Semiparametric Inference for Functional Survival Models

Hongyi Zhou, Wenqing Su, Qixian Zhong, Ying Yang

Tsinghua University, Xiamen University

Supplementary Material

This supplementary material contains implementation details of our estimators, some additional simulation results and the auxiliary lemmas and technical proofs for propositions and theorems of the paper.

S1 Estimation of Information matrix and Simultaneous Confidence Band

In this section, we provide some details on the estimation of information matrix I_n in Theorem 3 and J_n in Theorem 4. In applications, we first estimate the gradient of loglikelihood (given by equation (2.6) in the manuscript) with respect to all the coefficient, which involves a solution to the ODE system. Therefore, we first evaluate $\Lambda_{U_i}(Y_i;\theta)$ by calculating the solution of (2.7) given the parameter estimates θ and covariates U_i at time $t = Y_i$. Then, we obtain the derivatives of $\Lambda_{U_i}(Y_i;\theta)$ with respect to the parameters by solving another ODE, that is, $\Lambda'_\alpha(Y_i,\theta), \Lambda'_a(Y_i,\theta), \Lambda'_b(Y_i,\theta)$ and $\Lambda'_c(Y_i,\theta)$ is the solution of the following ODE at time Y_i:

$$\frac{d\Lambda'_\alpha(t)}{dt} = \Psi(t, X_i, Z_i, \theta)\left(X_i + \sum_{j=1}^{q_{n,3}} c_j B_j^g\left(\Lambda_{U_i}(t)\right)\Lambda'_\alpha(t)\right),$$

$$\frac{d\Lambda'_a(t)}{dt} = \Psi(t, X_i, Z_i, \theta)\Bigg(\left[\int_0^1 B_1^\beta(s)Z_i(s)ds, \ldots, \int_0^1 B_{q_{n,1}}^\beta(s)Z_i(s)ds\right]^T + \sum_{j=1}^{q_{n,3}} c_j B_j^g\left(\Lambda_{U_i}(t)\right)\Lambda'_a(t)\Bigg),$$

$$\frac{d\Lambda'_b(t)}{dt} = \Psi(t, X_i, Z_i, \theta)\left(\left[B_1^\gamma(t), \ldots, B_{q_{n,2}}^\gamma(t)\right]^T + \sum_{j=1}^{q_{n,3}} c_j B_j^g\left(\Lambda_{U_i}(t)\right)\Lambda'_b(t)\right),$$

$$\frac{d\Lambda'_c(t)}{dt} = \Psi(t, X_i, Z_i, \theta)\left(\left[B_1^g(\Lambda_{U_i}(t), \ldots, B_{q_{n,3}}^g(\Lambda_{U_i}(t)\right]^T + \sum_{j=1}^{q_{n,3}} c_j B_j^g\left(\Lambda_{U_i}(t)\right)\Lambda'_\alpha(t)\right) \tag{S1.1}$$

with initial value $\Lambda'_\alpha(\mathbf{0}) = \Lambda'_a(\mathbf{0}) = \Lambda'_b(\mathbf{0}) = \Lambda'_c(0) = \mathbf{0}$, where

$$\Psi(t, X, Z, \theta) = \exp\left(\alpha^T X + \sum_{j=1}^{q_{n,1}} a_j \int_0^1 B_j^\beta(s)Z(s)ds + \sum_{j=1}^{q_{n,2}} b_j B_j^\gamma(t) + \sum_{j=1}^{q_{n,3}} c_j B_j^g\left(\Lambda_U(t)\right)\right).$$

Then we plug in the value of $\Lambda'_\alpha(Y_i,\theta), \Lambda'_a(Y_i,\theta), \Lambda'_b(Y_i,\theta)$ to obtain the gradient of loglikelihood $\frac{\partial}{\partial\theta} l_n(\theta)$. It follows that the matrix $A = \frac{\partial}{\partial\theta} l_n(\theta)\left(\frac{\partial}{\partial\theta} l_n(\theta)\right)^T$ can be expressed as

$$A = \begin{pmatrix} \frac{\partial l_n}{\partial\alpha}\left(\frac{\partial l_n}{\partial\alpha}\right)^T & \frac{\partial l_n}{\partial\alpha}\left(\frac{\partial l_n}{\partial a}\right)^T & \frac{\partial l_n}{\partial\alpha}\left(\frac{\partial l_n}{\partial(b,c)}\right)^T \\ \frac{\partial l_n}{\partial a}\left(\frac{\partial l_n}{\partial\alpha}\right)^T & \frac{\partial l_n}{\partial a}\left(\frac{\partial l_n}{\partial a}\right)^T & \frac{\partial l_n}{\partial a}\left(\frac{\partial l_n}{\partial(b,c)}\right)^T \\ \frac{\partial l_n}{\partial(b,c)}\left(\frac{\partial l_n}{\partial\alpha}\right)^T & \frac{\partial l_n}{\partial(b,c)}\left(\frac{\partial l_n}{\partial a}\right)^T & \frac{\partial l_n}{\partial(b,c)}\left(\frac{\partial l_n}{\partial(b,c)}\right)^T \end{pmatrix}$$

Suppose the inverse of matrix A takes the form of

$$A^{-1} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}$$

Then, the inverse information matrix for the scalar parameter $I_n^{-1}(\alpha)$ can be estimated by the first diagnoal block A_{11} and the matrix J_n^{-1} in Theorem 4 can be estimated by the second diagnoal block A_{22}.

S2 Additional Simulation Result

S2.1 Estimation of nuisance parameter

This section presents the estimation of nuisance parameter under setting 3, 4, 5 of our simulation study. Figure 1 displays the estimates of functional parameters $\beta(\cdot), \gamma(\cdot)$ and $g(\cdot)$ with a sample size of $n = 800$. The dash line and dash-dotted line represent the true value and the point-wise mean estimates, respectively. The dotted line represents 95% point-wise confidence bands. Clearly, the mean of estimates approximates the true value well in all scenarios. Table 1 presents the empirical coverage probability of simultaneous confidence band. As shown in Table 1, the empirical coverage probability of simultaneous confidence band approaches its theoretical value 0.95 as sample size increases.

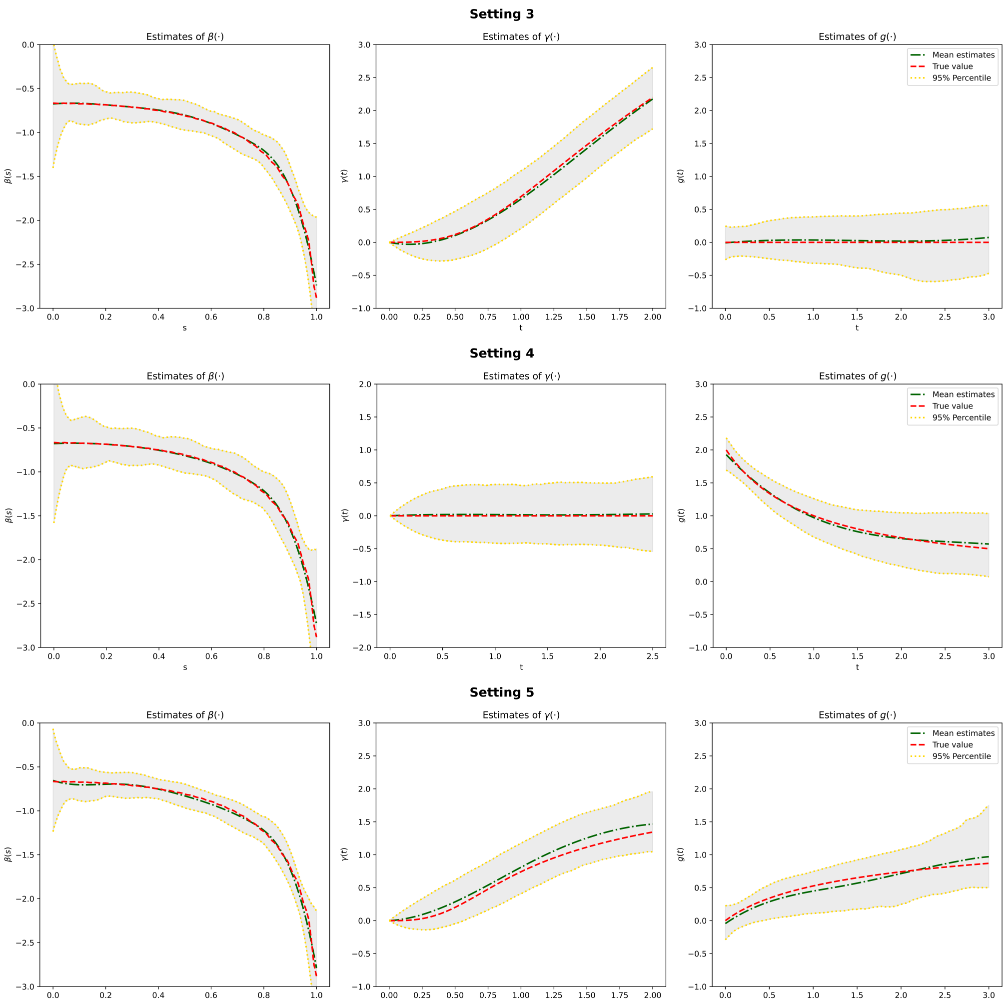

Figure 1: Estimates of functional parameters based on 1000 replications.

Table 1: Simulation results for functional parameter β under setting 3, 4 and 5, where CP represents the empirical coverage probability of simultaneous confidence band.

		Censoring rate = 15%		Censoring rate = 30%	
Setting	N	RIMSE	CP	RIMSE	CP
3	200	0.072	0.782	0.080	0.779
	400	0.034	0.842	0.035	0.844
	600	0.026	0.910	0.030	0.915
	800	0.019	0.947	0.022	0.938
4	200	0.115	0.779	0.102	0.788
	400	0.045	0.840	0.069	0.848
	600	0.035	0.903	0.042	0.907
	800	0.030	0.927	0.036	0.925
5	200	0.047	0.798	0.044	0.808
	400	0.023	0.880	0.024	0.874
	600	0.017	0.904	0.018	0.894
	800	0.013	0.940	0.015	0.936

S2.2 Sensitivity to spline basis

In this subsection, we provide some additional simulation results about the sensitivity to the spline basis. In this simulation, we considered "Setting 5" from the manuscript and let the order of the B-spline basis be either 2 or 3. We examined the impact of varying the number of interior knots—1, 2, 3, 5, 10, and 20—on the integrated mean square error of the functional parameter, β. The results are presented in the following table.

As shown in Table 2, there is no significant difference in the performance of the estimators when the number of interior knots is similar. However, we observed that when the number of interior knots is relatively large (e.g., 10 or 20), the estimated functional parameter may become quite unsmooth. Therefore, we recommend treating the number of basis functions as predetermined. In general, we suggest using cubic B-splines of order 3.

Additionally, based on our theoretical results, the optimal rate of convergence is achieved by choosing the number of interior knots to be $O(n^{-1/(2p+1)})$. Thus, we recommend using $\lceil n^{-1/(2p+1)} \rceil$ as the number of interior knots.

Table 2: Integrated mean square error of functional estimators

	order/interior knot	1	2	3	5	10	20
$n = 200$	$p = 2$	0.034	0.031	0.035	0.058	0.067	0.067
	$p = 3$	0.029	0.032	0.041	0.071	0.066	0.067
$n = 400$	$p = 2$	0.024	0.016	0.015	0.022	0.022	0.025
	$p = 3$	0.014	0.015	0.017	0.023	0.024	0.025
$n = 800$	$p = 2$	0.020	0.011	0.007	0.009	0.010	0.011
	$p = 3$	0.009	0.007	0.008	0.010	0.011	0.011

S2.3 Pointwise confidence interval vs simultaneous confidence band

Simultaneous confidence band is usually wider than the confidence interval. Here we provide some insights about pointwise confidence interval vs simultaneous confidence band. Figure 2, 3, 4 shows instances of confidence interval vs confidence band when sample size $n = 400, 800$ and 1600. The red curve represents the true function, the blue dash line represents confidence interval and the light blue dash line represents confidence band.

As shown in Figure 2, 3, 4 , although the true functional parameter may occasionally fall outside the pointwise confidence interval, it is covered by the confidence band. Meanwhile, the widths of both the confidence interval and the confidence band decrease as sample size n increases. The decay rate of the width of confidence band is slightly slower than $\sqrt{n}$, which aligns with our theoretical $O(n^{-\nu_1})$.

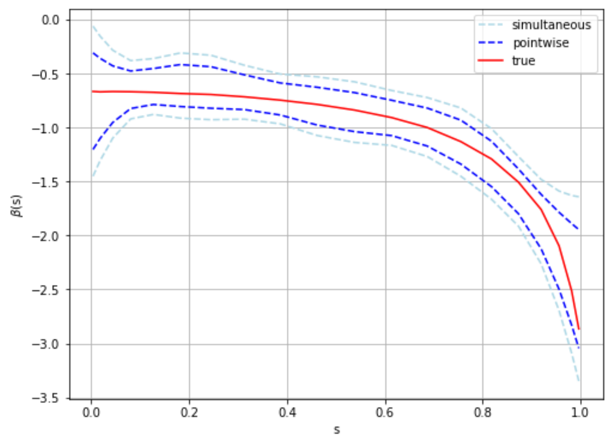

Figure 2: Confidence interval vs confidence band for $n = 400$

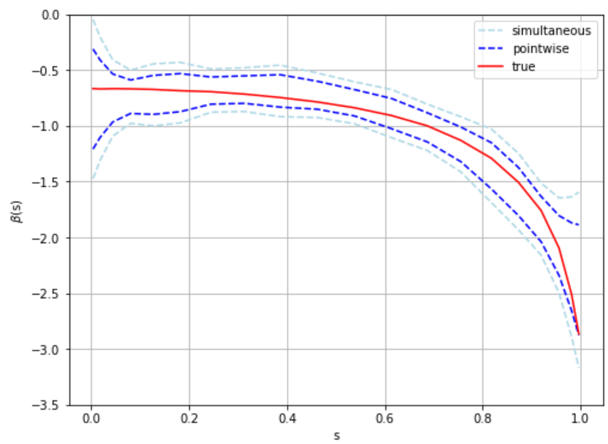

Figure 3: Confidence interval vs confidence band for $n = 800$

S2.4 Comparison of computation cost with MPLE based method

We have also conducted additional simulations to compare the computation efficiency. We consider the functional Cox model (Qu et al., 2016)

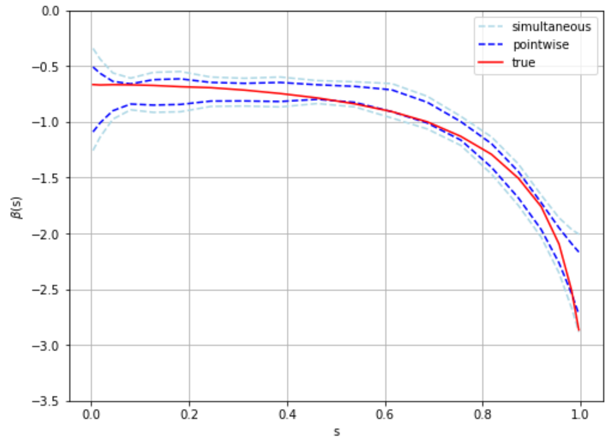

Figure 4: Confidence interval vs confidence band for $n = 1600$

setting and recorded the computation time of the two estimators under different sample sizes vary from $400, 800, 1200, 1600, 2000, 3000$. The results are shown in figure 5. As the figure indicates, the computation time our methods increases significantly slower than MPLE-based method as sample size grows.

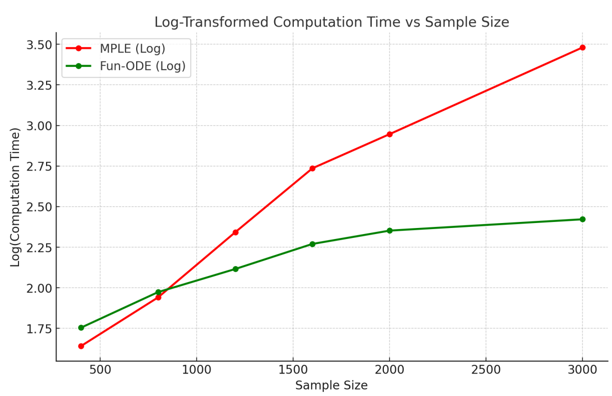

Figure 5: Log computation time of two methods under different sample sizes

S3 Further discussion about real data example

To demonstrate the advantage of our model, we compared its concordance index (C-index) with that of other alternative models. The C-index is a measure of model fitting for survival models, defined as the proportion of all usable subject pairs where the predictions and outcomes are concordant. It is calculated as follows:

$$\binom{m}{2}^{-1} \sum_{i \neq j} \mathbb{I}\{T_i < T_j\} \cdot \mathbb{I}\{\hat{T}_i < \hat{T}_j\}, \tag{S3.2}$$

where $\mathbb{I}(\cdot)$ is the indicator function and m is the number of uncensored event time and $T_i, \hat{T}_i, i = 1, \ldots, m$ are corresponding event time and expected event time. This concordance index allows us to justify the superior performance of our proposed method compared to other alternative methods including functional Cox model and functional accelerated failure time model (Qu et al., 2016; Liu et al., 2024). We considered the following three models: (i) Model 1, which includes only three scalar-type covariates—age, gender, and Charlson morbidity index; (ii) Model 2, which adds an individual's SOFA score on the first day as an additional covariate; and (iii) Model 3, which incorporates the same three scalar-type covariates plus the SOFA score from the first five days as a functional covariate. The results are presented in Table 3. These results indicate that the model including functional covariate significantly improves prediction accuracy, and our proposed model outperforms both the functional Cox Model and the functional Accelerated Failure Time Model.

Table 3: Comparison of C-index across different statistical models. "/" represents not available.

Model \ Covariate	**Model 1**	**Model 2**	**Model 3**
Cox Proportional Hazard model	0.57	0.63	/
Functional Cox model	/	/	0.71
Accelerated Failure Time model	0.60	0.63	/
Functional Accelerated Failure Time model	/	/	0.73
Functional ODE model	0.59	0.62	0.75

Model 1 includes only three scalar-type covariates—age, gender, and Charlson morbidity index.

Model 2 adds an individual's SOFA score on day 1 as an additional covariate.

Model 3 further incorporates the SOFA score from the first five days as a functional covariate.

S4 Proof of lemmas

We begin with some notations. Let $\mathcal{F}_p([a,b])$ as the class of functions f on $[a,b]$ with bounded derivatives $f^{(j)}, j = 1,\dots,k$ and the $k-$th derivative satisfies the m-Hölder continuity condition:

$$|f^{(k)}(s) - f^{(k)}(t)| \leq M|s-t|^m, \quad \forall s,t \in [a,b].$$

Let $\mathcal{F}^{p_1} := \mathcal{F}^{p_1}([0,1]), \Gamma^{p_2} := \{\gamma \in \mathcal{F}^{p_2}([0,\tau]) : \gamma(0) = 0\}$ and $\mathcal{G}^{p_3} := \mathcal{F}^{p_3}([0,\mu])$.

