P}_n\{\dot{\ell}_{\boldsymbol{\zeta}}(\hat{\boldsymbol{\theta}}_n;\boldsymbol{O}_i)\} = 0.$$

The remaining three equations are less obvious as $\boldsymbol{h}_{\psi}^*$ and $\boldsymbol{h}_{\mu}^*$ are only known to be in $\mathcal{H}_{\psi}$ and $\mathcal{H}_{\mu}$ which may not be the actual tangent sets of our sieve estimation with B-spline approximations. Fortunately, we can always approximate functions in $\mathcal{H}_{\psi}$ and $\mathcal{H}_{\mu}$ by B-splines with a decent accuracy. Since the proofs for the two equations are essentially the same, we only give the proof for

$$\mathbb{P}_n\{\dot{\ell}_{\psi}(\hat{\boldsymbol{\theta}}_n;\boldsymbol{O}_i)[\boldsymbol{h}_{\psi}^*]\} = o_P(n^{-1/2}).$$

The rest two equations can be proved similarly. According to Corollary 6.21 of Schumaker (1981), there exists an $h_{j,n}^* \in \mathcal{G}_n$ such that $\|h_{j,n}^* - h_{g,j}^*\|_\infty = O(n^{-p_\psi v_\psi})$. Note that in our sieve estimation the log-hazard rate function $\psi(t) = \log \lambda(t)$ is approximated by $\psi_n(t) = \Phi_n(t)^\top \boldsymbol{\gamma}$ where $\Phi_n(t)$ is the B-spline basis function for $\mathcal{G}_n^p$. Then by the score equation for $\boldsymbol{\gamma}$, we have

$$\begin{aligned}\mathbb{P}_n\dot{\ell}_{n,\boldsymbol{\gamma}}(\hat{\boldsymbol{\zeta}}_n,\hat{\gamma}_n,\hat{\boldsymbol{v}}_n;\boldsymbol{O}_i)\} &= \mathbb{P}_n\Bigg\{E^u_{i,\hat{\boldsymbol{\theta}}_n}\Bigg[\delta_i\Phi_n(\tilde{r}_i(\hat{\boldsymbol{\zeta}}_n;\boldsymbol{u}_i)) - \\ &\qquad \int_a^b I(\tilde{r}_i(\hat{\boldsymbol{\zeta}}_n,;\boldsymbol{u}_i)\geq s)\exp\{\hat{\psi}_n(s)\}\Phi_n(s)ds\Bigg]\Bigg\} \\ &= 0,\end{aligned}$$

where

$$E^u_{i,\boldsymbol{\theta}}\big[h(\boldsymbol{u}_i,\boldsymbol{O}_i)\big] = \frac{\int_{\boldsymbol{u}_i}\exp\{\ell_S(\boldsymbol{\theta}_S;\boldsymbol{u}_i)+\ell_y(\boldsymbol{\theta}_y;\boldsymbol{u}_i)+\ell_u(\Sigma_u;\boldsymbol{u}_i)\}\times h(\boldsymbol{u}_i,\boldsymbol{O}_i)d\boldsymbol{u}_i}{\int_{\boldsymbol{u}_i}\exp\{\ell_S(\boldsymbol{\theta}_S;\boldsymbol{u}_i)+\ell_y(\boldsymbol{\theta}_y;\boldsymbol{u}_i)+\ell_u(\Sigma_u;\boldsymbol{u}_i)\}d\boldsymbol{u}_i}.$$

Since $h_{j,n}^* \in \mathcal{G}_n^{\psi}$, we can write it as $h_{j,n}^*(t) = \Phi_n(t)^\top\boldsymbol{\gamma}_{j,n}^*$, i.e.,

$$\begin{aligned}0 &= \mathbb{P}_n\Bigg\{E^u_{i,\hat{\boldsymbol{\theta}}_n}\Bigg[\delta_i h_{j,n}^*(\tilde{r}_i(\hat{\boldsymbol{\zeta}}_n;\boldsymbol{u}_i)) - \int_a^b I(\tilde{r}_i(\hat{\boldsymbol{\zeta}}_n;\boldsymbol{u}_i)\geq s)\exp\{\hat{\psi}_n(s)\}h_{j,n}^*(s)ds\Bigg]\Bigg\} \\ &= \mathbb{P}_n\dot{\ell}_{\psi}(\hat{\boldsymbol{\theta}}_n;\boldsymbol{O}_i)[h_{j,n}^*]\end{aligned}$$

for $j = 1,...,q$. Thus it suffices to show that for each $j = 1,...,q$,

$$I_n = \mathbb{P}_n\dot{\ell}_{\psi}(\hat{\boldsymbol{\theta}}_n;\boldsymbol{O}_i)[h_{\psi,j}^* - h_{j,n}^*] = o_P(n^{-1/2}).$$