Lemma 1 (Existence and uniqueness)**.** *Under Conditions (A1)-(A5), For any* $x \in \mathbb{R}^d, z \in L^2([0,1]), \alpha \in \mathcal{B}, \beta \in \mathcal{F}^{p_1}, \gamma \in \Gamma^{p_2}, g \in \mathcal{G}^{p_3}$*, the initial problem*

$$\begin{cases} \Lambda_u'(t) = \exp\left(\alpha^T x + \int_K \beta(s)z(s)ds + \gamma(t) + g(\Lambda_u(t))\right) \\ \Lambda_u(0) = 0 \end{cases} \tag{S4.3}$$

has exactly one bounded and continuous solution $\Lambda_u(t)$ *on* $[0,\tau]$*. The first and second partial derivatives of* $\Lambda_u(t)$ *with respect to* $\alpha \in \mathcal{B}$ *and Fréchet derivatives with respect to* $\beta \in \mathcal{F}^{p_1}, \gamma \in \Gamma^{p_2}, g \in \mathcal{G}^{p_3}$ *are also bounded and continuous on* $[0,\tau]$*.*

Proof of Lemma 1

Let $f(t,\Lambda) = \exp\left(\alpha^T x + \int_K \beta(s)z(s)ds + \gamma(t) + g(\Lambda(t))\right)$, then followed by mean-value theorem, for any $(t, \Lambda_1), (t, \Lambda_2) \in [0, \tau] \times [0, \mu]$, we have

$$
\begin{aligned}
& |f(t, \Lambda_1) - f(t, \Lambda_2)| \\
\leq\ & \exp\left(\alpha^T x + \int_K \beta(s) z(s) ds + \gamma(t) + g(\tilde{\Lambda})\right) |g'(\tilde{\Lambda})| \cdot |\Lambda_1 - \Lambda_2| \\
\leq\ & L |\Lambda_1 - \Lambda_2|,
\end{aligned}
$$

where $\tilde{\Lambda}$ is a point between Λ_1 and Λ_2, $L < \infty$. This implies that $f(t, \Lambda)$ satisfies the Lipschitz condition with respect to Λ in $[0, \tau] \times [0, \mu]$. According to Theorem 10.VI in Walter (1998), there is exactly one solution to the initial value problem (S4.3). The solution $\Lambda(t, x, z, \theta)$ satisfies

$$
\Lambda(t, x, z, \theta) = \int_0^t \exp\left(\alpha^T x + \int_K \beta(u) z(u) du + \gamma(s) + g(\Lambda_u(s, x, z, \theta))\right) ds. \tag{S4.4}
$$

For simplicity of notation, we denote $\Lambda(t, x, z, \theta)$ by $\Lambda(t)$. Then the following initial value problems have unique, bounded and continuous solutions.

$$
\begin{aligned}
\frac{d\Lambda'_\alpha(t)}{dt} &= f(t, \Lambda(t))(x + g'(\Lambda(t))\Lambda'_\alpha(t)), \Lambda'_\alpha(t) = 0, && \text{(S4.5)} \\
\frac{d\Lambda'_\beta(t)[h]}{dt} &= f(t, \Lambda(t))(\int_K \beta(s) h(s) ds + g'(\Lambda(t))\Lambda'_\beta(t)[h]), \Lambda'_\beta(0)[h] = 0, && \text{(S4.6)} \\
\frac{d\Lambda'_\gamma(t)[v]}{dt} &= f(t, \Lambda(t))(v(t) + g'(\Lambda(t))\Lambda'_\gamma(t)[v]), \\
& \quad \Lambda'_\gamma(0)[v] = 0, && \text{(S4.7)}
\end{aligned}
$$

$$\frac{d\Lambda'_g(t)[w]}{dt} = f(t,\Lambda(t))(w(\Lambda(t)) + g'(\Lambda(t))\Lambda'_g(t)[w]), \Lambda'_g(0)[w] = 0, \tag{S4.8}$$

$$\begin{aligned}\frac{d\Lambda''_{\alpha\alpha}(t)}{dt} &= f(t,\Lambda(t))\big\{(x + g'(\Lambda(t))\Lambda'_\alpha(t))(x + g'(\Lambda(t))\Lambda'_\alpha(t))^T \\ &\quad + g''(\Lambda(t))\Lambda'_\alpha(t)\Lambda'_\alpha(t)^T + g'(\Lambda(t))\Lambda''_{\alpha\alpha}(t)\big\}, \Lambda''_{\alpha\alpha}(0) = 0,\end{aligned} \tag{S4.9}$$

$$\begin{aligned}\frac{d\Lambda''_{\alpha\beta}(t)[h]}{dt} &= f(t,\Lambda(t))\big\{(x + g'(\Lambda(t))\Lambda'_\alpha(t))(\int_K \beta(s)h(s)ds + g'(\Lambda(t))\Lambda'_\beta(t)[h]) \\ &\quad + g''(\Lambda(t))\Lambda'_\alpha(t)\Lambda'_\beta(t)[h] + g'(\Lambda(t))\Lambda''_{\alpha\beta}(t)[h]\big\}, \\ &\quad \Lambda''_{\alpha\beta}(0)[h] = 0,\end{aligned} \tag{S4.10}$$

$$\begin{aligned}\frac{d\Lambda''_{\alpha\gamma}(t)[h]}{dt} &= f(t,\Lambda(t))\big\{(x + g'(\Lambda(t))\Lambda'_\alpha(t))(v(t) + g'(\Lambda(t))\Lambda'_\gamma(t)[v]) \\ &\quad + g''(\Lambda(t))\Lambda'_\alpha(t)\Lambda'_\gamma(t)[v] + g'(\Lambda(t))\Lambda''_{\alpha\gamma}(t)[v]\big\}, \\ &\quad \Lambda''_{\alpha\gamma}(0)[v] = 0,\end{aligned} \tag{S4.11}$$

$$\begin{aligned}\frac{d\Lambda''_{\alpha g}(t)[w]}{dt} &= f(t,\Lambda(t))\big\{(x + g'(\Lambda(t))\Lambda'_\alpha(t))(w(\Lambda(t)) + g'(\Lambda(t))\Lambda'_g(t)[w]) \\ &\quad + g''(\Lambda(t))\Lambda'_\alpha(t)\Lambda'_g(t)[v] + g'(\Lambda(t))\Lambda''_{\alpha g}(t)[w]\big\}, \\ &\quad \Lambda''_{\alpha g}(0)[w] = 0.\end{aligned} \tag{S4.12}$$

Now we prove that the solution of (S4.5), $\Lambda'_\alpha(t)$, is the derivative of $\Lambda(t)$

with respect to α. Using equation (S4.4) and (S4.5), we have

$$\begin{aligned}
& \limsup_{\delta \to 0} \frac{1}{|\delta|} |\Lambda(t,x,z,\alpha+\delta,\beta,\gamma,g) - \Lambda(t,x,z,\alpha,\beta,\gamma,g) - \Lambda'_\alpha(t)^T\delta| \\
= & \limsup_{\delta \to 0} \frac{1}{|\delta|} \Big| \int_0^t \exp\left((\alpha+\delta)^T x + \int_K \beta(u)z(u)du + \gamma(s) + g(\Lambda(s,x,z,\alpha+\delta,\beta,\gamma,g)) \right) \\
& - \exp\left(\alpha^T x + \int_K \beta(u)z(u)du + \gamma(s) + g(\Lambda(s,x,z,\alpha,\beta,\gamma,g)) \right) \\
& - \exp\left(\alpha^T x + \int_K \beta(u)Z(u)ds + \gamma(s) + g(\Lambda(s)) \right) (x + g'(\Lambda(s))\Lambda'_\alpha(s))ds \Big| \\
\le & \limsup_{\delta \to 0} \frac{1}{|\delta|} \int_0^t \Big| \exp\left((\alpha+\delta)^T x + \int_K \beta(u)z(u)du + \gamma(s) + g(\Lambda(s,x,z,\alpha+\delta,\beta,\gamma,g)) \right) \\
& - \exp\left(\alpha^T x + \int_K \beta(u)z(u)du + \gamma(s) + g(\Lambda(s,x,z,\alpha,\beta,\gamma,g)) \right) \\
& - \exp\left(\alpha^T x + \int_K \beta(u)Z(u)ds + \gamma(s) + g(\Lambda(s)) \right) (x + g'(\Lambda(s))\Lambda'_\alpha(s))\delta \Big| ds \\
\le & \int_0^t \limsup_{\delta \to 0} \frac{1}{|\delta|} \Big| \exp\left((\alpha+\delta)^T x + \int_K \beta(u)z(u)du + \gamma(s) + g(\Lambda(s,x,z,\alpha+\delta,\beta,\gamma,g)) \right) \\
& - \exp\left(\alpha^T x + \int_K \beta(u)z(u)du + \gamma(s) + g(\Lambda(s)) \right) \\
& - \exp\left(\alpha^T x + \int_K \beta(u)Z(u)ds + \gamma(s) + g(\Lambda(s)) \right) (x + g'(\Lambda(s))\Lambda'_\alpha(s))\delta \Big| ds \\
= & \int_0^t \exp\left(\alpha^T x + \int_K \beta(u)z(u)du + \gamma(s) + g(\Lambda(s)) \right) \cdot g'(\Lambda(s)) \\
& \times \{ \limsup_{\delta \to 0} \frac{1}{|\delta|} |\Lambda(s,x,z,\alpha+\delta,\beta,\gamma,g) - \Lambda(s) - \Lambda'_\alpha(s)^T\delta| \} ds.
\end{aligned}$$

It then follows from Gronwall's Lemma (Walter, 1998) that

$$\limsup_{\delta \to 0} \frac{1}{|\delta|} |\Lambda(s,x,z,\alpha+\delta,\beta,\gamma,g) - \Lambda(s) - \Lambda'_\alpha(s)^T\delta| \le 0, \qquad \text{(S4.13)}$$

which indicates that the solution of (S4.5) is the derivative of $\Lambda(t)$ with respect to α. The other first and second order derivatives of $\Lambda(t)$ with respect to β, γ, g can be proved to be the solution of (S4.6)-(S4.9) using similar argument as before. The details is omitted.

Lemma 2 (Spline Approximation)**.** *Under Conditions (A5) and (A6), for any* $\beta_0 \in \mathcal{F}^{p_1}, \gamma_0 \in \Gamma^{p_2}$ *and* $g_0 \in \mathcal{G}^{p_3}$*, there exist functions* $\beta_{0n} \in \mathcal{F}_n^{p_1}, \gamma_{0n} \in \Gamma_n^{p_2}$ *and* $g_{0n} \in \mathcal{G}_n^{p_3}$ *such that*

$$\|\beta_{0n} - \beta_0\|_\infty = O(n^{-\nu_1 p_1}),$$

$$\|\gamma_{0n} - \gamma_0\|_\infty = O(n^{-\nu_2 p_2}),$$

$$\|g_{0n} - g_0\|_\infty = O(n^{-\nu_3 p_3}),$$

where ν_1, ν_2 *and* ν_3 *are defined in Assumption (A6).*

Proof of Lemma 2

This Lemma is a direct result according to Corollary 6.21 in Schumaker (2007).

Lemma 3 (Bracketing number)**.** *Let* $\theta_{0n} = (\alpha_0, \beta_{0n}, \gamma_{0n}, \zeta_{0n}(\cdot, \alpha_0, \beta_{0n}, \gamma_{0n}))$*, denote*

$$\mathcal{F}_n = \{l(\theta, W) - l(\theta_{0n}, W), \theta \in \Theta_n\}.$$

Then, under Conditions (A1)-(A6), the ε*-bracketing number associated with* $\|\cdot\|_\infty$ *norm of* $\mathcal{F}_n$ *satisfies*

$$N_{[\,]}(\varepsilon, \mathcal{F}_n, \|\cdot\|_\infty) \lesssim (1/\varepsilon)^{c_1 q_{n_1} + c_2 q_{n_2} + c_3 q_{n_3} + d}$$

for some positive constants $c_1, c_2, c_3 > 0$, where $q_{n_1}, q_{n_2}, q_{n_3}$ and d appeared in Assumption (A3) and (A6).

Proof of Lemma 3

Following the calculation of Shen and Wong (1994) on page 597, for any $\varepsilon > 0$, there exists sets of ε-brackets

$$\{[\beta_i^L, \beta_i^U], \|\beta_i^U - \beta_i^L\|_\infty \leq \varepsilon, i = 1, 2, \ldots, \lceil (1/\varepsilon)^{c_1 K_{n1}} \rceil\},$$

$$\{[\gamma_j^L, \gamma_j^U], \|\gamma_j^U - \gamma_j^L\|_\infty \leq \varepsilon, j = 1, 2, \ldots, \lceil (1/\varepsilon)^{c_2 K_{n2}} \rceil\},$$

$$\{[g_k^L, g_k^U], \|g_k^U - g_k^L\|_\infty \leq \varepsilon, k = 1, 2, \ldots, \lceil (1/\varepsilon)^{c_3 K_{n3}} \rceil\}$$

such that for any $\beta \in \mathcal{F}_n^{p_1}, \gamma \in \Gamma_n^{p_2}, g \in \mathcal{G}_n^{p_3}$,

$$\beta_i^L(t) \leq \beta(t) \leq \beta_i^U(t), \quad t \in [0, 1],$$

$$\gamma_j^L(t) \leq \gamma(t) \leq \gamma_j^U(t), \quad t \in [0, \tau],$$

$$g_k^L(t) \leq \beta(t) \leq g_k^U(t), \quad t \in [0, \mu]$$

holds for some i, j, k. This leads to $\int_0^1 \left|(\beta(s) - \beta_i^L(s))Z(s)\right| ds \leq c_5\varepsilon$ for some constant c_5. Moreover, under condition (A1), $\mathcal{B} \subseteq \mathbb{R}^d$ is compact. $\mathcal{B}$ can be covered by $\lceil c_4(1/\varepsilon)^d \rceil$ balls with radius ε. Hence, there exists $\{\alpha_l : l = 1, 2, \ldots \lceil c_4(1/\varepsilon)^d \rceil\}$, for any $\alpha \in \mathcal{B}$, there exists $l \in \{1, 2, \ldots \lceil c_4(1/\varepsilon)^d \rceil\}$ such that $|X^T(\alpha - \alpha_l)| \leq c_6\varepsilon$ for some constant c_6. For any $\eta = (\alpha, \beta)$, define $\eta_{l,i} = (\alpha_l, \beta_i^L)$. Next we verify that $|\Lambda(t, U, \alpha, \beta, \gamma, g) - \Lambda(t, U, \alpha_l, \beta_i^L, \gamma_j^L, g_k^L)| \leq C\epsilon$ for all $t \in [0, \tau]$. For notation simplicity, denote $\Lambda(t, U, \alpha, \beta, \gamma, g) =$

$\Lambda_\theta(t), \Lambda(t, U, \alpha_l, \beta_i^L, \gamma_j^L, g_k^L) = \Lambda_{ijkl}(t)$, then $\Lambda_\theta(t)$ and $\Lambda_{ijkl}(t)$ satisfies

$$\begin{aligned}
\int_0^t \exp\left(\alpha^T X + \int \beta(u)Z(u)du + \gamma(s)\right) ds &= \int_0^{\Lambda_\theta(t)} \exp(-g(s))ds, \\
\int_0^t \exp\left(\alpha_l^T X + \int \beta_i^L(u)Z(u)du + \gamma_j^L(s)\right) ds &= \int_0^{\Lambda_{ijkl}(t)} \exp(-g_k^L(s))ds,
\end{aligned}$$

which leads to

$$\begin{aligned}
\left|\int_{\Lambda_{ijkl}(t)}^{\Lambda_\theta(t)} \exp(-g_0(s))ds\right| &= \Bigg| \int_0^t \exp\left(\alpha^T X + \int \beta(u)Z(u)du + \gamma(s)\right) ds \\
&\quad - \int_0^t \exp\left(\alpha_l^T X + \int \beta_i^L(u)Z(u)du + \gamma_j^L(s)\right) ds \\
&\quad - \int_0^{\Lambda_\theta(t)} \exp(-g(s)) - \exp(-g_k^L(s))ds\Bigg|.
\end{aligned}$$

Under conditions (A1)-(A6), the right hand side is upper bounded by $\tau M_1(c_5 + c_6 + 1)\epsilon + \mu M_2\epsilon$, and the left hand side is no less than $e^c|\Lambda_{1\theta}(t) - \Lambda_\theta(t)|$,where $c = \min_{g\in\mathcal{G}_n^{p_3}, t\in[0,\tau]}(-g(t))$. Therefore, $|\Lambda(t, U, \alpha, \beta, \gamma, g) - \Lambda(t, U, \alpha_l, \beta_i^L, \gamma_j^L, g_k^L)| \leq c_7\epsilon$ for some constant c_7.

Next, we construct a set of ϵ-bracket for $\mathcal{F}_n$. Given $\theta \in \Theta_n$, denote

$$\begin{aligned}
m(\theta, W) &= l(\theta, W) - l(\theta_{0n}, W) \\
&= \Delta\left(\alpha^T X + \int \beta(s)Z(s)ds + \gamma(Y) + g(\Lambda(Y, U, \theta))\right) \\
&\qquad -\Lambda(Y) - l(\theta_{0n}, W).
\end{aligned}$$

Define

$$\begin{aligned}
m_{ijkl}^L(W) &= \Delta\left(\alpha_l^T X - c_6\varepsilon + \int \beta_i^L(s)Z(s)ds - c_5\varepsilon + \gamma_j^L(Y) + g_k^L(c_{ijlk}^L)\right) \\
&\qquad -\Lambda_{ijkl}(Y) - c_7\varepsilon - l(\theta_{0n}, W),
\end{aligned}$$

$$
\begin{aligned}
m^U_{ijkl}(W) &= \Delta\left(\alpha_l^T X + c_6\varepsilon + \int \beta_i^L(s)Z(s)ds + c_5\varepsilon + \gamma_j^U(Y) + g_k^U(c^U_{ijlk})\right) \\
&\quad -\Lambda_{ijkl}(Y) + c_7\varepsilon - l(\theta_{0n}, W),
\end{aligned}
$$

where c^L_{ijlk}, c^U_{ijlk} are the minimum and maximum point of g_k^L and g_k^U within the interval $[\Lambda_{ijkl}(Y) - c_7\varepsilon, \Lambda_{ijkl}(Y) + c_7\varepsilon]$. Noted that

$$
\begin{aligned}
& |m^U_{ijkl}(W) - m^L_{ijkl}(W)| \\
= \ & |\Delta(2(c_5 + c_6)\epsilon + \gamma_j^U(Y) - \gamma_j^L(Y) + g_k^U(c^U_{ijlk}) - g_k^L(c^L_{ijlk}) + 2c_7\varepsilon| \\
\leq \ & 2(c_5 + c_6 + c_7)\varepsilon + \|\gamma_j^U - \gamma_j^L\|_\infty + g_k^U(c^U_{ijlk}) - g_k^L(c^U_{ijlk}) \\
& \quad + g_k^L(c^U_{ijlk}) - g_k^L(c^L_{ijlk}) \\
\leq \ & 2(c_5 + c_6 + c_7)\varepsilon + \|\gamma_j^U - \gamma_j^L\|_\infty + \|g_k^U - g_k^L\|_\infty + \|\dot{g}_k^L\|_\infty |c^L_{ijlk} - c^U_{ijlk}| \\
\lesssim \ & \varepsilon.
\end{aligned}
$$

The last inequality follows from that $|c^L_{ijlk} - c^U_{ijlk}| < 2c_7\varepsilon$ and $\|\dot{g}_k^L\|_\infty$ is bounded by a constant under condition (A4). Therefore,

$$
\begin{aligned}
\Big\{ [m^L_{ijkl}(W), m^U_{ijkl}(W)] : \quad & i \in \{1, 2, \ldots, \lceil (1/\varepsilon)^{c_1 K_{n1}} \rceil\}; j \in \{1, 2, \ldots, \lceil (1/\varepsilon)^{c_2 K_{n2}} \rceil\} \\
& k \in \{1, 2, \ldots, \lceil (1/\varepsilon)^{c_3 K_{n3}} \rceil\}; l \in \{1, 2, \ldots \lceil c_4 (1/\varepsilon)^d \rceil\} \Big\}
\end{aligned}
$$

is a set of ε-bracket of the class $\mathcal{F}_n$ and the ε-bracketing number associated with $\|\cdot\|_\infty$ norm of $\mathcal{F}_n$ satisfies

$$
N_{[\,]}(\varepsilon, \mathcal{F}_n, \|\cdot\|_\infty) \leq (1/\varepsilon)^{c_1 K_{n1}} (1/\varepsilon)^{c_2 K_{n2}} (1/\varepsilon)^{c_3 K_{n3}} c_4 (1/\varepsilon)^d \lesssim (1/\varepsilon)^{c_1 q_{n_1} + c_2 q_{n_2} + c_3 q_{n_3} + d}.
$$

Lemma 4. *For* $1 \leq j \leq d$, *denote the following classes of functions* $\mathcal{F}^{\beta}_{n,j}(\eta) = \{l'_{\beta}(\theta, W)[h^*_{1j} - h_j] : \theta \in \Theta_n, h_j \in \mathcal{F}^2_n, d(\theta, \theta_0) \leq \eta, \|h^*_{1j} - h_j\|_\infty \leq \eta\}$, $\mathcal{F}^{\gamma}_{n,j}(\eta) = \{l'_{\gamma}(\theta, W)[h^*_{2j} - h_j] : \theta \in \Theta_n, h_j \in \Gamma^2_n, d(\theta, \theta_0) \leq \eta, \|h^*_{2j} - h_j\|_\infty \leq \eta\}$ *and* $\mathcal{F}^{\zeta}_{n,j}(\eta) = \{l'_{\zeta}(\theta, W)[\tilde{h}^*_{3j} - h_j] : \theta \in \Theta_n, h_j \in \mathcal{H}^2_n, d(\theta, \theta_0) \leq \eta, \|\tilde{h}^*_{3j} - h_j\|_\infty \leq \eta\}$, *where*

$$\tilde{h}^*_{3j}(\cdot, \alpha, \beta, \gamma) = h^*_{3j}(\Lambda(\cdot, \alpha, \beta, \gamma, g)) + g'(\Lambda(\cdot, \alpha, \beta, \gamma, g))\Lambda'_g(\cdot, \alpha, \beta, \gamma, g)$$

and $h^*_{1j}, h^*_{2j}, h^*_{3j}$ *are given in Theorem 2. Then under Conditions (A1)-(A6), we have*

$$\begin{aligned} N_{[\ ]}(\epsilon, \mathcal{F}^{\beta}_{n,j}(\eta), \|\cdot\|_\infty) &\lesssim \left(\frac{\eta}{\epsilon}\right)^{c_1 q_{n1} + c_2 q_{n2} + c_3 q_{n3} + d}, \\ N_{[\ ]}(\epsilon, \mathcal{F}^{\gamma}_{n,j}(\eta), \|\cdot\|_\infty) &\lesssim \left(\frac{\eta}{\epsilon}\right)^{c_1 q_{n1} + c_2 q_{n2} + c_3 q_{n3} + d}, \\ N_{[\ ]}(\epsilon, \mathcal{F}^{\zeta}_{n,j}(\eta), \|\cdot\|_\infty) &\lesssim \left(\frac{\eta}{\epsilon}\right)^{c_1 q_{n1} + c_2 q_{n2} + c_3 q_{n3} + d} \end{aligned}$$

for some positive constants c_1, c_2 *and* c_3.

Lemma 5. *For* $1 \leq j \leq d$, *define the following class of functions*

$$\begin{aligned} \mathcal{F}^{*\alpha}_{n,j}(\eta) &= \{l'_{\alpha_j}(\theta, W) - l'_{\alpha_j}(\theta_0, W) : d(\theta, \theta_0), \|g'(\Lambda(\cdot, \theta)) - g'_0(\Lambda_0(\cdot))\|_2 \leq \eta\}, \\ \mathcal{F}^{*\beta}_{n,j}(\eta) &= \{l'_{\beta}(\theta, W) - l'_{\beta}(\theta_0, W)[h^*_{1j}] : d(\theta, \theta_0), \|g'(\Lambda(\cdot, \theta)) - g'_0(\Lambda_0(\cdot))\|_2 \leq \eta\}, \\ \mathcal{F}^{*\gamma}_{n,j}(\eta) &= \{l'_{\gamma}(\theta, W) - l'_{\gamma}(\theta_0, W)[h^*_{2j}] : d(\theta, \theta_0), \|g'(\Lambda(\cdot, \theta)) - g'_0(\Lambda_0(\cdot))\|_2 \leq \eta\}, \\ \mathcal{F}^{*\zeta}_{n,j}(\eta) &= \{l'_{\zeta}(\theta, W) - l'_{\zeta}(\theta_0, W)[\tilde{h}^*_{3j}] : d(\theta, \theta_0), \|g'(\Lambda(\cdot, \theta)) - g'_0(\Lambda_0(\cdot))\|_2 \leq \eta\}, \end{aligned}$$

where $\tilde{h}^*_{3j}(\cdot, \alpha, \beta, \gamma) = h^*_{3j}(\Lambda(\cdot, \theta)) + g'(\Lambda(\cdot, \theta))\Lambda'_g(\cdot, \theta)[h^*_{3j}]$ *and* $h^*_{1j}, h^*_{2j}, h^*_{3j}$

are defined in Theorem 2. Then under Conditions (A1)-(A6), we have

$$
\begin{aligned}
N_{[\ ]}(\epsilon, \mathcal{F}^{*\alpha}_{n,j}(\eta), \|\cdot\|_\infty) &\lesssim (\frac{\eta}{\epsilon})^{c_1 q_{n1} + c_2 q_{n2} + c_3 q_{n3} + d}, \\
N_{[\ ]}(\epsilon, \mathcal{F}^{*\beta}_{n,j}(\eta), \|\cdot\|_\infty) &\lesssim (\frac{\eta}{\epsilon})^{c_1 q_{n1} + c_2 q_{n2} + c_3 q_{n3} + d}, \\
N_{[\ ]}(\epsilon, \mathcal{F}^{*\gamma}_{n,j}(\eta), \|\cdot\|_\infty) &\lesssim (\frac{\eta}{\epsilon})^{c_1 q_{n1} + c_2 q_{n2} + c_3 q_{n3} + d}, \\
N_{[\ ]}(\epsilon, \mathcal{F}^{*\zeta}_{n,j}(\eta), \|\cdot\|_\infty) &\lesssim (\frac{\eta}{\epsilon})^{c_1 q_{n1} + c_2 q_{n2} + c_3 q_{n3} + d}.
\end{aligned}
$$

for some positive constants c_1, c_2 and c_3.

Proof of Lemma 4 and 5

The proof of Lemma 4 and Lemma 5 are similar to the bracketing number calculation in Lemma 3, the details are omitted.

Lemma 6 (Bounded operator)**.** *Let $\psi(t, x, z, \alpha, \beta, \gamma, g) = \alpha^T x + \int_0^1 \beta(s) z(s) ds + \gamma(t) + g(\Lambda_{x,z}(t))$. Denote the derivatives of $\psi(t, x, z, \alpha, \beta, \gamma, g)$ with respect to β, γ, g at the true parameter $(\alpha_0, \beta_0, \gamma_0, g_0)$ by $\psi'_{0\beta}, \psi'_{0\gamma}, \psi'_{0g}$, respectively. For any*

$\psi'_{0\beta}(\cdot)[h] \in \mathcal{E}_\beta = \{\psi'_{0\beta}(\cdot)[h] : \psi'_{0\beta}(t, u)[h], t \in [0, \tau], u \in \mathcal{U} = \mathcal{X} \times L_2([0, 1]), h \in \mathcal{F}^{p_1}\}$,

the L_2-norm of $\psi'_{0\beta}(\cdot)[h]$ is defined as

$$
\|\psi'_{0\beta}(\cdot)[h]\|_2 = \left[\int_{\mathcal{U}} \int_0^\tau (\psi'_{0\beta}(t, u)[h])^2 d\Lambda_0(t, u) dF_U(u)\right]^{1/2}.
$$

The L_2-norm of $\psi'_{0\gamma}(\cdot)[v]$, $\psi'_{0g}(\cdot)[w]$ are similarly defined. Then, under Conditions (A1)-(A5),$\psi'_{0\beta}(\cdot) : h \to \psi'_{0\beta}(\cdot)[h]$, $\psi'_{0\gamma}(\cdot) : v \to \psi'_{0\gamma}(\cdot)[v]$ and

$\psi'_{0g}(\cdot) : w \to \psi'_{0g}(\cdot)[w]$ *are bounded linear operators and are bounded from below, i.e.*

$$\|\psi'_{0\beta}(\cdot)[h]\|_2 \gtrsim \|h\|_2, \quad \text{for any } h \in \mathcal{F}^{p_1},$$

$$\|\psi'_{0\gamma}(\cdot)[v]\|_2 \gtrsim \|v\|_2, \quad \text{for any } v \in \Gamma^{p_2},$$

$$\|\psi'_{0g}(\cdot)[w]\|_2 \gtrsim \|w\|_2, \quad \text{for any } w \in \mathcal{G}^{p_3}.$$

Proof of Lemma 6

By solving the initial problems in Lemma 1 and some direct calculations, we have

$$\begin{aligned}
\psi'_{0\beta}(t,u)[h] &= g'_0(\Lambda_0(t,u))\Lambda'_{0\beta}(t,u)[h] + \int z(s)h(s)ds \\
&= \left\{ g'_0(\Lambda_0(t,u)) \exp\left(\alpha_0^T x + \int z(s)\beta_0(s)ds + g_0(\Lambda_0(t,u))\right) \right. \\
&\qquad \left. \times \int_0^t \exp(\gamma_0(s))ds + 1 \right\} \int_K z(s)u(s)ds, \\
\psi'_{0\gamma}(t,u)[v] &= g'_0(\Lambda_0(t,u))\Lambda'_{0\gamma}(t,u)[v] + v(t) \\
&= g'_0(\Lambda_0(t,u)) \exp\left(\alpha_0^T x + \int z(s)\beta_0(s)ds + g_0(\Lambda_0(t,u))\right) \\
&\qquad \times \int_0^t \exp(\gamma_0(s))v(s)ds + v(t), \\
\psi'_{0g}(t,u)[w] &= g'_0(\Lambda_0(t,u))\Lambda'_{0g}(t,u) + w(\Lambda_0(t,u)) \\
&= g'_0(\Lambda_0(t,u)) \exp(g_0(\Lambda_0(t,u))) \int_0^{\Lambda_0(t,u)} w(s)\exp(-g_0(s))ds \\
&\qquad + w(\Lambda_0(t,u)).
\end{aligned}$$

We first prove that $\psi'_{0\gamma}(\cdot)[v]$ is bounded. Noted that

$$\|\psi'_{0\gamma}(\cdot)[v]\|_2^2 = \int_{\mathcal{U}}\int_0^\tau (\psi'_{0\gamma}(t,u)[v])^2 dt dF_U(u) \leq 2(I_1 + I_2), \tag{S4.14}$$

where

$$\begin{aligned} I_1 &= \int_{\mathcal{U}}\int_0^\tau \Big[g_0'(\Lambda_0(t,u)) \exp\left(\alpha_0^T x + \int z(s)\beta_0(s)ds + g_0(\Lambda_0(t,u))\right) \\ &\qquad \times \int_0^t \exp(\gamma_0(s))v(s)ds\Big]^2 d\Lambda_0(t,u) dF_{\mathcal{U}}(u), \\ I_2 &= \int_{\mathcal{U}}\int_0^\tau v^2(t) d\Lambda_0(t,u) dF_{\mathcal{U}}(u). \end{aligned}$$

Since the boundedness of α_0, β_0, x, z implies $\exp(\alpha_0^T x + \int z(s)\beta_0(s)ds + g_0(\Lambda_0(t,u)) < \infty$. It follows that $I_2 \lesssim \|v\|_2^2$. By Cauchy-Schwarz inequality, we have

$$\begin{aligned} \left(\int_0^t \exp(\gamma_0(s))v(s)ds\right)^2 &\leq \int_0^t \exp(2\gamma_0(s))ds \cdot \int_0^t v^2(s)ds \\ &\leq \tau \exp(\max_{s\in[0,\tau]} 2\gamma_0(s))\|v\|_2^2. \end{aligned}$$

Thus, based on the boundedness of $\alpha_0, \beta_0, x, z, g_0, g_0'$, we have $I_1 \lesssim \|v\|_2^2$. Therefore, $\|\psi'_{0\gamma}(\cdot)[v]\|_2 \lesssim \|v\|_2^2$. Similarily, we have

$$\begin{aligned} &\|\psi'_{0g}(\cdot)[w]\|_2^2 \\ = &\int_{\mathcal{U}}\int_0^\tau (\psi'_{0g}(t,u)[w])^2 dt dF_U(u) \end{aligned}$$

$$\leq 2\int_{\mathcal{U}}\int_0^{\tau}\left(g_0'(\Lambda_0(t,u))\exp(g_0(\Lambda_0(t,u)))\int_0^{\Lambda_0(t,u)} w(s)\exp(-g_0(s))ds\right)^2 d\Lambda_0(t,u)dF_U(u) + 2\int_{\mathcal{U}}\int_0^{\tau} w^2(\Lambda_0(t,u))d\Lambda_0(t,u)dF_U(u)$$

$$\leq \int_{\mathcal{U}}\int_0^{\Lambda_0(\tau,u)}\left(g_0'(t)\exp(g_0(t))\int_0^t w(s)\exp(-g_0(s))ds\right)^2 dtdF_U(u) + 2\int_{\mathcal{U}}\int_0^{\Lambda_0(\tau,u)} w^2(t)dtdF_U(u)$$

$$\leq \int_{\mathcal{U}}\int_0^{\mu}(g_0'(t))^2\exp(2g_0(t))\int_0^t \exp(-2g_0(s))ds\int_0^t w^2(s)dsdtdF_U(u) + 2\int_{\mathcal{U}}\int_0^{\mu} w^2(t)dtdF_U(u).$$

Combine with the boundedness of g_0 and g_0', $\|\psi_{0g}'(\cdot)[w]\|_2 \lesssim \|w\|_2$ follows. Next we show that $\psi_{0\gamma}'(\cdot)$ is bounded from below. Using the fact that $\psi_{0\gamma}'(t)[v] = v(t) + g_0'(\Lambda_0(t))\Lambda_{0\gamma}'(t)[v]$, we have

$$v(t) = \psi_{0\gamma}'(t)[v] - g_0'(\Lambda_0(t))\int_0^t \psi_{0\gamma}'(s)[v]d\Lambda_0(s).$$

Therefore, $\|\psi_{0\gamma}'(t)[v_1]-\psi_{0\gamma}'(t)[v_2]\|_2 = 0$ indicates $\int_{\mathcal{U}}\int_0^{\tau}(v_1(t)-v_2(t))^2dtdF_U(u) = 0$. Hence, $\psi_{0\gamma}'(\cdot)$ is a bijective operator. This indicates $\psi_{0\gamma}'(\cdot)$ is a bounded operator and is bounded from below. The boundedness of $\psi_{0\beta}'$ and ψ_{0g}' can be proved similarly, the details are omitted.

S5 Proof of Proposition 1

We first consider model (3) in the manuscript. If there exists at least one coordinate of X is continuous with a nonzero coefficient α, then follow the same argument as Horowitz (1996), we have $(\varphi(\cdot), \alpha, \beta, \varepsilon)$ and $(\tilde{\varphi}(\cdot), \tilde{\alpha}, \tilde{\beta}, \tilde{\varepsilon})$, specify the same distribution of T if and only if $\tilde{\varphi} = c_1\varphi + c$, $\tilde{\alpha} = c_1\alpha$, $\tilde{\beta} = c_1\beta$ and $\tilde{\varepsilon} = c_1\varepsilon + c$, for some constant c_1 and c. Then $\int_0^t \tilde{h}(u)du = \exp(\tilde{\varphi}(t)) = e^c \exp(c_1\varphi(t)) = e^c \left(\int_0^t h(u)du\right)^{c_1}$. Denote the survival function of $\exp(\tilde{\varepsilon})$ and $\exp(\varepsilon)$ by $\tilde{G}(t)$ and $G(t)$ respectively. Then we have

$$\tilde{G}(t) = \mathbb{P}\left(e^{\tilde{\varepsilon}} > t\right) = \mathbb{P}\left(e^{\varepsilon} > (te^{-c})^{1/c_1}\right) = G\left((te^{-c})^{1/c_1}\right),$$

It follows that $\tilde{G}^{-1}(t) = e^c\left(G^{-1}(t)\right)^{c_1}$. Using the fact that $\int_0^{-\ln t} \tilde{q}^{-1}(s)ds = \tilde{G}^{-1}(t)$ and $\int_0^{-\ln t} q^{-1}(s)ds = G^{-1}(t)$, we have

$$\int_0^t \tilde{q}^{-1}(s)ds = \tilde{G}^{-1}(e^{-t}) = e^c\left(G^{-1}(e^{-t})\right)^{c_1} = e^c\left(\int_0^t q^{-1}(s)ds\right)^{c_1}.$$

Take $c_2 = e^c$, then Proposition 1 follows.

S6 Proof of Theorem 2

The score vector for α_0, the score operator for coefficient function β_0, the score operator for γ_0 and g_0 are given by

$$\begin{aligned}
\dot{l}_{\alpha_0} &= X\int\left(g_0'(\tilde{\Lambda}_0(t))\exp\left(g_0(\tilde{\Lambda}_0(t))\right)t+1\right)dM(t) =: X\int\varepsilon_1(t)dM(t),\\
\dot{l}_{\beta_0}h_1 &= \left(\int_0^1 h_1(s)Z(s)ds\right)\int\left(g_0'(\tilde{\Lambda}_0(t))\exp\left(g_0(\tilde{\Lambda}_0(t))\right)t+1\right)dM(t)\\
&=: \left(\int_0^1 h_1(s)Z(s)ds\right)\int\varepsilon_1(t)dM(t),\\
\dot{l}_{\gamma_0}h_2 &= \int\left(g_0'(\tilde{\Lambda}_0(t))\exp\left(g_0(\tilde{\Lambda}_0(t))\right)\int_0^t h_2(L^{-1}(se^{-V}))ds + h_2(L^{-1}(te^{-V}))\right)dM(t),\\
&=: \int\varepsilon_2(t,V)[h_2]dM(t)\\
\dot{l}_{g_0}h_3 &= \int\left(g_0'(\tilde{\Lambda}_0(t))\exp\left(g_0(\tilde{\Lambda}_0(t))\right)\int_0^{\tilde{\Lambda}_0(t)}\exp(-g_0(s))h_3(s)ds + h_3(\tilde{\Lambda}_0(t))\right)dM(t)\\
&=: \int\varepsilon_3(t)[h_3]dM(t).
\end{aligned}$$

where $M(t) = \Delta I(R\le t) - \int_0^t I(R\ge s)d\tilde{\Lambda}_0(s)$ is a counting process martingale associated with the counting process $\left\{I_{\{R\le t\}}, t\ge 0\right\}$. Define

$$\begin{aligned}
\overline{\mathbb{T}}_{\beta_0} &= \left\{h_1\in\mathcal{F}^{p_1} : \mathbb{P}\left[\Delta\left(\int_0^1 h_1(s)Z(s)ds\right)^2\right] < \infty\right\},\\
\overline{\mathbb{T}}_{\gamma_0} &= \left\{h_2\in\Gamma^{p_2} : \mathbb{P}\left[\Delta\left(\varepsilon_2(R,V)[h_2]\right)^2\right] < \infty\right\},\\
\overline{\mathbb{T}}_{g_0} &= \left\{h_3\in\mathcal{G}^{p_3} : \mathbb{P}\left[\Delta\left(\varepsilon_3(R)[h_3]\right)^2\right] < \infty\right\}.
\end{aligned}$$