Since $P\dot{\ell}_\psi(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[h^*_{\psi,j}-h^*_{j,n}]=0$, we decompose I_n into $I_n=I_{n,1}+I_{n,2}$, where

$$\begin{aligned} I_{n,1} &= (\mathbb{P}_n-P)\dot{\ell}_\psi(\hat{\boldsymbol{\theta}}_n;\boldsymbol{O}_i)[h^*_{\psi,j}-h^*_{j,n}], \\ I_{n,2} &= P\big\{\dot{\ell}_\psi(\hat{\boldsymbol{\theta}}_n;\boldsymbol{O}_i)[h^*_{\psi,j}-h^*_{j,n}]-\dot{\ell}_\psi(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[h^*_{\psi,j}-h^*_{j,n}]\big\}. \end{aligned}$$

B4 is proved if both $I_{n,1}$ and $I_{n,2}$ are $o_P(n^{-1/2})$.

For $I_{n,1}$, the key is to study the following class of functions:

$$\mathcal{F}^j_n(\eta) \;=\; \{\dot{\ell}_\psi(\boldsymbol{\theta};\boldsymbol{O}_i)[h^*_j-h_j]:\boldsymbol{\theta}\in\Theta_n, d(\boldsymbol{\theta},\boldsymbol{\theta}_0)\le\eta, h_j\in\mathcal{H}_\psi, \|h^*_j-h_j\|_\infty\le\eta\}.$$

By a similar argument that we used to prove the Condition A2 in Theorem 1, we can prove that

$$\begin{aligned} & |\dot{\ell}_\psi(\boldsymbol{\theta};\boldsymbol{O}_i)[h^*_j-h_j]-\dot{\ell}_\psi(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[h^*_j-h_j]| \\ \lesssim\; & (\|\boldsymbol{\zeta}-\boldsymbol{\zeta}_0\|+\|\psi-\psi_{n,0}\|_{\Lambda_0}+\|\mu-\mu_{n,0}\|_{F_0})\times\|h^*_j-h_j\|_\infty. \end{aligned}$$

It implies that the ϵ-bracketing number of $\mathcal{F}^j_n(\eta)$ in terms of the L_∞ norm, denoted by $N_{[\,]}(\epsilon,\mathcal{F}^j_n(\eta),\|\cdot\|_\infty)$, is bounded by $(\eta/\epsilon)^{c(K_\mu,K_\psi)}$, where $c(K_\mu,K_\psi)=d+c_1(K_\mu+p_\mu+1)+c_2(K_\psi+p_\psi+1)$. Then it follows that

$$\log N_{[\,]}(\epsilon,\mathcal{F}^j_n(\eta),L_2(P))\le\log N_{[\,]}(\epsilon,\mathcal{F}^j_n(\eta),\|\cdot\|_\infty)\lesssim c(K_\mu,K_\psi)\log(\eta/\epsilon),$$

which leads to the bracketing integral

$$J_{[\,]}(\eta,\mathcal{F}^j_n(\eta),L_2(P))=\int_0^\eta\sqrt{1+\log N_{[\,]}(\epsilon,\mathcal{F}^j_n(\eta),L_2(P))}d\epsilon\lesssim c(K_\mu,K_\psi)^{1/2}\eta.$$

Note that $c(K_\mu,K_\psi)\asymp n^{\max\{v_\psi,v_\mu\}}$. We can pick η to be

$$\eta_n=O(n^{-\min\{p_\psi v_\psi,p_\mu v_\mu,(1-\max\{v_\psi,v_\mu\})/2\}}).$$

Then $\|h^*_{\psi,j}-h^*_{j,n}\|_\infty=O(n^{-p_\psi v_\psi})\le\eta_n$ and $d(\hat{\boldsymbol{\theta}}_n,\boldsymbol{\theta}_0)=O_P(n^{-\min\{p_\psi v_\psi,p_\mu v_\mu,(1-\max\{v_\psi,v_\mu\})/2\}})\asymp\eta_n$, which implies that $\dot{\ell}_\psi(\hat{\boldsymbol{\theta}}_n;\boldsymbol{O}_i)[h^*_{\psi,j}-h^*_{j,n}]\in\mathcal{F}^j_n(\eta_n)$. For any $\dot{\ell}_\psi(\boldsymbol{\theta};\boldsymbol{O})[h^*_j-h_j]\in\mathcal{F}^j_n(\eta_n)$, we have