The efficient score function for α_0 is

$$l^*_{\alpha_0} = \dot{l}_{\alpha_0} - \Pi(\dot{l}_{\alpha_0}|\dot{\mathbf{P}}_1+\dot{\mathbf{P}}_2+\dot{\mathbf{P}}_3), \tag{S6.15}$$

where $\dot{\mathbf{P}}_1 = \{\dot{l}_{\beta_0} h : h \in \overline{\mathbb{T}}_{\beta_0}\}$, $\dot{\mathbf{P}}_2 = \{\dot{l}_{\gamma_0} h : h \in \overline{\mathbb{T}}_{\gamma_0}\}$, $\dot{\mathbf{P}}_3 = \{\dot{l}_{\gamma_0} h : h \in \overline{\mathbb{T}}_{g_0}\}$. Finding $\Pi(\dot{l}_{\alpha_0} | \dot{\mathbf{P}}_1 + \dot{\mathbf{P}}_2 + \dot{\mathbf{P}}_3)$ is equivalent to finding the vector function $(\mathbf{h}_1^*, \mathbf{h}_2^*, \mathbf{h}_3^*)$ such that

$$
\begin{aligned}
&\mathbb{E}\left\{\left(\dot{l}_{\alpha_0} - \dot{l}_{\beta_0}\mathbf{h}_1^* - \dot{l}_{\gamma_0}\mathbf{h}_2^* - \dot{l}_{g_0}\mathbf{h}_3^*\right)\dot{l}_{\beta_0} h\right\} = 0 \quad \text{for all} \quad h \in \overline{\mathbb{T}}_{\beta_0}, \\
&\mathbb{E}\left\{\left(\dot{l}_{\alpha_0} - \dot{l}_{\beta_0}\mathbf{h}_1^* - \dot{l}_{\gamma_0}\mathbf{h}_2^* - \dot{l}_{g_0}\mathbf{h}_3^*\right)\dot{l}_{\gamma_0} h\right\} = 0 \quad \text{for all} \quad h \in \overline{\mathbb{T}}_{\gamma_0}, \\
&\mathbb{E}\left\{\left(\dot{l}_{\alpha_0} - \dot{l}_{\beta_0}\mathbf{h}_1^* - \dot{l}_{\gamma_0}\mathbf{h}_2^* - \dot{l}_{g_0}\mathbf{h}_3^*\right)\dot{l}_{g_0} h\right\} = 0 \quad \text{for all} \quad h \in \overline{\mathbb{T}}_{g_0}.
\end{aligned} \tag{S6.16}
$$

Then $\Pi(\dot{l}_{\alpha_0} | \dot{\mathbf{P}}_1 + \dot{\mathbf{P}}_2 + \dot{\mathbf{P}}_3) = \dot{l}_{\beta_0}\mathbf{h}_1^* + \dot{l}_{\gamma_0}\mathbf{h}_2^* + \dot{l}_{g_0}\mathbf{h}_3^*$. This implies that

$$
(\mathbf{h}_1^*, \mathbf{h}_2^*, \mathbf{h}_3^*) = \underset{(\mathbf{h}_1, \mathbf{h}_2, \mathbf{h}_3) \in \overline{\mathbb{T}}_{\beta_0}^d \times \overline{\mathbb{T}}_{\gamma_0}^d \times \overline{\mathbb{T}}_{g_0}^d}{\arg\min} \mathbb{E}\left[\left\|\dot{l}_{\alpha_0} - \dot{l}_{\beta_0}\mathbf{h}_1 - \dot{l}_{\gamma_0}\mathbf{h}_2 - \dot{l}_{g_0}\mathbf{h}_3\right\|^2\right], \tag{S6.17}
$$

which is equivalent to minimizing

$$
\begin{aligned}
&\mathbb{E}\Bigg[\Delta\Bigg\| \left(g_0'(\tilde{\Lambda}_0(R)) \exp\left(g_0(\tilde{\Lambda}_0(R))\right) R + 1\right)\left(X - \int_0^1 \mathbf{h}_1(s) Z(s) ds\right) \\
&-g_0'(\tilde{\Lambda}_0(R)) \exp\left(g_0(\tilde{\Lambda}_0(R))\right) \int_0^R \mathbf{h}_2(L^{-1}(se^{-V})) ds + \mathbf{h}_2(L^{-1}(Re^{-V})) \\
&g_0'(\tilde{\Lambda}_0(R)) \exp\left(g_0(\tilde{\Lambda}_0(R))\right) \int_0^{\tilde{\Lambda}_0(R)} \exp(-g_0(s)) \mathbf{h}_3(s) ds + \mathbf{h}_3(\tilde{\Lambda}_0(R)) \Bigg\|^2\Bigg].
\end{aligned}
$$

Since the space $\overline{\mathbb{T}}_{\beta_0} \times \overline{\mathbb{T}}_{\gamma_0} \times \overline{\mathbb{T}}_{g_0}$ is closed, so that the solution is well defined and the efficient score is

$$\begin{aligned} \dot{l}^*_{\alpha_0} &= \int \Big\{ \left(g_0'(\tilde{\Lambda}_0(t)) \exp\left(g_0(\tilde{\Lambda}_0(t))\right) t + 1\right) \left(X - \int_0^1 \mathbf{h}_1^*(s)\beta_0(s) ds \right) \\ &\quad - g_0'(\tilde{\Lambda}_0(t)) \exp\left(g_0(\tilde{\Lambda}_0(t))\right) \int_0^t \mathbf{h}_2^*(L^{-1}(se^{-V}))ds + \mathbf{h}_2^*(L^{-1}(se^{-V})) \\ &\quad - g_0'(\tilde{\Lambda}_0(t)) \exp\left(g_0(\tilde{\Lambda}_0(t))\right) \int_0^{\tilde{\Lambda}_0(t)} \exp(-g_0(s))\mathbf{h}_3^*(s)ds + \mathbf{h}_3^*(\tilde{\Lambda}_0(t)) \Big\} dM(t), \end{aligned}$$

and the information matrix is $I(\alpha_0) = \mathbb{P}\left\{ \dot{l}^{*\otimes 2}_{\alpha_0} \right\}$.

S7 Proof of Theorem 3

For any $\theta_1 = (\alpha_1, \beta_1, \gamma_1, \zeta_1), \theta_2 = (\alpha_2, \beta_2, \gamma_2, \zeta_2)$, define a pseudometric $d(\cdot, \cdot)$ as

$$\begin{aligned} d(\theta_1, \theta_2)^2 &= E\left[(\alpha_1 - \alpha_2)^T X + \int Z(s)(\beta_1 - \beta_2)(s) ds \right]^2 \\ &\quad + \|\gamma_1 - \gamma_2\|_2^2 + \|\zeta_1(\cdot, \alpha_1, \beta_1, \gamma_1) - \zeta_2(\cdot, \alpha_2, \beta_2, \gamma_2)\|_2^2 . \end{aligned}$$

We first prove that $d(\hat{\theta}, \theta_0) = O_p(n^{-c})$. Following the proof of theorem 1 in Shen and Wong(1994), the proof proceeds by verifying their conditions (C1)-(C3).

First we verify condition (C1). Denote

$$A = (\alpha - \alpha_0)^T X + \int_0^1 (\beta - \beta_0)(s) Z(s) ds + \gamma(Y) - \gamma_0(Y) + g\left(\Lambda(Y, U, \theta)\right) - g_0\left(\Lambda_0(Y, U)\right).$$

Direct calculation yields

$$Pl_0(\theta_0, Y) - Pl(\theta, Y) = P\left(\Delta\left\{\exp(A) - 1 - A\right\}\right).$$

Using Taylor expansion, we have

$$Pl_0(\theta_0, W) - Pl(\theta, W) = \frac{1}{2}P\left\{\Delta A^2\right\} + o\left(P\left\{\Delta A^2\right\}\right).$$

Noted that

$$\begin{aligned}
P\left\{\Delta A^2\right\} &= P\Delta\bigg[(\alpha-\alpha_0)^T X + \int_0^1 (\beta-\beta_0)(s)Z(s)ds + \gamma(Y) - \gamma_0(Y) \\
&\quad + g\left(\Lambda(Y,U,\theta)\right) - g\left(\Lambda_0(Y,U,\theta)\right) + g\left(\Lambda_0(Y,U,\theta)\right) - g_0\left(\Lambda_0(Y,U)\right)\bigg]^2 \\
&= P\bigg\{\Delta\bigg[(g'(\Lambda_0(Y,U))\Lambda'_{0\alpha}(Y,U) + X)^T(\alpha-\alpha_0) \\
&\quad + g'_0(\Lambda_0(Y,U))\Lambda'_{0\beta}(Y,U)[\beta-\beta_0] + \int_0^1 Z(s)(\beta-\beta_0)(s)ds \\
&\quad + g'_0(\Lambda_0(Y,U))\Lambda'_{0\gamma}(Y,U)[\gamma-\gamma_0] + \gamma(Y) - \gamma_0(Y) \\
&\quad + g'_0(\Lambda_0(Y,U))\Lambda'_{0g}(Y,U)[g-g_0] + g(\Lambda_0(Y,U)) - g_0(\Lambda_0(Y,U)) \\
&\quad + o(\|\alpha-\alpha_0\|_2) + o(\|\beta-\beta_0\|_C) + o(\|\gamma-\gamma_0\|_2) + o(\|g-g_0\|_2)\bigg]^2\bigg\} \\
&\gtrsim P\bigg\{\Delta\bigg[(g'_0(\Lambda_0(Y,U))\Lambda'_{0\alpha}(Y,U) + X)^T(\alpha-\alpha_0) \\
&\quad + g'_0(\Lambda_0(Y,U))\Lambda'_{0\beta}(Y,U)[\beta-\beta_0] + \int_0^1 Z(s)(\beta-\beta_0)(s)ds \\
&\quad + g'_0(\Lambda_0(Y,U))\Lambda'_{0\gamma}(Y,U)[\gamma-\gamma_0] + \gamma(Y) - \gamma_0(Y) \\
&\quad + g'_0(\Lambda_0(Y,U))\Lambda'_{0g}(Y,U)[g-g_0] + g(\Lambda_0(Y,U)) - g_0(\Lambda_0(Y,U))\bigg]^2\bigg\} \\
&\quad + o(d^2(\theta_0,\theta)), \qquad \text{(S7.18)}
\end{aligned}$$

where the first equality is obtained by Taylor expasion, the second equality holds under condition (A1)-(A5) and the fact that $\Lambda'_{0\beta}, \Lambda'_{0\gamma}, \Lambda'_{0g}$ are bounded operators.

Recall that $L(t) = \int_0^t \exp(\gamma_0(s))ds, V = \alpha_0^T X + \int_0^1 \beta_0(s)Z(s)ds, R = e^V L(Y)$, $\psi(t, u, \alpha, \beta, \gamma, g) = \alpha^T X + \int_0^1 \beta(s)Z(s)ds + \gamma(t) + g(\Lambda(t, u, \alpha, \beta, \gamma, g))$ and $\psi'_{0\alpha}(Y, U)$ is the partial derivative of ψ with respect to α and $\psi'_{0\beta}(Y, U)$, $\psi'_{0\gamma}(Y, U)$, $\psi'_{0g}(Y, U)$ are the Fréchet derivatives of ψ with respect to β, γ, g at the true parameter $(\alpha_0, \beta_0, \gamma_0, g_0)$. By solving the initial problem in Lemma 1.1, we have

$$\begin{aligned}\psi'_{0\alpha}(Y,U) =& g'_0(\Lambda_0(Y,U))\Lambda'_{0\alpha}(Y,U) + X \\ =& \Big(g'_0(\Lambda_0(Y,U)) \exp\left(\alpha_0^T X + \int \beta_0(s)Z(s)ds\right) \\ & \times \exp(g_0(\Lambda_0(Y,U)))L(Y) + 1\Big) \\ =& \left(g'_0(\tilde{\Lambda}_0(R)) \exp(g_0(\tilde{\Lambda}_0(R)))R + 1\right) X \\ =:& \varepsilon_1(R)X, \end{aligned} \tag{S7.19}$$

where $\varepsilon_1(\cdot)$ is a deterministic function of R. Similarly, we also have

$$\begin{aligned} & \psi'_{0\beta}(Y,U)[\beta - \beta_0] \\ =& \left(g'_0(\tilde{\Lambda}_0(R)) \exp(g_0(\tilde{\Lambda}_0(R)))R + 1\right) \int Z(s)(\beta - \beta_0)(s)ds \\ =:& \varepsilon_1(R) \int Z(s)(\beta - \beta_0)(s)ds, \end{aligned} \tag{S7.20}$$

and

$$
\begin{aligned}
&\psi'_{0\gamma}(Y,U)[\gamma-\gamma_0]\\
=&g'_0(\Lambda_0(Y,U))\exp\left(\alpha_0^TX+\int\beta_0(s)Z(s)ds+g_0(\Lambda_0(Y,U))\right)\\
&\quad\times\int_0^Y\exp(\gamma_0(t))(\gamma-\gamma_0)(t)dt+\gamma(Y)-\gamma_0(Y)\\
=&g'_0(\tilde\Lambda_0(R))\exp(g_0(\tilde\Lambda_0(R)))\int_0^R(\gamma-\gamma_0)(L^{-1}(se^{-V}))ds\\
&\quad+(\gamma-\gamma_0)(L^{-1}(Re^{-V}))\\
=:&\varepsilon_2(R,V)[\gamma-\gamma_0],
\end{aligned}
\tag{S7.21}
$$

where $\varepsilon_2(\cdot)$ is a deterministic function of R,V and the second equality is derived by variable transformation $s=e^VL(t)$.

$$
\begin{aligned}
&\psi'_{0g}(Y,U)[g-g_0]\\
=\;&g'_0(\Lambda_0(Y,U))\exp(g(\Lambda_0(Y,U)))\int_0^{\Lambda_0(Y,U)}(g-g_0)(s)\exp(-g_0(s))ds\\
&\quad+(g-g_0)(\Lambda_0(Y,U))\\
=\;&g'_0(\tilde\Lambda_0(R))\exp(g_0(\tilde\Lambda_0(R)))\int_0^{\tilde\Lambda_0(R)}(g-g_0)(s)\exp(-g_0(s))ds\\
&\quad+(g-g_0)(\tilde\Lambda_0(R))\\
=\;&\varepsilon_3(R)[g-g_0],
\end{aligned}
\tag{S7.22}
$$

where $\varepsilon_3(\cdot)$ is a deterministic function of R. Plugging (S7.19), (S7.20), (S7.21) and (S7.22) into (S7.18) yields

$$\begin{aligned}
&P\left\{\Delta A^2\right\}\\
\gtrsim &P\Bigg\{\Delta\Big[\varepsilon_1(R)X^T(\alpha-\alpha_0)+\varepsilon_1\int Z(s)(\beta-\beta_0)(s)ds+\varepsilon_2(R,V)[\gamma-\gamma_0]\\
&+\varepsilon_3(R)[g-g_0]\Big]^2\Bigg\}+o\left(d^2(\theta,\theta_0)\right)\\
=&P\left\{\Delta\left[\varepsilon_1(R)\left\{X^T(\alpha-\alpha_0)+\int Z(s)(\beta-\beta_0)(s)ds\right\}\right]^2\right\}\\
&+2P\Bigg\{\Delta\left[\varepsilon_1(R)\left\{X^T(\alpha-\alpha_0)+\int Z(s)(\beta-\beta_0)(s)ds\right\}\right]\\
&\qquad\times\left[\varepsilon_2(R,V)[\gamma-\gamma_0]+\varepsilon_3(R)[g-g_0]\right]\Bigg\}\\
&+P\left\{\Delta\left[\varepsilon_2(R,V)[\gamma-\gamma_0]+\varepsilon_3(R)[g-g_0]\right]^2\right\}+o\left(d^2(\theta,\theta_0)\right)\\
\geq &P\left\{\Delta\left[\varepsilon_1(R)\left\{X^T(\alpha-\alpha_0)+\int Z(s)(\beta-\beta_0)(s)ds\right\}\right]^2\right\}\\
&-2\Bigg|P\Bigg\{\Delta\left[\varepsilon_1(R)\left\{X^T(\alpha-\alpha_0)+\int Z(s)(\beta-\beta_0)(s)ds\right\}\right]\\
&\qquad\times\left[\varepsilon_2(R,V)[\gamma-\gamma_0]+\varepsilon_3(R)[g-g_0]\right]\Bigg\}\Bigg|\\
&+P\left\{\Delta\left[\varepsilon_2(R,V)[\gamma-\gamma_0]+\varepsilon_3(R)[g-g_0]\right]^2\right\}+o\left(d^2(\theta,\theta_0)\right). \qquad \text{(S7.23)}
\end{aligned}$$

Noted that $P\{\Delta f(R,U)\}=P\{\int_0^Y f\left(L(t)e^V,U\right)d\Lambda_0(t,U)\}=P\{\int_0^R f(t,U)d\tilde{\Lambda}_0(t)\}$ holds for any measurable function f, together with Cauchy inequality we have

$$
\begin{aligned}
&\Bigg| P\Bigg\{ \Delta \left[\varepsilon_1(R) \left\{ X^T(\alpha-\alpha_0) + \int Z(s)(\beta-\beta_0)(s)ds \right\} \right] \\
&\qquad \times \left[\varepsilon_2(R,V)[\gamma-\gamma_0] + \varepsilon_3(R)[g-g_0] \right] \Bigg\} \Bigg|^2 \\
=& \Bigg(P\Bigg\{ \int_0^R \left[\varepsilon_1(R) \left\{ X^T(\alpha-\alpha_0) + \int Z(s)(\beta-\beta_0)(s)ds \right\} \right] \\
&\qquad \times \left[\varepsilon_2(t,V)[\gamma-\gamma_0] + \varepsilon_3(t)[g-g_0] \right] d\tilde{\Lambda}_0(t) \Bigg\} \Bigg)^2 \\
=& \Bigg(P\Bigg\{ \int_0^R \varepsilon_1(t) P \left[X^T(\alpha-\alpha_0) + \int Z(s)(\beta-\beta_0)(s)ds | R, V \right] \\
&\qquad \times \left[\varepsilon_2(t,V)[\gamma-\gamma_0] + \varepsilon_3(t)[g-g_0] \right] d\tilde{\Lambda}_0(t) \Bigg\} \Bigg)^2 \\
=& P\left\{ \int_0^R (\varepsilon_1(t))^2 \left(P\left[X^T(\alpha-\alpha_0) \int Z(s)(\beta-\beta_0)(s)ds | R, V \right] \right)^2 d\tilde{\Lambda}_0(t) \right\} \\
&\qquad \times P\left\{ \varepsilon_2(t,V)[\gamma-\gamma_0] + \varepsilon_3(t)[g-g_0])^2 d\tilde{\Lambda}_0(t) \right\}.
\end{aligned}
$$

Since Condition (A7) implies that

$$
\begin{aligned}
&\left(P\left[X^T(\alpha-\alpha_0) \int Z(s)(\beta-\beta_0)(s)ds | R, V \right] \right)^2 \\
\leq\ & (1-\eta_1) P\left[\left\{ X^T(\alpha-\alpha_0) \int Z(s)(\beta-\beta_0)(s)ds \right\}^2 \Big| R, V \right]
\end{aligned}
$$

holds for some $\eta_1 \in (0,1)$, hence,

$$\begin{aligned}
&\left| P\left\{ \Delta \left[\varepsilon_1(R) \left\{ X^T(\alpha - \alpha_0) + \int Z(s)(\beta - \beta_0)(s)ds \right\} \right] \right.\right. \\
&\qquad \left.\left. \times \left[\varepsilon_2(R,V)[\gamma - \gamma_0] + \varepsilon_3(R)[g - g_0] \right] \right\} \right|^2 \\
\leq\ & (1-\eta_1) P\left\{ \int_0^R \varepsilon_2(t,V)[\gamma - \gamma_0] + \varepsilon_3(t)[g - g_0])^2 d\tilde{\Lambda}_0(t) \right\} \\
& \times P\left\{ \int_0^R (\varepsilon_1(t))^2 P\left[\left(X^T(\alpha - \alpha_0) \int Z(s)(\beta - \beta_0)(s)ds \right)^2 \Big| R, V \right] d\tilde{\Lambda}_0(t) \right\} \\
=\ & (1-\eta_1) P\left\{ \int_0^R (\varepsilon_1(t))^2 \left(X^T(\alpha - \alpha_0) \int Z(s)(\beta - \beta_0)(s)ds \right)^2 d\tilde{\Lambda}_0(t) \right\} \\
& \times P\left\{ \int_0^R \varepsilon_2(t,V)[\gamma - \gamma_0] + \varepsilon_3(t)[g - g_0])^2 d\tilde{\Lambda}_0(t) \right\} \\
=\ & (1-\eta_1) P\left\{ \Delta \left[\varepsilon_1(R) \left\{ X^T(\alpha - \alpha_0) + \int Z(s)(\beta - \beta_0)(s)ds \right\} \right]^2 \right\} \\
& \times P\left\{ \Delta \left(\varepsilon_2(R,V)[\gamma - \gamma_0] + \varepsilon_3(R)[g - g_0] \right)^2 \right\}
\end{aligned}$$