$$\begin{aligned} & P\left\{\dot{\ell}_\psi(\boldsymbol{\theta};\boldsymbol{O})[h^*_j-h_j]\right\}^2 \\ =\; & P\left\{E^x_{i,\boldsymbol{\theta}}\left[\delta_i[h^*_j-h_j](\tilde{r}_i(\boldsymbol{\zeta};\boldsymbol{u}_i))-\int_a^b I(\tilde{r}_i(\boldsymbol{\zeta};\boldsymbol{u}_i)\ge s)\exp\{\psi(s)\}[h^*_j-h_j](s)\,ds\right]\right\}^2 \\ \lesssim\; & PE^x_{i,\boldsymbol{\theta}}\left[[h^*_j-h_j]^2(\tilde{r}_i(\boldsymbol{\zeta};\boldsymbol{u}_i))+\int_a^b\exp\{2\psi(s)\}[h^*_j-h_j]^2(s)\,ds\right] \\ \lesssim\; & \|h^*_j-h_j\|^2_\infty. \end{aligned}$$

Note that $\|h_j^* - h_j\|_\infty \leq \eta_n$. It is easy to see that $\|\dot{\ell}_\psi(\boldsymbol{\theta}; \boldsymbol{O})[h_j^* - h_j]\|_\infty$ can be taken to be bounded by some constant $0 < M < \infty$. By the maximal inequality in Lemma 3.4.2 of van der Vaart and Wellner (1996), it follows that

$$\begin{aligned} E_P\|\mathbb{G}_n\|_{\mathcal{F}_n^j(\eta_n)} &\lesssim J_{[\,]}(\eta, \mathcal{F}_n^j(\eta), L_2(P))\left(1 + \frac{J_{[\,]}(\eta, \mathcal{F}_n^j(\eta), L_2(P))}{\eta_n^2\sqrt{n}}M\right) \\ &\lesssim c(K_\mu, K_\psi)^{1/2}\eta_n + c(K_\mu, K_\psi)n^{-1/2} \\ &= O\left(n^{\max\{v_\psi, 2v_\mu\}/2 - \min\{p_\psi v_\psi, p_\mu v_\mu, (1-\max\{v_\psi, 2v_\mu\})/2\}}\right) + O\left(n^{\max\{v_\psi, 2v_\mu\} - 1/2}\right) \\ &= o(1), \end{aligned}$$

where the last equality follows from $1/(2+2p_\psi) < v_\psi < 1/(2p_\psi)$, $1/(2+2p_\mu) < v_\mu < 1/(2p_\mu)$, $p_\psi \geq 3$ and $p_\mu \geq 2$. By Markov's inequality, $I_{1n} = n^{-1/2}\mathbb{G}_n\dot{\ell}_\psi(\boldsymbol{\theta}; \boldsymbol{O})[h_j^* - h_j] = o_P(n^{-1/2})$.

For $I_{n,2}$, given any $\boldsymbol{\theta}_n \in \Theta_n$, by Taylor's expansion, we have,

$$\begin{aligned} &\dot{\ell}_\psi(\boldsymbol{\theta}_n; \boldsymbol{O}_i)[h_j^* - h_j] - \dot{\ell}_\psi(\boldsymbol{\theta}_0; \boldsymbol{O}_i)[h_j^* - h_j] \\ = \;& (\boldsymbol{\zeta}_n - \boldsymbol{\zeta}_0)^\top \ddot{\ell}_{\psi\boldsymbol{\zeta}}(\tilde{\boldsymbol{\theta}}_n; \boldsymbol{O}_i)[h_j^* - h_j] + \ddot{\ell}_{\psi\psi}(\tilde{\boldsymbol{\theta}}_n; \boldsymbol{O}_i)[h_j^* - h_j, \psi_n - \psi_0] \\ &+ \ddot{\ell}_{\psi\mu}(\tilde{\boldsymbol{\theta}}_n; \boldsymbol{O}_i)[h_j^* - h_j, \mu_n - \mu_0], \end{aligned}$$