Using the elementary inequality $2ab \leq a^2 + b^2$, we obtain

$$P\left\{ \Delta A^2 \right\} \gtrsim A_1 + A_2,$$

where

$$\begin{aligned}
A_1 &= P\left\{ \Delta \left[\varepsilon_1(R) \left\{ X^T(\alpha - \alpha_0) + \int Z(s)(\beta - \beta_0)(s)ds \right\} \right]^2 \right\}, \\
A_2 &= P\left\{ \Delta \left(\varepsilon_2(R,V)[\gamma - \gamma_0] + \varepsilon_3(R)[g - g_0] \right)^2 \right\}.
\end{aligned}$$

For A_1, we have

$$
\begin{aligned}
A_1 =&P\left\{\int_0^Y \left(\varepsilon_1(e^V L(t))\left(X^T(\alpha-\alpha_0)+\int Z(s)(\beta-\beta_0)(s)ds\right)\right)^2 d\Lambda_0(t)\right\}\\
=&P\left\{\int_0^\tau E[I(Y\geq t)|U]\left(\varepsilon_1(e^V L(t))\left(X^T(\alpha-\alpha_0)+\int Z(s)(\beta-\beta_0)(s)ds\right)\right)^2 d\Lambda_0(t)\right\}\\
\geq&\delta_0 P\bigg\{\int_0^\tau \left(\varepsilon_1(e^V L(t))\left(X^T(\alpha-\alpha_0)+\int Z(s)(\beta-\beta_0)(s)ds\right)\right)^2\\
&\qquad\times\exp\left(V+\gamma_0(t)+g_0(\Lambda_0(t,U))\right)dt\bigg\},
\end{aligned}
$$

where the last inequality holds under Condition (A4). Take the variable transformation $u=e^V L(t)$ and denote $c=\min_{x\in\mathcal{X},z\in\mathcal{Z}}\exp(\alpha_0^T X+\int\beta_0(s)z(s)ds)$, then

$$
\begin{aligned}
A_1 &\geq \delta_0 P\left\{\int_0^{e^V L(\tau)}\left(\varepsilon_1(u)X^T(\alpha-\alpha_0)\right)^2\exp\left(g_0(\tilde\Lambda_0(u))\right)dt\right\}\\
&\geq \delta_0 P\left(X^T(\alpha-\alpha_0)+\int Z(s)(\beta-\beta_0)(s)ds\right)^2\\
&\qquad\int_0^{e^c L(\tau)}\left(\varepsilon_1(u)\right)^2\exp\left(g_0(\tilde\Lambda_0(u))\right)dt.
\end{aligned}
$$

Since $\exp\left(g_0(\tilde\Lambda_0(u))\right)$ is strictly positive and $\varepsilon_1(t)$ satisfies $\int_0^t\exp\left(g_0(\tilde\Lambda_0(s))\right)\varepsilon_1(s)ds = t\exp\left(g_0(\tilde\Lambda_0(t))\right)$, thus $\varepsilon_1(t)$ cannot be a constant zero. Hence,

$$
\int_0^{e^c L(\tau)}\left(\varepsilon_1(u)\right)^2\exp\left(g_0(\tilde\Lambda_0(u))\right)dt
$$

is bounded away from 0 below and $A_1 \gtrsim P\left(X^T(\alpha-\alpha_0)+\int Z(s)(\beta-\beta_0)(s)ds\right)^2$. For A_2, we have

$$
\begin{aligned}
A_2 &= P\left\{\Delta\left[\varepsilon_2(R,V)+\varepsilon_3(R)[g-g_0]\right]^2\right\} \\
&= P\left\{\Delta(\epsilon_2(R,V)[\gamma-\gamma_0])^2\right\}+P\left\{\Delta(\epsilon_3(R)[g-g_0])^2\right\} \\
&\quad +2P\left\{\Delta\epsilon_2(R,V)[\gamma-\gamma_0]\epsilon_3(R)[g-g_0]\right\} \\
&\geq P\left\{\Delta(\epsilon_2(R,V)[\gamma-\gamma_0])^2\right\}+P\left\{\Delta(\epsilon_3(R)[g-g_0])^2\right\} \\
&\quad -2\left|P\left\{\Delta\epsilon_2(R,V)[\gamma-\gamma_0]\epsilon_3(R)[g-g_0]\right\}\right| \\
&\geq P\left\{\Delta(\epsilon_2(R,V)[\gamma-\gamma_0])^2\right\}+P\left\{\Delta(\epsilon_3(R)[g-g_0])^2\right\} \\
&\quad -2\eta_3^{1/2}P\{\Delta\}\left(P\left\{\Delta(\epsilon_2(R,V)[\gamma-\gamma_0])^2\right\}\right)^{1/2}\left(P\left\{\Delta(\epsilon_3(R)[g-g_0])^2\right\}\right)^{1/2} \\
&\gtrsim A_3+A_4,
\end{aligned}
$$

where $A_3 = P\{\Delta(\epsilon_2(R,V)[\gamma-\gamma_0])^2\}$, $A_4 = P\{\Delta(\epsilon_3(R)[g-g_0])^2\}$. The fourth inequality holds under Condition (A8), and the last inequality is obtained by $2ab \leq a^2+b^2$. Then, under Condition (A1)-(A6),

$$
\begin{aligned}
A_3 &= P\left\{\int_0^Y (\psi'_{0\gamma}(t,U)[\gamma-\gamma_0])^2 d\Lambda_0(t,U)\right\} \\
&\geq P\left\{\int_0^\tau P(Y\geq t|U)(\psi'_{0\gamma}(t,U)[\gamma-\gamma_0])^2 d\Lambda_0(t,U)\right\} \\
&\geq \delta_0 P\left\{\int_0^\tau (\psi'_{0\gamma}(t,U)[\gamma-\gamma_0])^2 d\Lambda_0(t,U)\right\} \\
&\gtrsim \|\gamma-\gamma_0\|_2^2,
\end{aligned}
$$

where the last equality is obtained by Lemma 1.6 and the fact that $\gamma-\gamma_0 \in \Gamma^{p_2}$, and

$$
\begin{aligned}
A_4 &= P\left\{\int_0^Y (\psi'_{0g}(Y,U)[g-g_0])^2 d\Lambda_0(t,U)\right\} \\
&\geq P\left\{\int_0^\tau P(Y\geq t|U)(\psi'_{0g}(Y,U)[g-g_0])^2 d\Lambda_0(t,U)\right\} \\
&\geq \delta_0 P\left\{\int_0^\tau (\psi'_{0g}(Y,U)[g-g_0])^2 d\Lambda_0(t,U)\right\} \\
&\gtrsim \|g-g_0\|^2.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& Pl_0(\theta_0,W)-Pl(\theta,W) \\
=\ & \frac{1}{2}P\{\Delta A^2\}+o(P\{A^2\}) \\
\gtrsim\ & P\left(X^T(\alpha-\alpha_0)+\int Z(s)(\beta-\beta_0)(s)ds\right)^2+\|\gamma-\gamma_0\|_2^2+\|g-g_0\|_2^2 \\
\gtrsim\ & d(\theta,\theta_0)^2,
\end{aligned}
$$

which implies

$$
\inf_{d(\theta,\theta_0)\geq\epsilon,\theta\in\Theta_n} Pl_0(\theta_0,W)-Pl(\theta,W)\gtrsim \epsilon^2.
$$

Thus the condition (C1) in theorem 1 of Shen and Wong (1994) holds with $\alpha=1$ in their notation.

Next we examine the condition (C2) of Shen and Wong (1994). By direct calculation,

$$
\begin{aligned}
& \left(l(\theta;W)-l(\theta_0;W)\right)^2 \\
= & \Big\{\Delta\big[(\alpha-\alpha_0)^T X+\int(\beta-\beta_0)(s)Z(s)ds+(\gamma-\gamma_0)(Y)+g(\Lambda(Y,U,\theta)) \\
& -g_0(\Lambda_0(Y,U))\big]-\int_0^Y\Big[\exp(\alpha^T X+\int\beta(s)Z(s)ds+\gamma(t)+g(\Lambda(Y,U,\theta))) \\
& \quad -\exp(\alpha_0^T X+\int\beta_0(s)Z(s)ds+\gamma_0(t)+g_0(\Lambda_0(Y,U)))\Big]dt\Big\}^2.
\end{aligned}
$$

Therefore,

$$
\begin{aligned}
& P\left(l(\theta;W)-l(\theta_0;W)\right)^2 \\
\lesssim & P\left(X^T(\alpha-\alpha_0)+\int Z(s)(\beta-\beta_0)(s)ds\right)^2+P\{\Delta(\gamma(Y)-\gamma_0(Y))^2\} \\
& +P\{\Delta(g(\Lambda(Y,U,\theta))-g_0(\Lambda_0(Y,U)))^2\} \\
& +P\{\int_0^Y(\exp(\alpha^T X+\int\beta(s)Z(s)ds+\gamma(t)+g(\Lambda(Y,U,\theta))) \\
& \quad -\exp(\alpha_0^T X+\int\beta_0(s)Z(s)ds+\gamma_0(t)+g_0(\Lambda_0(Y,U))))^2dt\}.
\end{aligned}
$$

For the second term $P\{\Delta(\gamma(Y)-\gamma_0(Y))^2\}$, we have

$$
\begin{aligned}
& P\{\Delta(\gamma(Y)-\gamma_0(Y))^2\} \\
= & P\{\int_0^Y(\gamma(t)-\gamma_0(t))^2 d\Lambda_0(t)\} \\
\le & \int_0^\tau P\{\exp(\alpha_0^T X+\int\beta_0(s)Z(s)ds+\gamma_0(t)+g(\Lambda_0(Y,U)))((\gamma-\gamma_0)(t))^2dt\} \\
\lesssim & \|\gamma-\gamma_0\|_2^2,
\end{aligned}
$$

where the last inequality holds since $\exp(\alpha_0^T X+\int\beta_0(s)Z(s)ds+\gamma_0(t)+g(\Lambda_0(Y,U)))$ is bounded above under conditions (A1)-(A5). For the third term $P\{\Delta(g(\Lambda(Y,U,\theta)) - g_0(\Lambda_0(Y,U)))^2\}$, we have

$$
\begin{aligned}
& P\{\Delta(g(\Lambda(Y,U,\theta)) - g_0(\Lambda_0(Y,U)))^2\} \\
= \;& P\left\{\int_0^Y (g(\Lambda(t,U,\theta)) - g_0(\Lambda_0(Y,U)))^2 d\Lambda_0(t)\right\} \\
\leq \;& P\left\{\int_0^\tau (g(\Lambda(t,U,\theta)) - g_0(\Lambda_0(Y,U)))^2 d\Lambda_0(t)\right\} \\
= \;& \|\zeta(\cdot,\alpha,\beta,\gamma) - \zeta_0(\cdot,\alpha_0,\beta_0,\gamma_0)\|.
\end{aligned}
$$

For the last term, mean value theorem yields

$$
\begin{aligned}
& P\Big\{\int_0^Y (\exp(\alpha^T X + \int \beta(s)Z(s)ds + \gamma(t) + g(\Lambda(Y,U,\theta))) \\
& \quad - \exp(\alpha_0^T X + \int \beta_0(s)Z(s)ds + \gamma_0(t) + g_0(\Lambda_0(Y,U))))^2 dt\Big\} \\
= \;& P\Big\{\int_0^Y \exp(2\tilde{\psi}(t,U))\Big((\alpha-\alpha_0)^T X + \int (\beta-\beta_0)(s)Z(s)ds + \gamma(t) - \gamma_0(t) \\
& + g(\Lambda(Y,U,\theta)) - g_0(\Lambda_0(Y,U))\Big)^2 dt\Big\},
\end{aligned}
$$

where $\tilde{\psi}(t,U)$ is some point between $\alpha_0^T X + \int \beta_0(s)Z(s)ds + \gamma_0(t) + g_0(\Lambda_0(Y,U))$ and $\alpha^T X + \int \beta(s)Z(s)ds + \gamma(t) + g(\Lambda(Y,U,\theta))$. Thus, under Conditions (A5) and (A6), $\tilde{\psi}(t,U)$ is bounded (or growing with n slowly enough so it can be treated as bounded based on the same argument of Shen and Wong on page 591). It follows that the last term is bounded by a constant multiple of $P\left(X^T(\alpha-\alpha_0) + \int Z(s)(\beta-\beta_0)(s)ds\right)^2 + \|\gamma-\gamma_0\|_2^2 + \|\zeta(\cdot,\alpha,\beta,\gamma) - \zeta_0(\cdot,\alpha_0,\beta_0,\gamma_0)\|$.

Consequently,

$$\begin{aligned}
P\left(l(\theta;W)-l(\theta_0;W)\right)^2 &\leq P\left(X^T(\alpha-\alpha_0)+\int Z(s)(\beta-\beta_0)(s)ds\right)^2 \\
&\quad +\|\gamma-\gamma_0\|_2^2+\|\zeta(\cdot,\alpha,\beta,\gamma)-\zeta_0(\cdot,\alpha_0,\beta_0,\gamma_0)\| \\
&\lesssim d^2(\theta,\theta_0),
\end{aligned}$$

which implies that

$$\begin{aligned}
& \sup_{d(\theta_0,\theta)\leq\epsilon,\theta\in\Theta_n} Var\left(l(\theta;W)-l(\theta_0;W)\right) \\
\leq & \sup_{d(\theta_0,\theta)\leq\epsilon,\theta\in\Theta_n} P\left(l(\theta;W)-l(\theta_0;W)\right)^2 \\
\lesssim & \ \epsilon^2.
\end{aligned}$$

Thus, condition (C2) of Shen and Wong (1994) holds with $\beta = 1$ in their notation.

Finally we check condition (C3) of Shen and Wong (1994). The result of lemma 1.3 shows that

$$\begin{aligned}
H(\varepsilon,\mathcal{F}_n,\|\cdot\|_\infty) &= \log N(\varepsilon,\mathcal{F}_n,\|\cdot\|_\infty) \\
&\lesssim (c_1q_{n1}+c_2q_{n2}+c_3q_{n3}+d)\log(\frac{1}{\varepsilon}) \\
&\lesssim (n^{\nu_1}+n^{\nu_2}+n^{\nu_3})\log(\frac{1}{\varepsilon}) \\
&\lesssim n^{\max\{\nu_1,\nu_2,\nu_3\}}\log(\frac{1}{\varepsilon}).
\end{aligned}$$

Thus, condition (C3) of Shen and Wong (1994) holds with $r_0 = \frac{1}{2}\max\{\nu_1,\nu_2,\nu_3\}$,

$r = 0^+$ in their notation. It follows that the constant τ in Theorem 1 of Shen and Wong (1994) is $\frac{1-\max\{\nu_1,\nu_2,\nu_3\}}{2} - \frac{\log\log n}{2\log n}$. Using the same argument as in the proof of Condition (C2), we have $K(\theta_{0n}, \theta_0) \lesssim O(d^2(\theta_{0n}, \theta_0))$. By Lemma 1.2, $d(\theta_{0n}, \theta_0) = O(n^{-\min\{p_1\nu_1, p_2\nu_2, p_3\nu_3\}})$, therefore, By theorem 1 of Shen and Wong (1994), we have

$$\begin{aligned} d(\hat{\theta}_n, \theta_0) &= O_P(\max\{n^{-\frac{1-\max\{\nu_1,\nu_2,\nu_3\}}{2}}, n^{-\min\{p_1\nu_1,p_2\nu_2,p_3\nu_3\}}\}) \\ &= O_p(n^{-\min\{p_1\nu_1,p_2\nu_2,p_3\nu_3,\frac{1-\max\{\nu_1,\nu_2,\nu_3\}}{2}\}}) =: O_p(n^{-c}). \end{aligned}$$

It follows that $\|\hat{\gamma}-\gamma\|_2 = O_p(n^{-c})$, $\|\hat{\zeta}(\cdot, \hat{\alpha}, \hat{\beta}, \hat{\gamma})-\zeta_0(\cdot, \alpha_0, \beta_0, \gamma_0)\|_2 = O_p(n^{-c})$ and

$$P\left\{\left(X^T(\alpha-\alpha_0) + \int Z(s)(\beta-\beta_0)(s)ds\right)^2\right\} = O_p(n^{-c}). \qquad \text{(S7.24)}$$

From the proof of Theorem 2, we have

$$\begin{aligned} & P\left\{\Delta\varepsilon_1^2(R)\left(X^T(\alpha-\alpha_0) + \int Z(s)(\beta-\beta_0)(s)ds\right)^2\right\} \\ = \ & P\left[\Delta\left(\varepsilon_1(R)(X - \int Z(s)\mathbf{h}_1^*(s)ds) - \varepsilon_2(R,V)[\mathbf{h}_2^*] - \varepsilon_3(R)[\mathbf{h}_3^*]\right)^T(\alpha-\alpha_0)\right]^2 \\ & + P\left[\Delta\left(\varepsilon_1(R)\int Z(s)\mathbf{h}_1^*(s)ds + \varepsilon_2(R,V)[\mathbf{h}_2^*] + \varepsilon_3(R)[\mathbf{h}_3^*]\right)^T(\alpha-\alpha_0) \right. \\ & \qquad \left. + \int Z(s)(\beta-\beta_0)(s)ds\right]^2, \end{aligned}$$

where $(\mathbf{h}_1^*, \mathbf{h}_2^*, \mathbf{h}_3^*)$ are defined in (13). When $I(\alpha_0)$ is non-singular, (S7.24) leads to $|\alpha-\alpha_0| = O_p(n^{-c})$. This in turn implies $\|\beta-\beta_0\|_C = O_p(n^{-c})$.

S8 Proof of Theorem 4

We prove the theorem by verifying assumptions (A1)-(A6) of Theorem 3 in Tang et al. (2022). By Theorem 3 we know that assumption (A1) holds with

$$\xi = \min\{p_1\nu_1, p_2\nu_2, p_3\nu_3, \frac{1-\max\{\nu_1,\nu_2,\nu_3\}}{2}\}.$$

Assumption (A2) holds since the score functions have zero mean. For assumption (A3), following the line of proving Theorem 2, there exists $(\mathbf{h}_1^*, \mathbf{h}_2^*, \mathbf{h}_3^*)$ such that

$$\begin{aligned}
&\mathbb{E}\left\{\left(\dot{l}_{\alpha_0} - \dot{l}_{\beta_0}\mathbf{h}_1^* - \dot{l}_{\gamma_0}\mathbf{h}_2^* - \dot{l}_{g_0}\mathbf{h}_3^*\right)\dot{l}_{\beta_0}h\right\} = 0 \quad \text{for all} \quad h \in \overline{\mathbb{T}}_{\beta_0},\\
&\mathbb{E}\left\{\left(\dot{l}_{\alpha_0} - \dot{l}_{\beta_0}\mathbf{h}_1^* - \dot{l}_{\gamma_0}\mathbf{h}_2^* - \dot{l}_{g_0}\mathbf{h}_3^*\right)\dot{l}_{\gamma_0}h\right\} = 0 \quad \text{for all} \quad h \in \overline{\mathbb{T}}_{\gamma_0},\\
&\mathbb{E}\left\{\left(\dot{l}_{\alpha_0} - \dot{l}_{\beta_0}\mathbf{h}_1^* - \dot{l}_{\gamma_0}\mathbf{h}_2^* - \dot{l}_{g_0}\mathbf{h}_3^*\right)\dot{l}_{g_0}h\right\} = 0 \quad \text{for all} \quad h \in \overline{\mathbb{T}}_{g_0}. \quad \text{(S8.25)}
\end{aligned}$$

Take $\tilde{\mathbf{h}}_3^* = \mathbf{h}_3^*(\Lambda_0(\cdot)) + g_0'(\Lambda_0(\cdot))\Lambda_{0g}'(\cdot)[\mathbf{h}_3^*]$, then for any $h \in \mathbb{T}_1$,

$$\begin{aligned}
& P\left\{l''_{\alpha\beta}(\theta_0)[h] - l''_{\beta\beta}(\theta_0)[\mathbf{h}_1^*, h] - l''_{\beta\gamma}(\theta_0)[\mathbf{h}_2^*, h] - l''_{\beta\zeta}(\theta_0)[\tilde{\mathbf{h}}_3^*, h]\right\}\\
= \; & P\left\{\left(\dot{l}_{\alpha_0} - \dot{l}_{\beta_0}\mathbf{h}_1^* - \dot{l}_{\gamma_0}\mathbf{h}_2^* - \dot{l}_{g_0}\mathbf{h}_3^*\right)\dot{l}_{\beta_0}h\right\} = 0,
\end{aligned}$$

for any $h \in \mathbb{T}_2$,

$$\begin{aligned}
& P\left\{l''_{\alpha\gamma}(\theta_0)[h] - l''_{\beta\gamma}(\theta_0)[\mathbf{h}_1^*, h] - l''_{\gamma\gamma}(\theta_0)[\mathbf{h}_2^*, h] - l''_{\zeta\gamma}(\theta_0)[\tilde{\mathbf{h}}_3^*, h]\right\}\\
= \; & P\left\{\left(\dot{l}_{\alpha_0} - \dot{l}_{\beta_0}\mathbf{h}_1^* - \dot{l}_{\gamma_0}\mathbf{h}_2^* - \dot{l}_{g_0}\mathbf{h}_3^*\right)\dot{l}_{\gamma_0}h\right\} = 0,
\end{aligned}$$

for any $h \in \mathbb{T}_3$,

$$\begin{aligned} & P\left\{l''_{\alpha\gamma}(\theta_0)[h] - l''_{\beta\gamma}(\theta_0)[\mathbf{h}_1^*, h] - l''_{\gamma\gamma}(\theta_0)[\mathbf{h}_2^*, h] - l''_{\zeta\gamma}(\theta_0)[\tilde{\mathbf{h}}_3^*, h]\right\} \\ = \; & P\left\{\left(\dot{l}_{\alpha_0} - \dot{l}_{\beta_0}\mathbf{h}_1^* - \dot{l}_{\gamma_0}\mathbf{h}_2^* - \dot{l}_{g_0}\mathbf{h}_3^*\right)\dot{l}_{\gamma_0} h\right\} = 0. \end{aligned}$$

The matrix A in assumption (A3) is given by

$$\begin{aligned} A &= P\left\{(l'_\alpha(\theta_0, W) - l'_\beta(\theta_0, W) - l'_\gamma(\theta_0, W) - l'_\zeta(\theta_0, W))^{\otimes 2}\right\} \\ &= I(\alpha_0), \end{aligned}$$

which is nonsingular under the assumption of Theorem 4. Hence, assumption (A3) holds.