where $\tilde{\boldsymbol{\theta}}_n$ is between $\boldsymbol{\theta}_n$ and $\boldsymbol{\theta}_0$, and

$$\begin{aligned} \ddot{\ell}_{\alpha\upsilon}(\boldsymbol{\theta}; \boldsymbol{O}_i) &= -E_{i,\boldsymbol{\theta}}^u\left[\dot{\tilde{\ell}}_\alpha(\boldsymbol{\theta}; \boldsymbol{u}_i)\right]E_{i,\boldsymbol{\theta}}^u\left[\dot{\tilde{\ell}}_\upsilon(\boldsymbol{\theta}; \boldsymbol{u}_i)\right] \\ &\quad + E_{i,\boldsymbol{\theta}}^u\left[\dot{\tilde{\ell}}_\alpha(\boldsymbol{\theta}; \boldsymbol{u}_i)\dot{\tilde{\ell}}_\upsilon(\boldsymbol{\theta}; \boldsymbol{u}_i) + \ddot{\tilde{\ell}}_{\alpha\upsilon}(\boldsymbol{\theta}; \boldsymbol{u}_i)\right] \\ &= -\dot{\ell}_\alpha(\boldsymbol{\theta}; \boldsymbol{O}_i)\dot{\ell}_\upsilon(\boldsymbol{\theta}; \boldsymbol{O}_i) + E_{i,\boldsymbol{\theta}}^u\left[\dot{\tilde{\ell}}_\alpha(\boldsymbol{\theta}; \boldsymbol{u}_i)\dot{\tilde{\ell}}_\upsilon(\boldsymbol{\theta}; \boldsymbol{u}_i) + \ddot{\tilde{\ell}}_{\alpha\upsilon}(\boldsymbol{\theta}; \boldsymbol{u}_i)\right], \end{aligned}$$

where

$$\tilde{\ell}(\boldsymbol{\theta}; \boldsymbol{u}_i) = \ell_S(\boldsymbol{\theta}_S; \boldsymbol{u}_i) + \ell_y(\boldsymbol{\theta}_y; \boldsymbol{u}_i) + \ell_u(\Sigma_u; \boldsymbol{u}_i),$$

and α, υ represent any two of the parameters in $(\boldsymbol{\zeta}, \psi, \mu)$. It is worth noting that in general

$$\ddot{\ell}_{\alpha\upsilon}(\boldsymbol{\theta}) \neq E_{i,\boldsymbol{\theta}}^u\left[\ddot{\tilde{\ell}}_{\alpha\upsilon}(\boldsymbol{\theta}; \boldsymbol{u}_i)\right],$$

since the operator $E_{i,\boldsymbol{\theta}}^u[\cdot]$ itself also depends on $\boldsymbol{\theta}$. It follows from direct calculations that,

$$\begin{aligned} \ddot{\tilde{\ell}}_{\psi\boldsymbol{\zeta}}(\boldsymbol{\theta}; \boldsymbol{u}_i)[h_\psi] &= -\left[Z_i^\top, \boldsymbol{u}_i^\top\int_0^1 B(\tau)B(\tau)d\tau\right]\Big\{\delta_i\dot{h}_\psi(\tilde{r}_i(\boldsymbol{\zeta}; \boldsymbol{u}_i)) \\ &\qquad - \int_a^b I(\tilde{r}_i(\boldsymbol{\zeta}; \boldsymbol{u}_i) \geq s)\exp\{\psi(s)\}\left[\dot{h}_\psi(s) + \dot{\psi}(s)\dot{h}_\psi(s)\right]ds\Big\}, \\ \ddot{\tilde{\ell}}_{\psi\psi}(\boldsymbol{\theta}; \boldsymbol{u}_i)[h_\psi, \hbar_\psi] &= -\int_a^b I(\tilde{r}_i(\boldsymbol{\zeta}; \boldsymbol{u}_i) \geq s)\exp\{\psi(s)\}h_\psi(s)\hbar_\psi(s)ds, \end{aligned}$$

and $\ddot{\tilde{\ell}}_{\psi\mu}(\boldsymbol{\theta};\boldsymbol{u}_i)[h_\psi,h_\mu]=0$. Following a similar argument used in proving the Condition A2 in Theorem 1, it can be proved that

$$\begin{aligned}
& P\left\{\left|\dot{\ell}_\psi(\hat{\boldsymbol{\theta}}_n;\boldsymbol{O})[h^*_{\psi,j}-h^*_{j,n}]-\dot{\ell}_\psi(\boldsymbol{\theta}_0;\boldsymbol{O})[h^*_{\psi,j}-h^*_{j,n}]\right|^2\right\} \\
\lesssim\ & d(\hat{\boldsymbol{\theta}}_n,\boldsymbol{\theta}_0)^2\times(\|h^*_{\psi,j}-h^*_{j,n}\|_\infty+\|\dot{h}^*_{g,j}-\dot{h}^*_{j,n}\|_\infty)^2 \\
=\ & o(n^{-1}),
\end{aligned}$$

which implies that $I_{n,2}=o_P(n^{-1/2})$. Thus $I_n=I_{n,1}+I_{n,2}=o_P(n^{-1/2})$ and Condition B4 holds.

For B5, the three equations are proved by studying the following three classes of functions:

$$\begin{aligned}
\mathcal{F}^{\boldsymbol{\zeta}}_{n,j}(\eta) &= \{\dot{\ell}_{\eta_j}(\boldsymbol{\theta};\boldsymbol{O})-\dot{\ell}_{\zeta_j}(\boldsymbol{\theta}_0;\boldsymbol{O}):\boldsymbol{\theta}\in\Theta_n,d(\boldsymbol{\theta},\boldsymbol{\theta}_0)\le\eta,\|\dot{\psi}-\dot{\psi}_0\|_{\Lambda_0}\le\eta\}, \\
\mathcal{F}^{\psi}_{n,j}(\eta) &= \{\dot{\ell}_\psi(\boldsymbol{\theta};\boldsymbol{O})[h^*_{j,n}]-\dot{\ell}_\psi(\boldsymbol{\theta}_0;\boldsymbol{O})[h^*_{\psi,j}]:\boldsymbol{\theta}\in\Theta_n,d(\boldsymbol{\theta},\boldsymbol{\theta}_0)\le\eta\}, \\
\mathcal{F}^{\mu}_{n,j}(\eta) &= \{\dot{\ell}_\mu(\boldsymbol{\theta};\boldsymbol{O})[h^*_{\mu,j}]-\dot{\ell}_\mu(\boldsymbol{\theta}_0;\boldsymbol{O})[h^*_{\mu,j}]:\boldsymbol{\theta}\in\Theta_n,d(\boldsymbol{\theta},\boldsymbol{\theta}_0)\le\eta\},
\end{aligned}$$

for $j=1,...,q$, where $\dot{\ell}_{\eta_j}(\boldsymbol{\theta};\boldsymbol{O})$ is the jth component of $\dot{\ell}_\eta(\boldsymbol{\theta};\boldsymbol{O})$. The rest of the proof for B5 is essentially very similar to B4. The η is picked to be

$$\eta_n=O(n^{-\min\{(p_\psi-1)v_\psi,p_\mu v_\mu,(1-\max\{v_\psi,2v_\mu\})/2\}}).$$

More detailed discussion can be found in Ding and Nan (2011), page 3055-3057.

Finally, we verify B6. We just provide the details for the third equation regarding $\dot{\ell}_\psi(\boldsymbol{\theta};\boldsymbol{O})[\boldsymbol{h}^*_\psi]$, as the proofs for the three equations are essentially the same. To avoid confusion in later discussions, we use $\boldsymbol{O}_i=(r_i,\delta_i,Z_i,\boldsymbol{y}_i,\boldsymbol{t}_i)$ in place of $\boldsymbol{O}$. By Taylor's expansion, we have,

$$\begin{aligned}
& P\Big\{\dot{\ell}_\psi(\boldsymbol{\theta};\boldsymbol{O})[\boldsymbol{h}^*_\psi]-\dot{\ell}_\psi(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}^*_\psi]-\ \ddot{\ell}_{\psi\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}^*_\psi](\boldsymbol{\zeta}-\boldsymbol{\zeta}_0) \\
&\quad -\ddot{\ell}_{\psi\mu}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}^*_\psi,\mu-\mu_0]-\ddot{\ell}_{\psi\psi}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}^*_\psi,\psi-\psi_0]\Big\} \\
=\ & P\Big\{\ddot{\ell}_{\psi\boldsymbol{\zeta}}(\tilde{\boldsymbol{\theta}};\boldsymbol{O}_i)[\boldsymbol{h}^*_\psi]-\ddot{\ell}_{\psi\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\boldsymbol{h}^*_\psi]\Big\}(\boldsymbol{\zeta}-\boldsymbol{\zeta}_0) \\
& +P\Big\{\ddot{\ell}_{\psi\mu}(\tilde{\boldsymbol{\theta}};\boldsymbol{O}_i)[\boldsymbol{h}^*_\psi,\mu-\mu_0]-\ddot{\ell}_{\psi\mu}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\boldsymbol{h}^*_\psi,\mu-\mu_0]\Big\} \\
& +P\Big\{\ddot{\ell}_{\psi\psi}(\tilde{\boldsymbol{\theta}};\boldsymbol{O}_i)[\boldsymbol{h}^*_\psi,\psi-\psi_0]-\ddot{\ell}_{\psi\psi}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\boldsymbol{h}^*_\psi,\psi-\psi_0]\Big\}
\end{aligned}$$