To verify assumption (A4), we need to show that $\mathbb{P}_n\left\{l'_\alpha(\hat{\theta}_n, W)\right\} = o_p(n^{-1/2})$, $\mathbb{P}_n\left\{l'_\beta(\hat{\theta}_n, W)[\mathbf{h}_1^*]\right\} = o_p(n^{-1/2})$, $\mathbb{P}_n\left\{l'_\gamma(\hat{\theta}_n, W)[\mathbf{h}_2^*]\right\} = o_p(n^{-1/2})$ and $\mathbb{P}_n\left\{l'_\zeta(\hat{\theta}_n, W)[\tilde{\mathbf{h}}_3^*]\right\} = o_p(n^{-1/2})$. Since $\hat{\alpha}_n$ satisfies $\mathbb{P}_n\left\{l'_\alpha(\hat{\theta}_n, W)\right\} = 0 = o_p(n^{-1/2})$, the first equality naturally holds. Next we show $\mathbb{P}_n\left\{l'_\beta(\hat{\theta}_n, W)[h^*_{1j}]\right\} = o_p(n^{-1/2})$ for each $j \in \{1, 2, \ldots, d\}$. By Lemma 2, there exists $h_{1j,n} \in \mathcal{F}_n^2$, such that $\|h^*_{1j} - h_{1j,n}\|_\infty = O(n^{-2\nu_1})$. Based on the fact that θ_n maximizes the loglikelihood on the sieve space, we have $\mathbb{P}_n\left\{l'_\beta(\hat{\theta}_n, W)[h_{1j,n}]\right\} = 0$. Additionally, $Pl'_\beta(\theta_0, W)[h^*_{1j}] = 0$ holds automatically, hence, it suffices to study $\mathbb{P}_n l'_\beta(\hat{\theta}_n, W)[h^*_{1j} - h_{1j,n}]$, which can be further decomposed as

$$\mathbb{P}_n l'_\beta(\hat{\theta}_n, W)[h^*_{1j} - h_{1j,n}] = B_{1n} + B_{2n},$$

where

$$B_{1n} = (\mathbb{P}_n - P)\{l'_\beta(\hat\theta_n, W)[h^*_{1j} - h_{1j,n}]\},$$

$$B_{2n} = P\{l'_\beta(\hat\theta_n, W)[h^*_{1j} - h_{1j,n}]\} - P\{l'_\beta(\theta_0, W)[h^*_{1j} - h_{1j,n}]\}.$$

We first show that $B_{1n} = o_p(n^{-1/2})$. By Lemma 4, the ε-bracketing number associated with $\|\cdot\|_\infty$ norm for the class $\mathcal{F}^\beta_{n,j}(\eta)$ is bounded by $(\eta/\varepsilon)^{c_1 q_{n1} + c_2 q_{n2} + c_3 q_{n3} + d}$, thus the bracketing integral satisfies

$$\begin{aligned} J_{[\,]}(\eta, \mathcal{F}^\beta_{n,j}(\eta), L_2(P)) &= \int_0^\eta \sqrt{1 + \log N_{[\,]}(\varepsilon, \mathcal{F}^\beta_{n,j}(\eta), L_2(P))} d\varepsilon \\ &\le \int_0^\eta \sqrt{1 + \log N_{[\,]}(\varepsilon, \mathcal{F}^\beta_{n,j}(\eta), \|\cdot\|_\infty)} d\varepsilon \\ &\lesssim (c_1 q_{n1} + c_2 q_{n2} + c_3 q_{n3})^{1/2} \eta. \end{aligned}$$

Pick $\eta_n = O(n^{-\min\{2\nu_1, p_2\nu_2, p_3\nu_3, (1-\max\{\nu_1,\nu_2,\nu_3\})/2\}})$, then $\|h^*_{1j} - h_{1j,n}\|_\infty = O(n^{-2\nu_1}) \le \eta_n$ and $d(\theta_n, \theta_0) = O_p(n^{-c}) \le \eta_n$, hence, $l'_\beta(\hat\theta_n, W)[h^*_{1j} - h_{1j,n}] \in \mathcal{F}^\beta_{n,j}(\eta_n)$. Using the boundedness of Z, g' and Λ'_β, we have for any $\theta \in \Theta_n$,

$$\begin{aligned} & P\{l'_\beta(\theta, W)[h^*_{1j} - h_{1j,n}]\}^2 \\ = \; & P\Big\{\Delta\Big(\int_0^1 (h^*_{1j} - h_{1j,n})(s) Z(s) ds + g'(\Lambda(Y,U,\theta))\Lambda'_\beta(Y,U,\theta)[h^*_{1j} - h_{1j,n}]\Big) \\ & \quad -\Lambda'_\beta(Y,U,\theta)[h^*_{1j} - h_{1j,n}]\Big\}^2 \\ \lesssim \; & \|h^*_{1j} - h_{1j,n}\|^2_\infty, \end{aligned}$$

and $\sup_{d(\theta,\theta_0)\le\eta_n, \|h^*_{1j} - h_{1j,n}\|_\infty \le \eta_n} |l'_\beta(\theta, W)[h^*_{1j} - h_{1j,n}]|$ is bounded by some constant $M < \infty$. Thus, the prerequisites of Lemma 3.4.2 of Van der Vaart and Wellner (1996) are satisfied. By Lemma 3.4.2 of Van der Vaart and Wellner (1996),

$$\begin{aligned} E\|\mathbb{G}_n\|_{\mathcal{F}^{\beta}_{n,j}(\eta_n)} &\lesssim J_{[\,]}(\eta_n, \mathcal{F}^{\beta}_{n,j}(\eta_n), L_2(P))\left(1+\frac{J_{[\,]}(\eta, \mathcal{F}^{\beta}_{n,j}(\eta_n), L_2(P))}{\eta_n^2\sqrt{n}}\right) \\ &\lesssim (c_1 q_{n1}+c_2 q_{n2}+c_3 q_{n3})^{1/2}\eta_n + (c_1 q_{n1}+c_2 q_{n2}+c_3 q_{n3})\sqrt{n} \\ &\lesssim O(n^{\frac{\max\{\nu_1,\nu_2,\nu_3\}}{2}})O(n^{-\min\{2\nu_1, p_2\nu_2, p_3\nu_3, (1-\max\{\nu_1,\nu_2,\nu_3\})/2\}}) \\ &\quad + O(n^{\max\{\nu_1,\nu_2,\nu_3\}-1/2}) \\ &= o_p(1), \end{aligned}$$

where $\mathbb{G}_n = \sqrt{n}(\mathbb{P}_n - \mathbb{P})$ and the last equality holds because $\max\{\nu_1, \nu_2, \nu_3\} < 1/2$ and $\nu_{\max} < 4\nu_{\min}$. Then by Markov's inequality, we have

$$B_{1n} = (\mathbb{P}_n - \mathbb{P})\{l'_{\beta}(\hat{\theta}_n, W)[h^*_{1j} - h_{1j,n}]\} = o_p(n^{-1/2}).$$

Next we consider B_{2n}. Denote $\bar{h} = h^*_{1j} - h_{1j,n}$. By adding and subtracting some terms, we have

$$\begin{aligned} &\|\hat{\zeta}'_{n,\beta}(\cdot, \hat{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_n, \hat{g}_n)[\bar{h}] - \zeta'_{0\beta}(\cdot, \alpha_0, \beta_0, \gamma_0, g_0)[\bar{h}]\|_2 \\ =\; &\|\hat{g}'_n(\Lambda(\cdot, \hat{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_n, \hat{g}_n))\Lambda'_{\beta}(\cdot, \hat{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_n, \hat{g}_n)[\bar{h}] - g'_0(\Lambda_0(\cdot))\Lambda'_{0\beta}(\cdot)[\bar{h}]\|_2 \\ \le\; &J_1 + J_2 + J_3 + J_4 + J_5, \end{aligned}$$

where

$$
\begin{aligned}
J_1 &= \|\hat{g}_n'(\Lambda(\cdot,\hat{\alpha}_n,\hat{\beta}_n,\hat{\gamma}_n,\hat{g}_n))\Lambda_\beta'(\cdot,\hat{\alpha}_n,\hat{\beta}_n,\hat{\gamma}_n,\hat{g}_n)[\bar{h}] \\
&\qquad -g_0'(\Lambda_0(\cdot))\Lambda_\beta'(\cdot,\hat{\alpha}_n,\hat{\beta}_n,\hat{\gamma}_n,\hat{g}_n)[\bar{h}]\|_2, \\
J_2 &= \|g_0'(\Lambda_0(\cdot))\Lambda_\beta'(\cdot,\hat{\alpha}_n,\hat{\beta}_n,\hat{\gamma}_n,\hat{g}_n)[\bar{h}] - g_0'(\Lambda_0(\cdot))\Lambda_\beta'(\cdot,\alpha_0,\hat{\beta}_n,\hat{\gamma}_n,\hat{g}_n)[\bar{h}]\|_2, \\
J_3 &= \|g_0'(\Lambda_0(\cdot))\Lambda_\beta'(\cdot,\alpha_0,\hat{\beta}_n,\hat{\gamma}_n,\hat{g}_n)[\bar{h}] - g_0'(\Lambda_0(\cdot))\Lambda_\beta'(\cdot,\alpha_0,\beta_0,\hat{\gamma}_n,\hat{g}_n)[\bar{h}]\|_2, \\
J_4 &= \|g_0'(\Lambda_0(\cdot))\Lambda_\beta'(\cdot,\alpha_0,\beta_0,\hat{\gamma}_n,\hat{g}_n)[\bar{h}] - g_0'(\Lambda_0(\cdot))\Lambda_\beta'(\cdot,\alpha_0,\beta_0,\gamma_0,\hat{g}_n)[\bar{h}]\|_2, \\
J_5 &= \|g_0'(\Lambda_0(\cdot))\Lambda_\beta'(\cdot,\alpha_0,\beta_0,\gamma_0,\hat{g}_n)[\bar{h}] - g_0'(\Lambda_0(\cdot))\Lambda_{0\beta}'(\cdot)[\bar{h}]\|_2.
\end{aligned}
$$

We first consider J_1. Based on the boundedness of $\hat{g}_n, \hat{g}_n'$, we have $\|\Lambda_\beta'(\cdot,\hat{\alpha}_n,\hat{\beta}_n,\hat{\gamma}_n,\hat{g}_n)[\bar{h}]\|_\infty \lesssim \|\bar{h}\|_\infty$ and it follows that

$$
\begin{aligned}
J_1 &\leq \|\hat{g}_n'(\Lambda(\cdot,\hat{\alpha}_n,\hat{\beta}_n,\hat{\gamma}_n,\hat{g}_n)) - g_0'(\Lambda_0(\cdot))\|_2 \cdot \|\Lambda_\beta'(\cdot,\hat{\beta}_n,\hat{\gamma}_n,\hat{g}_n)[\bar{h}]\|_\infty \\
&\lesssim \|\hat{g}_n'(\Lambda(\cdot,\hat{\alpha}_n,\hat{\beta}_n,\hat{\gamma}_n,\hat{g}_n)) - g_0'(\Lambda_0(\cdot))\|_2 \cdot \|\bar{h}\|_\infty \\
&= O_p(n^{-\min\{p_1\nu_1, p_2\nu_2, (p_3-1)\nu_3, (1-\max\{\nu_1,\nu_2,\nu_3\})/2\}}) \cdot O(n^{-2\nu_1}) \\
&= O_p(n^{-\min\{(p_1+2)\nu_1, p_2\nu_2+2\nu_1, (p_3-1)\nu_3+2\nu_1, (1-\max\{\nu_1,\nu_2,\nu_3\})/2+2\nu_1\}}),
\end{aligned}
$$

where the third equality holds based on the same argument of Ding and

Nan (2011) on page 3058. Next we consider J_2. By mean value theorem,

$$\begin{aligned} J_2 &= \|g_0'(\Lambda_0(\cdot))(\Lambda''_{\alpha\beta}(\cdot, \tilde{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_n, \hat{g}_n)[\bar{h}])^T(\hat{\alpha}_n - \alpha_0)\|_2 \\ &\lesssim \|\Lambda''_{\alpha\beta}(\cdot, \tilde{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_n, \hat{g}_n)[\bar{h}]\|_2 \|\hat{\alpha}_n - \alpha_0\| \\ &\lesssim \|\bar{h}\|_\infty \cdot \|\hat{\alpha}_n - \alpha_0\| \\ &= O_p(n^{-c+2\nu_1}), \end{aligned}$$

where $\tilde{\alpha}_n$ is a point between α_0 and $\hat{\alpha}_n$ and the third inequality can be derived by solving the initial problem in Lemma 1 and similar argument as in Lemma 6 based on the boundedness of $\hat{\beta}_n, \hat{\gamma}_n, \hat{g}_n, \hat{g}_n', \hat{g}_n''$. As for J_3, J_4 and J_5, similar argument leads to $J_3 = O_p(n^{-c+2\nu_1})$, $J_4 = O_p(n^{-c+2\nu_1})$ and $J_5 = O_p(n^{-c+2\nu_1})$. Thus, we have

$$\begin{aligned} &\|\hat{\zeta}'_{n,\beta}(\cdot, \hat{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_n, \hat{g}_n)[\bar{h}] - \zeta'_{0\beta}(\cdot, \alpha_0, \beta_0, \gamma_0, g_0)[\bar{h}]\|_2 \\ \lesssim\ & O_p(n^{-\min\{(p_1+2)\nu_1, p_2\nu_2+2\nu_1, (p_3-1)\nu_3+2\nu_1, (1-\max\{\nu_1,\nu_2,\nu_3\})/2+2\nu_1\}}). \end{aligned}$$

Now we turn to B_{2n}. Direct calculation yields that

$$\begin{aligned} B_{2n}^2 = \Bigg(& P\Bigg\{\Delta\left(\hat{\zeta}'_{n,\beta}(Y, U, \hat{\alpha}_n, \hat{\beta}_n, \hat{\gamma}_n)[\bar{h}] - \zeta'_{0\beta}(Y, U, \alpha_0, \beta_0, \gamma_0)[\bar{h}]\right) \\ & - \int_0^Y \exp\left(\hat{\alpha}_n^T X + \int \hat{\beta}_n(s) Z(s) ds + \hat{\gamma}_n(t) + \hat{g}(\Lambda(t, U, \hat{\theta}_n))\right) \\ & \quad \times \left(\int \bar{h}(s) Z(s) ds + \hat{\zeta}'_\beta(t, U, \hat{\theta}_n)[\bar{h}]\right) dt \\ & + \int_0^Y \exp\left(\alpha_0^T X + \int \beta_0(s) Z(s) ds + \gamma_0(t) + g_0(\Lambda_0(t, U))\right) \\ & \quad \times \left(\int \bar{h}(s) Z(s) ds + \zeta'_{0\beta}(t, U, \theta_0)[\bar{h}]\right) dt \Bigg\}\Bigg)^2 \end{aligned}$$

$$
\begin{aligned}
\leq\ & P\Bigg\{\Delta\left(\hat{\zeta}'_{n,\beta}(Y,U,\hat{\alpha}_n,\hat{\beta}_n,\hat{\gamma}_n)[\bar{h}]-\zeta'_{0\beta}(Y,U,\alpha_0,\beta_0,\gamma_0)[\bar{h}]\right)^2 \\
& +\int_0^\tau I(Y\geq t)\exp\left(2\psi(t,U,\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{g})\right)\left(\hat{\zeta}'_{\beta}(t,U,\hat{\theta}_n)[\bar{h}]-\zeta'_{0\beta}(t,U,\theta_0)[\bar{h}]\right)^2 dt \\
& +\int_0^\tau I(Y\geq t)\left[\exp(\psi(t,U,\hat{\alpha},\hat{\beta},\hat{\gamma},\hat{g}))-\exp\left(\psi(t,U,\alpha_0,\beta_0,\gamma_0,g_0)\right)\right]^2 \\
& \quad\cdot\left(\int\bar{h}(s)Z(s)ds+\zeta'_{0\beta}(t,U,\theta_0)[\bar{h}]\right)^2 dt\Bigg\}
\end{aligned}
$$

$$
\begin{aligned}
\lesssim\ & \|\hat{\zeta}'_{n,\beta}(\cdot,\hat{\alpha}_n,\hat{\beta}_n,\hat{\gamma}_n,\hat{g}_n)[\bar{h}]-\zeta'_{0\beta}(\cdot,\alpha_0,\beta_0,\gamma_0,g_0)[\bar{h}]\|_2^2+d^2(\theta,\theta_0)\cdot\|\bar{h}\|_\infty^2 \\
=\ & O_p(n^{-2\min\{(p_1+2)\nu_1,p_2\nu_2+2\nu_1,(p_3-1)\nu_3+2\nu_1,(1-\max\{\nu_1,\nu_2,\nu_3\})/2+2\nu_1\}}),
\end{aligned}
$$

where the third inequality is obtained by the boundedness of ψ and mean value theorem. Thus, we have $B_{2n}=o_p(n^{-1/2})$ under the restrictions of Theorem 4. By combining $B_{1n}=o_p(n^{-1/2})$ and $B_{2n}=o_p(n^{-1/2})$, we have $\mathbb{P}_n l'_\beta(\hat{\theta}_n,W)[h^*_{1j}-h_{1j,n}]=o_p(n^{-1/2})$ for $j=1,2,\ldots,d$, which leads to $\mathbb{P}_n\left\{l'_\beta(\hat{\theta}_n,W)[\mathbf{h}^*_1]\right\}=o_p(n^{-1/2})$. Similarly, $\mathbb{P}_n\left\{l'_\gamma(\hat{\theta}_n,W)[\mathbf{h}^*_2]\right\}$ can be proved to be $o_p(n^{-1/2})$. To verify $\mathbb{P}_n\left\{l'_\zeta(\hat{\theta}_n,W)[\tilde{\mathbf{h}}^*_3]\right\}=o_p(n^{-1/2})$ for $\tilde{\mathbf{h}}^*_3=\mathbf{h}^*_3(\hat{\Lambda}(\cdot))+\hat{g}'_n(\hat{\Lambda}(\cdot))\hat{\Lambda}'_g(\cdot)[\mathbf{h}^*_3]$. By Lemma 2, for each $j\in\{1,2,\ldots d\}$, there exists $h_{3j,n}\in\mathcal{G}_n^2$ such that $\|h_{3j,n}-h^*_{3j}\|_\infty=O(n^{-2\nu_3})$ and let $\tilde{h}_{3j,n}=h_{3j,n}(\hat{\Lambda}(\cdot))+\hat{g}'_n(\hat{\Lambda}(\cdot))\hat{\Lambda}'_g(\cdot)[h_{3j,n}]$. Then, $\mathbb{P}_n\left\{l'_\zeta(\hat{\theta}_n,W)[\tilde{h}^*_{3j}]\right\}$ can be decomposed as

$$
\mathbb{P}_n l'_\zeta(\hat{\theta}_n,W)[\tilde{h}^*_{3j}-\tilde{h}_{3j,n}]=B_{3n}+B_{4n},
$$

where

$$B_{3n} = (\mathbb{P}_n - P)\{l'_\zeta(\hat{\theta}_n, W)[\tilde{h}^*_{3j} - \tilde{h}_{3j,n}]\},$$

$$B_{4n} = P\{l'_\zeta(\hat{\theta}_n, W)[\tilde{h}^*_{3j} - \tilde{h}_{3j,n}]\} - P\{l'_\zeta(\theta_0, W)[\tilde{h}^*_{3j} - \tilde{h}_{3j,n}]\}.$$