where $\tilde{\boldsymbol{\theta}}$ is between $\boldsymbol{\theta}$ and $\boldsymbol{\theta}_0$. It follows from direct calculations that

$$\begin{aligned}
& \left|P\Big\{\ddot{\ell}_{\psi\boldsymbol{\zeta}}(\tilde{\boldsymbol{\theta}};\mathcal{O}_i)[\boldsymbol{h}_\psi^*]-\ddot{\ell}_{\psi\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\mathcal{O}_i)[\boldsymbol{h}_\psi^*]\Big\}(\boldsymbol{\zeta}-\boldsymbol{\zeta}_0)\right| \\
= & \left|P\Big\{E^u_{i,\tilde{\boldsymbol{\theta}}}\left[\ddot{\ell}_{S,\psi\boldsymbol{\zeta}}(\tilde{\boldsymbol{\theta}};\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{h}_\psi^*,\boldsymbol{\zeta}-\boldsymbol{\zeta}_0]\right]-E^u_{i,\boldsymbol{\theta}_0}\left[\ddot{\ell}_{S,\psi\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{h}_\psi^*,\boldsymbol{\zeta}-\boldsymbol{\zeta}_0]\right]\right. \\
& \quad + E^u_{i,\tilde{\boldsymbol{\theta}}}\left[\dot{\ell}_{S,\psi}(\tilde{\boldsymbol{\theta}};\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{h}_\psi^*]\dot{\ell}_{S,\psi}(\tilde{\boldsymbol{\theta}};\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{\zeta}-\boldsymbol{\zeta}_0]\right] \\
& \quad - E^u_{i,\boldsymbol{\theta}_0}\left[\dot{\ell}_{S,\psi}(\boldsymbol{\theta}_0;\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{h}_\psi^*]\dot{\ell}_{S,\psi}(\boldsymbol{\theta}_0;\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{\zeta}-\boldsymbol{\zeta}_0]\right] \\
& \quad - E^u_{i,\tilde{\boldsymbol{\theta}}}\left[\dot{\ell}_{S,\psi}(\tilde{\boldsymbol{\theta}};\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{h}_\psi^*]\right]E^u_{i,\tilde{\boldsymbol{\theta}}}\left[\dot{\ell}_{S,\psi}(\tilde{\boldsymbol{\theta}};\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{\zeta}-\boldsymbol{\zeta}_0]\right] \\
& \quad \left. + E^u_{i,\boldsymbol{\theta}_0}\left[\dot{\ell}_{S,\psi}(\boldsymbol{\theta}_0;\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{h}_\psi^*]\right]E^u_{i,\boldsymbol{\theta}_0}\left[\dot{\ell}_{S,\psi}(\boldsymbol{\theta}_0;\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{\zeta}-\boldsymbol{\zeta}_0]\right]\Big\}\right| \\
\lesssim & \left|P\Big\{E^u_{i,\tilde{\boldsymbol{\theta}}}\left[\ddot{\ell}_{S,\psi\boldsymbol{\zeta}}(\tilde{\boldsymbol{\theta}};\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{h}_\psi^*,\boldsymbol{\zeta}-\boldsymbol{\zeta}_0]\right]-E^u_{i,\boldsymbol{\theta}_0}\left[\ddot{\ell}_{r,\psi\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{h}_\psi^*,\boldsymbol{\zeta}-\boldsymbol{\zeta}_0]\right]\Big\}\right| \\
& + \left|P\Big\{E^u_{i,\tilde{\boldsymbol{\theta}}}\left[\dot{\ell}_{S,\psi}(\tilde{\boldsymbol{\theta}};\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{h}_\psi^*]\dot{\ell}_{S,\psi}(\tilde{\boldsymbol{\theta}};\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{\zeta}-\boldsymbol{\zeta}_0]\right]\right. \\
& \quad \left. - E^u_{i,\boldsymbol{\theta}_0}\left[\dot{\ell}_{S,\psi}(\boldsymbol{\theta}_0;\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{h}_\psi^*]\dot{\ell}_{S,\psi}(\boldsymbol{\theta}_0;\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{\zeta}-\boldsymbol{\zeta}_0]\right]\Big\}\right| \\
& + \left|P\Big\{E^u_{i,\tilde{\boldsymbol{\theta}}}\left[\dot{\ell}_{S,\psi}(\tilde{\boldsymbol{\theta}};\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{h}_\psi^*]\right]E^u_{i,\tilde{\boldsymbol{\theta}}}\left[\dot{\ell}_{S,\psi}(\tilde{\boldsymbol{\theta}};\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{\zeta}-\boldsymbol{\zeta}_0]\right]\right. \\
& \quad \left. - E^u_{i,\boldsymbol{\theta}_0}\left[\dot{\ell}_{S,\psi}(\boldsymbol{\theta}_0;\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{h}_\psi^*]\right]E^u_{i,\boldsymbol{\theta}_0}\left[\dot{\ell}_{S,\psi}(\boldsymbol{\theta}_0;\mathcal{O}_i,\boldsymbol{u}_i)[\boldsymbol{\zeta}-\boldsymbol{\zeta}_0]\right]\Big\}\right| \\
:= & \ B_1+B_2+B_3.
\end{aligned}$$

By a similar argument that we used to verify the Condition B4 above and the Condition A2 in Theorem 1, we can prove that

$$\begin{aligned}
B_1 & \lesssim (d(\boldsymbol{\theta},\boldsymbol{\theta}_0)+\|\dot{\psi}-\dot{\psi}_0\|_{\Lambda_0}+\|\ddot{\psi}-\ddot{\psi}_0\|_{\Lambda_0})\times\|\boldsymbol{\zeta}-\boldsymbol{\zeta}_0\| \\
& \lesssim O(n^{-\min\{(p_\psi-2)v_\psi,p_\mu v_\mu,(1-\max\{v_\psi,2v_\mu\})/2\}})\times O(n^{-\kappa}) \\
& = O(n^{-\min\{(p_\psi-2)v_\psi,p_\mu v_\mu,(1-\max\{v_\psi,2v_\mu\})/2\}})\times O(n^{-\min\{p_\psi v_\psi,p_\mu v_\mu,(1-\max\{v_\psi,2v_\mu\})/2\}}) \\
& = o(n^{-1/2}),
\end{aligned}$$

where the last equality requires the condition $(p_\psi-2)v_\psi > v_\mu$. And

$$\begin{aligned}
B_2+B_3 & \lesssim (d(\boldsymbol{\theta},\boldsymbol{\theta}_0)+\|\dot{\psi}-\dot{\psi}_0\|_{\Lambda_0})\times\|\boldsymbol{\zeta}-\boldsymbol{\zeta}_0\| \\
& \lesssim O(n^{-\min\{(p_\psi-1)v_\psi,p_\mu v_\mu,(1-\max\{v_\psi,2v_\mu\})/2\}})\times O(n^{-\min\{p_\psi v_\psi,p_\mu v_\mu,(1-\max\{v_\psi,2v_\mu\})/2\}}) \\
& = o(n^{-1/2}).
\end{aligned}$$

Let $\tilde{a}=\min\{(p_\psi-2)v_\psi,p_\mu v_\mu,(1-\max\{v_\psi,v_\mu\})/2\}$. We have

$$P\Big\{\ddot{\ell}_{\psi\boldsymbol{\zeta}}(\tilde{\boldsymbol{\theta}};\mathcal{O}_i)[\boldsymbol{h}_\psi^*]-\ddot{\ell}_{\psi\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\mathcal{O}_i)[\boldsymbol{h}_\psi^*]\Big\}(\boldsymbol{\zeta}-\boldsymbol{\zeta}_0) \lesssim O(n^{-a\kappa})=o(n^{-1/2}),$$

where $a = (\tilde{a}+\kappa)/\kappa > 1$ and $a\kappa > 1/2$. Similarly, we can prove that