We first verify $B_{3n} = o_p(n^{-1/2})$. By Lemma 4, we have

$$\begin{aligned} \log N_{[\,]}(\epsilon, \mathcal{F}^{\zeta}_{n,j}(\eta), L_2(P)) &\leq \log N_{[\,]}(\epsilon, \mathcal{F}^{\zeta}_{n,j}(\eta), \|\cdot\|_\infty) \\ &\lesssim (c_1 q_{n1} + c_2 q_{n2} + c_3 q_{n3}) \log(\eta/\epsilon). \end{aligned}$$

The corresponding bracketing integral satisfies

$$\begin{aligned} J_{[\,]}(\eta, \mathcal{F}^{\zeta}_{n,j}(\eta), L_2(P)) &= \int_0^\eta \sqrt{1 + \log N_{[\,]}(\epsilon, \mathcal{F}^{\zeta}_{n,j}(\eta), L_2(P))}\, d\epsilon \\ &\lesssim (c_1 q_{n1} + c_2 q_{n2} + c_3 q_{n3})^{1/2} \eta. \end{aligned} \quad \text{(S8.26)}$$

Pick $\eta_n = O(n^{-\min\{p_1\nu_1, p_2\nu_2, 2\nu_3, (1-\max\{\nu_1,\nu_2,\nu_3\})/2\}})$, then $\|h_{3j,n} - h^*_{3j}\|_\infty = O(n^{-2\nu_3}) \leq \eta_n$, $d(\hat{\theta}_n, \theta_0) \leq \eta_n$. By Lemma 3.4.2 of Van der Vaart and Wellner (1996), we have

$$\begin{aligned} E_P \|\mathbb{G}_n\|_{\mathcal{F}^{\zeta}_{n,j}(\eta_n)} &\lesssim (c_1 q_{n1} + c_2 q_{n2})^{1/2} \eta_n + (c_1 q_{n1} + c_2 q_{n2}) n^{-1/2} \\ &= O(n^{(\max\{\nu_1,\nu_2,\nu_3\})/2}) \cdot O(n^{-\min\{p_1\nu_1, p_2\nu_2, 2\nu_3, (1-\max\{\nu_1,\nu_2,\nu_3\})/2\}}) \\ &\quad + O(n^{\max\{\nu_1,\nu_2,\nu_3\} - 1/2}) \\ &= o(1). \end{aligned}$$

where the last equality holds under the restrictions of Theorem 4. Thus, by Markov's inequality, we have $B_{3n} = o_p(n^{-1/2})$. Next, we show $B_{4n} =$

$o_p(n^{-1/2})$. Denote $\tilde{h} = \tilde{h}^*_{3j} - \tilde{h}_{3j,n}$, then according to mean value theorem,

$$
\begin{aligned}
& \tilde{h}(t,u,\hat{\alpha}_n,\hat{\beta}_n,\hat{\gamma}_n) - \tilde{h}(t,u,\alpha_0,\beta_0,\gamma_0) \\
= \;& (\hat{\alpha}_n - \alpha_0)^T \tilde{h}'_\alpha(t,u,\tilde{\alpha}_n,\tilde{\beta}_n,\tilde{\gamma}_n) + \tilde{h}'_\beta(t,u,\tilde{\alpha}_n,\tilde{\beta}_n,\tilde{\gamma}_n)[\hat{\beta}_n - \beta_0] \\
& + \tilde{h}'_\gamma(t,u,\tilde{\alpha}_n,\tilde{\beta}_n,\tilde{\gamma}_n)[\hat{\gamma}_n - \gamma_0].
\end{aligned}
$$

Based on the boundedness of $\tilde{\gamma}_n, \tilde{g}_n, \tilde{g}'_n, \tilde{g}''_n, \tilde{\Lambda}'_\alpha(\cdot), \tilde{\Lambda}'_\beta(\cdot), \tilde{\Lambda}'_\gamma(\cdot)$, we have

$$
\begin{aligned}
& \|\tilde{h}(\cdot,\tilde{\alpha}_n,\tilde{\beta}_n,\tilde{\gamma}_n)\|_\infty \\
= \;& \|(h_{3j,n} - h^*_{3j})(\tilde{\Lambda}(\cdot)) + \tilde{g}'_n(\tilde{\Lambda}(\cdot))\tilde{\Lambda}'_g(\cdot)[h_{3j,n} - h^*_{3j}]\|_\infty \\
\lesssim \;& \|h_{3j,n} - h^*_{3j}\|_\infty, \\
& \|\tilde{h}'_\alpha(\cdot,\tilde{\alpha}_n,\tilde{\beta}_n,\tilde{\gamma}_n)\|_\infty \\
= \;& \|(h_{3j,n} - h^*_{3j})'(\tilde{\Lambda}(\cdot))\tilde{\Lambda}_\alpha(\cdot) + \tilde{g}'_n(\tilde{\Lambda}(\cdot))\tilde{\Lambda}''_{g\alpha}(\cdot)[h_{3j,n} - h^*_{3j}] \\
& + \tilde{g}''_n(\tilde{\Lambda}(\cdot))\tilde{\Lambda}'_g(\cdot)[h_{3j,n} - h^*_{3j}]\tilde{\Lambda}'_\alpha(\cdot)\|_\infty \\
\lesssim \;& \|h_{3j,n} - h^*_{3j}\|_\infty + \|(h_{3j,n} - h^*_{3j})'\|_\infty, \\
& \|\tilde{h}'_\beta(\cdot,\tilde{\alpha}_n,\tilde{\beta}_n,\tilde{\gamma}_n)[\hat{\beta}_n - \beta_0]\|_2 \\
= \;& \|(h_{3j,n} - h^*_{3j})'(\tilde{\Lambda}(\cdot))\tilde{\Lambda}'_\beta(\cdot) + \tilde{g}'_n(\tilde{\Lambda}(\cdot))\tilde{\Lambda}''_{g\beta}(\cdot)[h_{3j,n} - h^*_{3j}, \hat{\beta}_n - \beta_0] \\
& + \tilde{g}''_n(\tilde{\Lambda}(\cdot))\tilde{\Lambda}'_g(\cdot)[h_{3j,n} - h^*_{3j}]\tilde{\Lambda}'_\beta(\cdot)[\hat{\beta}_n - \beta_0]\|_2 \\
\lesssim \;& (\|h_{3j,n} - h^*_{3j}\|_\infty + \|(h_{3j,n} - h^*_{3j})'\|_\infty)\|\hat{\beta}_n - \beta_0\|_2,
\end{aligned}
$$

$$
\begin{aligned}
& \|\tilde{h}'_{\gamma}(\cdot,\tilde{\alpha}_n,\tilde{\beta}_n,\tilde{\gamma}_n)[\hat{\gamma}_n-\gamma_0]\|_2 \\
= \; & \|(h_{3j,n}-h^*_{3j})'(\tilde{\Lambda}(\cdot))\tilde{\Lambda}'_{\gamma}(\cdot)+\tilde{g}'_n(\tilde{\Lambda}(\cdot))\tilde{\Lambda}''_{g\gamma}(\cdot)[h_{3j,n}-h^*_{3j},\hat{\gamma}_n-\gamma_0] \\
& +\tilde{g}''_n(\tilde{\Lambda}(\cdot))\tilde{\Lambda}'_{g}(\cdot)[h_{3j,n}-h^*_{3j}]\tilde{\Lambda}'_{\gamma}(\cdot)[\hat{\gamma}_n-\gamma_0]\|_2 \\
\lesssim \; & (\|h_{3j,n}-h^*_{3j}\|_\infty+\|(h_{3j,n}-h^*_{3j})'\|_\infty)\|\hat{\gamma}_n-\gamma_0\|_2.
\end{aligned}
$$

Hence, $\tilde{h}(t,u,\hat{\alpha}_n,\hat{\beta}_n,\hat{\gamma}_n)-\tilde{h}(t,u,\alpha_0,\beta_0,\gamma_0)\lesssim(\|h_{3j,n}-h^*_{3j}\|_\infty+\|(h_{3j,n}-h^*_{3j})'\|_\infty)d(\hat{\theta}_n,\theta_0)$. It follows that

$$
\begin{aligned}
B_{4n}^2 \;\le\; & P\left\{\left(l'_\zeta(\hat{\theta}_n,W)-l'_\zeta(\theta_0,W)\right)[\tilde{h}]\right\}^2 \\
= \; & P\Bigg\{\Delta\left(\tilde{h}(Y,U,\hat{\alpha}_n,\hat{\beta}_n,\hat{\gamma}_n)-\tilde{h}(Y,U,\alpha_0,\beta_0,\gamma_0)\right) \\
& -\int_0^\tau I(Y\ge t)\exp\left(\psi(t,U,\hat{\theta}_n)\right)\bar{h}(t,U,\hat{\alpha}_n,\hat{\beta}_n,\hat{\gamma}_n)dt \\
& +\int_0^\tau I(Y\ge t)\exp\left(\psi(t,U,\theta_0)\right)\bar{h}(t,U,\alpha_0,\beta_0,\gamma_0)dt\Bigg\}^2 \\
\lesssim \; & P\Bigg\{\Delta\left(\tilde{h}(Y,U,\hat{\alpha}_n,\hat{\beta}_n,\hat{\gamma}_n)-\tilde{h}(Y,U,\alpha_0,\beta_0,\gamma_0)\right) \\
& -\int_0^\tau\left(\exp(\psi(t,U,\hat{\theta}_n))-\exp\left(\psi(t,U,\theta_0)\right)\right)\bar{h}(t,U,\hat{\alpha}_n,\hat{\beta}_n,\hat{\gamma}_n)dt \\
& -\int_0^\tau\exp\left(\psi(t,U,\theta_0)\right)\left(\bar{h}(t,U,\hat{\alpha}_n,\hat{\beta}_n,\hat{\gamma}_n)-\bar{h}(t,U,\alpha_0,\beta_0,\gamma_0)\right)dt\Bigg\}^2 \\
\lesssim \; & (\|h_{3j,n}-h^*_{3j}\|_\infty+\|(h_{3j,n}-h^*_{3j})'\|_\infty)^2 d^2(\hat{\theta}_n,\theta_0).
\end{aligned}
$$

Using Corollary 6.21 in Schumaker (2007), we have $\|(h_{3j,n}-h^*_{3j})'\|_\infty=O(n^{-\nu_3})$, hence, $B_{4n}=o_p(n^{-1/2})$ under the restrictions of Theorem 4. Combining $B_{3n}=o_p(n^{-1/2})$ and $B_{4n}=o_p(n^{-1/2})$, we have $\mathbb{P}_n\left\{l'_\zeta(\hat{\theta}_n,W)[\tilde{h}^*_{3j}]\right\}=o_p(n^{-1/2})$ for each $j\in\{1,\ldots,d\}$, which indicates $\mathbb{P}_n\left\{l'_\zeta(\hat{\theta}_n,W)[\tilde{\mathbf{h}}^*_3]\right\}=$

$o_p(n^{-1/2})$. Thus, Assumption (A4) holds.

Now we verify assumption (A5). First, by Lemma 5, the ϵ-bracketing numbers associated with $\|\cdot\|_\infty$ norm for the classes of functions $\mathcal{F}^{*\alpha}_{n,j}(\eta), \mathcal{F}^{*\beta}_{n,j}(\eta)$, $\mathcal{F}^{*\gamma}_{n,j}(\eta), \mathcal{F}^{*\zeta}_{n,j}(\eta)$ are all bounded by $(\eta/\epsilon)^{c_1q_{n1}+c_2q_{n2}+c_3q_{n3}+d}$, which implies that the corresponding ϵ-bracketing integrals are all bounded by $(c_1q_{n1} + c_2q_{n2}+c_3q_{n3})^{1/2}\eta$, that is, $J_{[\,]}(\eta, \mathcal{F}^{*\alpha}_{n,j}(\eta), \|\cdot\|_\infty) \lesssim (c_1q_{n1}+c_2q_{n2}+c_3q_{n3})^{1/2}\eta$, $J_{[\,]}(\eta, \mathcal{F}^{*\beta}_{n,j}(\eta), \|\cdot\|_\infty) \lesssim (c_1q_{n1}+c_2q_{n2}+c_3q_{n3})^{1/2}\eta$, $J_{[\,]}(\eta, \mathcal{F}^{*\gamma}_{n,j}(\eta), \|\cdot\|_\infty) \lesssim (c_1q_{n1}+c_2q_{n2}+c_3q_{n3})^{1/2}\eta$ and $J_{[\,]}(\eta, \mathcal{F}^{*\zeta}_{n,j}(\eta), \|\cdot\|_\infty) \lesssim (c_1q_{n1}+c_2q_{n2}+c_3q_{n3})^{1/2}\eta$. Then for $l'_{\alpha_j}(\theta, W) - l'_{\alpha_j}(\theta_0, W)$, by applying Cauchy-Schwarz inequality together with subtracting and adding $g'_0(\Lambda_0(Y))\Lambda'_{\alpha_j}(Y, U, \theta)$, $\int_0^Y (X + g'(\Lambda(t, U, \theta))\Lambda'_{\alpha_j}(t, U, \theta))d\Lambda_0(t, U)$, we have

$$
\begin{aligned}
& \{l'_{\alpha_j}(\theta, W) - l'_{\alpha_j}(\theta_0, W)\}^2 \\
= \; & \Big\{\Delta\Big(g'(\Lambda(Y,U,\theta))\Lambda'_{\alpha_j}(Y,U,\theta) - g'_0(\Lambda_0(Y,U))\Lambda'_{0\alpha_j}(Y,U)\Big) \\
& - \int_0^\tau I(Y \ge t)\Big(X_j + g'(\Lambda(t,U,\theta))\Lambda'_{\alpha_j}(t,U,\theta)\Big)\,d\Lambda(t,U,\theta) \\
& + \int_0^\tau I(Y \ge t)\Big(X_j + g'_0(\Lambda_0(t,U))\Lambda'_{0\alpha_j}(t,U)\Big)\,d\Lambda_0(t,U)\Big\}^2 \\
\lesssim \; & C_1 + C_2 + C_3 + C_4,
\end{aligned}
$$

where

$$
\begin{aligned}
C_1 &= \Delta\left(g'(\Lambda(Y,U,\theta))\Lambda'_{\alpha_j}(Y,U,\theta) - g'_0(\Lambda_0(Y,U))\Lambda'_{\alpha_j}(Y,U,\theta)\right)^2, \\
C_2 &= \Delta\left(g'_0(\Lambda_0(Y,U))\Lambda'_{\alpha_j}(Y,U,\theta) - g'_0(\Lambda_0(Y,U))\Lambda'_{0\alpha_j}(Y,U)\right)^2, \\
C_3 &= \int_0^\tau \left\{(\exp(\psi(t,U,\theta)) - \exp(\psi(t,U,\theta_0)))\left(X_j + g'(\Lambda(t,U,\theta))\Lambda'_{\alpha_j}(t,U,\theta)\right)\right\}^2 dt, \\
C_4 &= \int_0^\tau \left(g'(\Lambda_0(t,U,\theta))\Lambda'_{\alpha_j}(t,U,\theta) - g'_0(\Lambda_0(t,U))\Lambda'_{0\alpha_j}(t,U)\right)^2 \exp(2\psi(t,U,\theta_0))dt.
\end{aligned}
$$

Apparently, $PC_1 \le \eta^2$. For C_2, by the boundedness of $g'_0, \Lambda'_{\alpha_j\theta}$ and mean value theorem, we have

$$
\begin{aligned}
PC_2 &\lesssim P\left(\Lambda'_{\alpha_j}(Y,U,\theta) - \Lambda'_{0\alpha_j}(Y,U)\right)^2 \\
&= P\left(\Lambda'_{\theta\alpha_j}(Y,U,\tilde{\theta})[\theta-\theta_0]\right)^2 \\
&\lesssim d^2(\theta,\theta_0) \lesssim \eta^2,
\end{aligned}
$$

where $\tilde{\theta}$ is a point between θ and θ_0. For C_3, by the boundedness of ψ, X_j, g' and Λ_{α_j} and mean value theorem, we have

$$
\begin{aligned}
PC_3 &\lesssim P\int_0^\tau \exp(2\tilde{\psi})(\psi(t,U,\theta) - \psi(t,U,\theta))^2 dt \\
&\lesssim d^2(\theta,\theta_0) \lesssim \eta^2,
\end{aligned}
$$

and for C_4, we have

$$\begin{aligned} PC_4 &\lesssim \int_0^\tau \Big(g'(\Lambda(t,U,\theta))\Lambda'_{\alpha_j}(t,U,\theta) - g'_0(\Lambda_0(t,U,\theta))\Lambda'_{\alpha_j}(t,U)\Big)^2 dt \\ &\quad + \int_0^\tau \Big(g'_0(\Lambda_0(t,U))\Lambda'_{\alpha_j}(t,U,\theta) - g'_0(\Lambda_0(t,U))\Lambda'_{0\alpha_j}(t,U)\Big)^2 dt \\ &\lesssim d^2(\theta,\theta_0) + \|g'(\Lambda(\cdot,\theta)) - g'_0(\Lambda_0(\cdot))\|_2 \lesssim \eta^2. \end{aligned}$$

Therefore, we have $P\{l'_{\alpha_j}(\theta,W)-l'_{\alpha_j}(\theta_0,W)\}^2 \le \eta^2$. Using the similar argument, we can show that $P\{l'_\beta(\theta,W)[h^*_{1j}]-l'_\beta(\theta_0,W)[h^*_{1j}]\}^2$, $P\{l'_\gamma(\theta,W)[h^*_{2j}]-l'_\gamma(\theta_0,W)[h^*_{2j}]\}^2$, $P\{l'_\zeta(\theta,W)[\tilde{h}^*_{3j}] - l'_\zeta(\theta_0,W)[\tilde{h}^*_{3j}]\}^2$ are all bounded by η^2. We also have $\|l'_{\alpha_j}(\theta,W) - l'_{\alpha_j}(\theta_0,W)\|_\infty$, $\|l'_\beta(\theta,W)[h^*_{1j}] - l'_\beta(\theta_0,W)[h^*_{1j}]\|_\infty$, $\|l'_\gamma(\theta,W)[h^*_{2j}]-l'_\gamma(\theta_0,W)[h^*_{2j}]\|_\infty$ and $\|l'_\zeta(\theta,W)[\tilde{h}^*_{3j}]-l'_\zeta(\theta_0,W)[\tilde{h}^*_{3j}]\|_\infty$ are all bounded. Now we choose $\eta_n = O(n^{-\min\{p_1\nu_1, p_2\nu_2, (p_3-1)\nu_3, (1-\max\{\nu_1,\nu_2,\nu_3\})/2\}})$, By the maximal inequality in Lemma 3.4.2 of Van der Vaart and Wellner (1996) and the restrictions of Theorem 4, it follows that

$$\begin{aligned} E_P\|\mathbb{G}_n\|_{\mathcal{F}^{*\alpha_j}_{n,j}(\eta_n)} &\lesssim (c_1q_{n1}+c_2q_{n2}+c_3q_{n3})^{1/2}\eta_n + (c_1q_{n1}+c_2q_{n2}+c_3q_{n3})n^{-1/2} \\ &= O(n^{\max\{\nu_1,\nu_2,\nu_3\}})\eta_n + O(n^{\max\{\nu_1,\nu_2,\nu_3\}-1/2}) \\ &= o(1). \end{aligned}$$

Similarly, we have $E_P\|\mathbb{G}_n\|_{\mathcal{F}^{*\beta_j}_{n,j}(\eta_n)} = o(1)$, $E_P\|\mathbb{G}_n\|_{\mathcal{F}^{*\gamma_j}_{n,j}(\eta_n)} = o(1)$ and $E_P\|\mathbb{G}_n\|_{\mathcal{F}^{*\zeta_j}_{n,j}(\eta_n)} = o(1)$. Thus for $\xi = c$ and $Cn^{-\xi} = O(n^{-c})$, by Markov's inequality, we have

$$\begin{aligned}
\sup_{d(\theta,\theta_0)\le Cn^{-\xi},\theta\in\Theta_n} |\mathbb{G}_n\{l'_{\alpha_j}(\theta,W)-l'_{\alpha_j}(\theta_0,W)\}| &= o_p(1),\\
\sup_{d(\theta,\theta_0)\le Cn^{-\xi},\theta\in\Theta_n} |\mathbb{G}_n\{l'_{\beta_j}(\theta,W)[h^*_{1j}]-l'_{\beta_j}(\theta_0,W)[h^*_{1j}]\}| &= o_p(1),\\
\sup_{d(\theta,\theta_0)\le Cn^{-\xi},\theta\in\Theta_n} |\mathbb{G}_n\{l'_{\gamma_j}(\theta,W)[h^*_{2j}]-l'_{\gamma_j}(\theta_0,W)[h^*_{2j}]\}| &= o_p(1),\\
\sup_{d(\theta,\theta_0)\le Cn^{-\xi},\theta\in\Theta_n} |\mathbb{G}_n\{l'_{\zeta_j}(\theta,W)[\tilde{h}^*_{3j}]-l'_{\zeta_j}(\theta_0,W)[\tilde{h}^*_{3j}]\}| &= o_p(1).
\end{aligned}$$

This completes the verification of assumption (A5).