$$\begin{aligned} P\Big\{\ddot{\ell}_{\psi\mu}(\tilde{\boldsymbol{\theta}};\boldsymbol{O}_i)[\boldsymbol{h}^*_\psi, \mu-\mu_0] - \ddot{\ell}_{\psi\mu}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\boldsymbol{h}^*_\psi, \mu-\mu_0]\Big\} &\lesssim O(n^{-a\kappa}) = o(n^{-1/2}), \\ P\Big\{\ddot{\ell}_{\psi\psi}(\tilde{\boldsymbol{\theta}};\boldsymbol{O}_i)[\boldsymbol{h}^*_\psi, \psi-\psi_0] - \ddot{\ell}_{\psi\psi}(\boldsymbol{\theta}_0;\boldsymbol{O}_i)[\boldsymbol{h}^*_\psi, \psi-\psi_0]\Big\} &\lesssim O(n^{-a\kappa}) = o(n^{-1/2}), \end{aligned}$$

and thus

$$\begin{aligned} P\Big\{&\dot{\ell}_\psi(\boldsymbol{\theta};\boldsymbol{O})[\boldsymbol{h}^*_\psi] - \dot{\ell}_\psi(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}^*_\psi] - \ddot{\ell}_{\psi\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}^*_\psi](\boldsymbol{\zeta}-\boldsymbol{\zeta}_0) \\ &- \ddot{\ell}_{\psi\mu}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}^*_\psi, \mu-\mu_0] - \ddot{\ell}_{\psi\psi}(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}^*_\psi, \psi-\psi_0]\Big\} \lesssim O(n^{-a\kappa}) = o(n^{-1/2}). \end{aligned}$$

Therefore, we have verified all six conditions B1-B6. Following Theorem 2.1 in Ding and Nan (2011), we have

$$\sqrt{n}(\hat{\boldsymbol{\zeta}}_n - \boldsymbol{\zeta}_0) \to N(0, \Omega_{\boldsymbol{\zeta}}^{-1}\tilde{\Omega}_{\boldsymbol{\zeta}}(\Omega_{\boldsymbol{\zeta}}^{-1})'),$$

where $\Omega_{\boldsymbol{\zeta}}$ is as defined in B3 and

$$\tilde{\Omega}_{\boldsymbol{\zeta}} = P\Big\{\Big(\dot{\ell}_{\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O}) - \dot{\ell}_\beta(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}^*_\beta] - \dot{\ell}_\psi(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}^*_\psi] - \dot{\ell}_\mu(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}^*_\mu]\Big)^{\otimes 2}\Big\}.$$

By the proofs for B3, we have $\Omega_{\boldsymbol{\zeta}} = \tilde{\Omega}_{\boldsymbol{\zeta}}$, i.e., $\sqrt{n}(\hat{\boldsymbol{\zeta}}_n - \boldsymbol{\zeta}_0) \to N(0, \Omega_{\boldsymbol{\zeta}}^{-1})$. Since

$$\Big(\dot{\ell}_{\boldsymbol{\zeta}}(\boldsymbol{\theta}_0;\boldsymbol{O}) - \dot{\ell}_\beta(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}^*_\beta] - \dot{\ell}_\psi(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}^*_\psi] - \dot{\ell}_\mu(\boldsymbol{\theta}_0;\boldsymbol{O})[\boldsymbol{h}^*_\mu]\Big)$$

is the efficient score function for $\boldsymbol{\zeta}_0$, it follows that $\Omega_{\boldsymbol{\zeta}}^{-1}$ achieves the semiparametric efficiency bound. □

References

Ding, Y., and Nan, B. (2011), "A sieve M-theorem for bundled parameters in semiparametric models, with application to the efficient estimation in a linear model for censored data," *The Annals of Statistics*, 39, 3032–3061.

Gervini, D., and Gasser, T. (2004), "Self-modelling warping functions," *Journal of the Royal Statistical Society, Series B*, 66(4), 959–971.

Lipsitz, S. R., Fitzmaurice, G. M., Ibrahim, J. G., Gelber, R., and Lipshultz, S. (2002), "Parameter Estimation in Longitudinal Studies with Outcome-Dependent Follow-Up," *Biometrics*, 58(3), 621–630.

Schumaker, L. L. (1981), *Spline Functions: Basic Theory*, New York: Wiley.

Shen, X., and Wong, W, H. (1994), "Convergence Rate of Sieve Estimates," *The Annals of Statistics*, 22(2), 580–615.