Finally, assumption (A6) can be verified using Taylor expansion. Since the proof of three equations are essentially identical, we only present the proof of the first equation. In a neighborhood of θ_0 : $\{\theta : d(\theta,\theta_0) \le Cn^{-\xi}, \theta\in\Theta_n\}$ with $\xi=-c$, the Taylor expansion for $l_\alpha(\theta,W)$ yields

$$\begin{aligned}
l'_\alpha(\theta,W) &= l'_\alpha(\theta_0,W)+l''_{\alpha\alpha}(\tilde{\theta},W)(\alpha-\alpha_0)+l''_{\alpha\beta}(\tilde{\theta},W)[\beta-\beta_0]\\
&\quad +l''_{\alpha\gamma}(\tilde{\theta},W)[\gamma-\gamma_0]+l''_{\alpha\zeta}(\tilde{\theta},W)[\zeta-\zeta_0],
\end{aligned}$$

where $\tilde{\theta}$ is a point between θ and θ_0. So

$$\begin{aligned}
& P\{l'_\alpha(\theta,W)\}-P\{l'_\alpha(\theta_0,W)\}-P\{l''_{\alpha\alpha}(\theta_0,W)(\alpha-\alpha_0)\}-P\{l''_{\alpha\beta}(\theta_0,W)[\beta-\beta_0]\}\\
&\quad -P\{l''_{\alpha\gamma}(\theta_0,W)[\gamma-\gamma_0]\}-P\{l''_{\alpha\zeta}(\theta_0,W)[\zeta-\zeta_0]\}\\
=\ & P\{(l''_{\alpha\alpha}(\tilde{\theta},W)-l''_{\alpha\alpha}(\theta_0,W))(\alpha-\alpha_0)\}+P\{(l''_{\alpha\beta}(\tilde{\theta},W)-l''_{\alpha\beta}(\theta_0,W))[\beta-\beta_0]\}\\
& +P\{(l''_{\alpha\gamma}(\tilde{\theta},W)-l''_{\alpha\gamma}(\theta_0,W))[\gamma-\gamma_0]\}+P\{(l''_{\alpha\zeta}(\tilde{\theta},W)-l''_{\alpha\zeta}(\theta_0,W))[\zeta-\zeta_0]\}.
\end{aligned}$$

Direct caluclation yileds

$$\begin{aligned}
& |P\{(l''_{\alpha\alpha}(\tilde{\theta}, W) - l''_{\alpha\alpha}(\theta_0, W))\}| \\
\leq\ & P\left\{\int_0^Y \left|\tilde{\zeta}''_{\alpha\alpha}(t,U)(\exp(\psi(\tilde{\theta}, W)) - \exp(\psi(\theta_0, W)))\right| dt\right\} \\
& + P\left\{\int_0^Y \left|(X + \tilde{\zeta}'_{\alpha}(t,U))(X + \tilde{\zeta}'_{\alpha}(t,U))^T \left(\exp(\psi(\tilde{\theta}, W)) - \exp(\psi(\theta_0, W))\right)\right| dt\right\} \\
& + P\left\{\int_0^Y \left|(X + \zeta'_{0\alpha}(t,U))(X + \zeta'_{0\alpha}(t,U))^T - (X + \tilde{\zeta}'_{\alpha}(t,U))(X + \tilde{\zeta}'_{\alpha}(t,U))^T\right| d\Lambda_0(t)\right\} \\
=\ & D_1 + D_2 + D_3.
\end{aligned}$$

By applying mean value theorem and the Cauchy-Schwarz inequality, we can show that

$$\begin{aligned}
D_1 &= P\bigg\{\int_0^Y \left((\tilde{\alpha} - \alpha_0)^T X + \int (\tilde{\beta} - \beta_0)(s) Z(s) ds + (\tilde{\gamma} - \gamma_0)(t) + (\tilde{\zeta} - \zeta_0)(t,U)\right) \\
&\qquad \times \tilde{\zeta}''_{\alpha\alpha}(t,U) \exp(\tilde{\psi}(t,U)) dt\bigg\} \\
&\lesssim \|\tilde{\alpha} - \alpha_0\|_2 + \|\tilde{\beta} - \beta_0\|_C + \|\tilde{\gamma} - \gamma_0\|_2 + \|\tilde{\zeta} - \zeta_0\|_2 \leq d(\theta_0, \theta) \\
&\leq O(n^{-c}).
\end{aligned}$$

For D_2, based on the boundedness of X and $\tilde{\zeta}'_{\alpha}$, we can similarly show that

$$D_2 \lesssim \|\tilde{\alpha} - \alpha_0\|_2 + \|\tilde{\beta} - \beta_0\|_C + \|\tilde{\gamma} - \gamma_0\|_2 + \|\tilde{\zeta} - \zeta_0\|_2 = O(n^{-c}).$$

For D_3, using a similar argument that we used before for verifying assumption (A5), we can show that

$$D_3 \lesssim \|\zeta'_{0\alpha}(\cdot) - \tilde{\zeta}'_{\alpha}(\cdot)\|_2 = O(n^{-\min\{p_1\nu_1, p_2\nu_2, (p_3-1)\nu_3, (1-\max\{\nu_1,\nu_2,\nu_3\})/2\}}).$$

Therefore,

$$\begin{aligned}
& |P\{(l''_{\alpha\alpha}(\tilde{\theta}, W) - l''_{\alpha\alpha}(\theta_0, W))(\alpha - \alpha_0)\}| \\
= \; & O(n^{-c}) \cdot O(n^{-\min\{p_1\nu_1, p_2\nu_2, (p_3-1)\nu_3, (1-\max\{\nu_1,\nu_2,\nu_3\})/2\}}) \\
= \; & o(n^{-1/2}).
\end{aligned}$$

Similarly, we can show that

$$\begin{aligned}
& P|(l''_{\alpha\beta}(\tilde{\theta}, W) - l''_{\alpha\beta}(\theta_0, W))[\beta - \beta_0]| \\
= \; & O(n^{-c}) \cdot O(n^{-\min\{p_1\nu_1, p_2\nu_2, (p_3-1)\nu_3, (1-\max\{\nu_1,\nu_2,\nu_3\})/2\}}) = o(n^{-1/2}), \\
& P|(l''_{\alpha\gamma}(\tilde{\theta}, W) - l''_{\alpha\gamma}(\theta_0, W))[\gamma - \gamma_0]| \\
= \; & O(n^{-c}) \cdot O(n^{-\min\{p_1\nu_1, p_2\nu_2, (p_3-1)\nu_3, (1-\max\{\nu_1,\nu_2,\nu_3\})/2\}}) = o(n^{-1/2}),
\end{aligned}$$

and

$$\begin{aligned}
& P\{(l''_{\alpha\zeta}(\tilde{\theta}, W) - l''_{\alpha\zeta}(\theta_0, W))[\zeta - \zeta_0]\} \\
= \; & O(n^{-c}) \cdot O(n^{-\min\{p_1\nu_1, p_2\nu_2, (p_3-1)\nu_3, (1-\max\{\nu_1,\nu_2,\nu_3\})/2\}}) = o(n^{-1/2}).
\end{aligned}$$

Thus, we have

$$\begin{aligned}
& P\{l'_{\alpha}(\theta, W)\} - P\{l'_{\alpha}(\theta_0, W)\} - P\{l''_{\alpha\alpha}(\theta_0, W)(\alpha - \alpha_0)\} \\
& \quad -P\{l''_{\alpha\beta}(\theta_0, W)[\beta - \beta_0]\} - P\{l''_{\alpha\gamma}(\theta_0, W)[\gamma - \gamma_0]\} - P\{l''_{\alpha\zeta}(\theta_0, W)[\zeta - \zeta_0]\} \\
= \; & O(n^{-c-\min\{p_1\nu_1, p_2\nu_2, (p_3-1)\nu_3, (1-\max\{\nu_1,\nu_2,\nu_3\})/2\}}) = O(n^{-\alpha\xi}),
\end{aligned}$$

where $\alpha = 1 + (\min\{p_1\nu_1, p_2\nu_2, (p_3 - 1)\nu_3, (1 - \max\{\nu_1, \nu_2, \nu_3\})/2\})/c > 1$

and $\alpha\xi > 1/2$.

Therefore, we have verified all six assumptions. By Theorem 3 of Tang (2021), we have

$$\sqrt{n}(\hat{\alpha}_n - \alpha_0) = A^{-1}\sqrt{n}\mathbb{P}_n \mathbf{l}^*(\alpha_0, \beta_0, \gamma_0, \zeta_0, W) + o_p(1) \to N(0, A^{-1}B(A^{-1})^T),$$

where $\mathbf{l}^*(\alpha_0, \beta_0, \gamma_0, \zeta_0, W)$ is the efficient score function for α_0 and $A = P\{\mathbf{l}^*(\alpha_0, \beta_0, \gamma_0, \zeta_0, W)^{\otimes 2}\} = I(\alpha_0)$, which is shown when verifying assumption (A3). Thus, $A = B = I(\alpha_0)$ and $A^{-1}B(A^{-1})^T = I^{-1}(\alpha_0)$, and

$$\sqrt{n}(\hat{\alpha}_n - \alpha_0) \to N(0, I^{-1}(\alpha_0)).$$

This completes the proof.

S9 Proof of Theorem 5

We first introduce some notations. Let $\mathbf{B}_n^{\beta} = \left(B_1^{\beta}, \ldots, B_{q_{n,1}}^{\beta}\right)^T$ be a B-spline basis for estimating β_0 and $\mathbf{v}_1^* \in \mathcal{B}^{q_{n,1}}, \mathbf{v}_2^* \in \overline{\mathbb{T}}_{\gamma_0}^{q_{n,1}}, \mathbf{v}_3^* \in \overline{\mathbb{T}}_{g_0}^{q_{n,1}}$ satisfies

$$\begin{aligned}
&\mathbb{E}\left\{\left(\dot{l}_{\beta_0}[\mathbf{B}_n^{\beta}] - \dot{l}_{\alpha_0}\mathbf{v}_1^* - \dot{l}_{\gamma_0}\mathbf{v}_2^* - \dot{l}_{g_0}\mathbf{v}_3^*\right)\dot{l}_{\alpha_0}^T v\right\} = 0 \quad \text{for all} \quad v \in \mathcal{B}, \\
&\mathbb{E}\left\{\left(\dot{l}_{\beta_0}[\mathbf{B}_n^{\beta}] - \dot{l}_{\alpha_0}\mathbf{v}_1^* - \dot{l}_{\gamma_0}\mathbf{v}_2^* - \dot{l}_{g_0}\mathbf{v}_3^*\right)\dot{l}_{\gamma_0} v\right\} = 0 \quad \text{for all} \quad v \in \overline{\mathbb{T}}_{\gamma_0}, \\
&\mathbb{E}\left\{\left(\dot{l}_{\beta_0}[\mathbf{B}_n^{\beta}] - \dot{l}_{\alpha_0}\mathbf{v}_1^* - \dot{l}_{\gamma_0}\mathbf{v}_2^* - \dot{l}_{g_0}\mathbf{v}_3^*\right)\dot{l}_{g_0} v\right\} = 0 \quad \text{for all} \quad v \in \overline{\mathbb{T}}_{g_0}
\end{aligned} \tag{S9.27}$$

Similar as the line of proving Theorem 2, $\mathbf{v}_1^*, \mathbf{v}_2^*, \mathbf{v}_3^*$ can be found by minimizing

$$\mathbb{E}\Bigg[\Delta\Big\| \Big(g_0'(\tilde{\Lambda}_0(R))\exp\Big(g_0(\tilde{\Lambda}_0(R))\Big)R+1\Big)\Big(\int_0^1 \mathbf{B}_n^\beta(s)Z(s)ds-\mathbf{v}_1X\Big)$$
$$-g_0'(\tilde{\Lambda}_0(R))\exp\Big(g_0(\tilde{\Lambda}_0(R))\Big)\int_0^R \mathbf{v}_2(L^{-1}(se^{-V}))ds+\mathbf{v}_2(L^{-1}(Re^{-V}))$$
$$g_0'(\tilde{\Lambda}_0(R))\exp\Big(g_0(\tilde{\Lambda}_0(R))\Big)\int_0^{\tilde{\Lambda}_0(R)} \exp(-g_0(s))\mathbf{v}_3(s)ds+\mathbf{v}_3(\tilde{\Lambda}_0(R))\Big\|^2\Bigg] \tag{S9.28}$$

Let $\dot{l}^*_{\beta_0 n}=\dot{l}_{\beta_0}[\mathbf{B}_n^\beta]-\dot{l}_{\alpha_0}\mathbf{v}_1^*-\dot{l}_{\gamma_0}\mathbf{v}_2^*-\dot{l}_{g_0}\mathbf{v}_3^*$ and $J_n=\mathbb{E}[\dot{l}^{*\otimes 2}_{\beta_0 n}]$. We first investigate the eigenvalue of J_n. For any vector $v\in\mathbb{R}^{q_{n,1}}$,

$$\begin{aligned}
& v^TJ_nv \\
= \;& \mathbb{E}\Bigg[\Delta\Big| \Big(g_0'(\tilde{\Lambda}_0(R))\exp\Big(g_0(\tilde{\Lambda}_0(R))\Big)R+1\Big)\Big(\int_0^1 v^T\mathbf{B}_n^\beta(s)Z(s)ds-v^T\mathbf{v}_1^*X\Big) \\
& -g_0'(\tilde{\Lambda}_0(R))\exp\Big(g_0(\tilde{\Lambda}_0(R))\Big)\int_0^R v^T\mathbf{v}_2^*(L^{-1}(se^{-V}))ds+v^T\mathbf{v}_2^*(L^{-1}(Re^{-V})) \\
& g_0'(\tilde{\Lambda}_0(R))\exp\Big(g_0(\tilde{\Lambda}_0(R))\Big)\int_0^{\tilde{\Lambda}_0(R)} \exp(-g_0(s))v^T\mathbf{v}_3^*(s)ds+v^T\mathbf{v}_3^*(\tilde{\Lambda}_0(R))\Big|^2\Bigg] \\
\lesssim \;& \mathbb{E}\left\{\int_0^1 v^T\mathbf{B}_n^\beta(s)Z(s)ds\right\}^2 \\
= \;& v^T\left\{\int_0^1\int_0^1 \mathbf{B}_n^\beta(s)E[Z(s)Z(t)]\mathbf{B}_n^\beta(t)^Tdsdt\right\}v \\
\lesssim \;& \int_0^1 (v^T\mathbf{B}_n^\beta(s))^2ds=\|v\|_2^2O(n^{\nu_1}).
\end{aligned}$$

On the other hand, follow a similar argument in proving Theorem 3, we have

$$
\begin{aligned}
v^T J_n v &\gtrsim \mathbb{E}\left\{\int_0^1 v^T \mathbf{B}_n^\beta(s) Z(s) ds - v^T \mathbf{v}_1^* X\right\}^2 \\
&= \mathbb{E}\left\{\int_0^1 v^T \mathbf{B}_n^\beta(s)(Z(s) - K(s)) ds\right\}^2 \\
&\quad + \mathbb{E}\left\{\int_0^1 v^T \mathbf{B}_n^\beta(s) K(s) ds - v^T \mathbf{v}_1^* X\right\}^2 \\
&\geq \int_0^1 \int_0^1 v^T \mathbf{B}_n^\beta(s) K(s,t) \mathbf{B}_n^\beta(t)^T v ds dt.
\end{aligned}
$$

Based on the fact that the covariance kernel $K(s,t)$ is symmetric and semi-positive definite, we may denote $\mathcal{H}(K)$ as the reproducing kernel Hilbert space induced by kernel function K. Let

$$\mathcal{H}(P) = \left\{f \in \mathcal{W}_2^1[0,1] : f(0) = f(1) = 0\right\},$$

where $\mathcal{W}_2^1[0,1]$ is the Sobolev space defined as

$$\mathcal{W}_2^1[0,1] = \left\{f \in C^1([0,1]) : f \text{ absolutely continuous }, f' \in \mathcal{L}^2[0,1]\right\}.$$

Noted that $\mathcal{H}(P)$ is also a reproducing kernel Hilbert space with kernel function $G(s,t) = \min\{s,t\}$. According to Corollary 1 in Ritter et al. (1995), we have $\mathcal{H}(P) \subseteq \mathcal{H}(K) \subseteq \mathcal{W}_2^1[0,1]$. Therefore, for any $f \in \mathcal{H}(K)$,

$$|f(s)| = |\langle f, K(s,\cdot)\rangle_{\mathcal{H}(K)}| \leq \|K(s,\cdot)\|_{\mathcal{H}(K)} \|f\|_{\mathcal{H}(K)} \lesssim \sqrt{K(s,s)} \|f\|_K,$$

with $\|f\|_K^2 = \int_0^1 \int_0^1 f(s)K(s,t)f(t)dtds$ and the last inequality follows from Corollary 2 in Ritter et al. (1995). It follows that

$$\int_0^1 f^2(s)ds \lesssim \|f\|_K^2 \int_0^1 K(s,s)ds.$$

Since $v^T \mathbf{B}_n^{\beta}(s) \in \mathcal{H}(P) \subseteq \mathcal{H}(K)$, we have

$$\int_0^1 \int_0^1 v^T \mathbf{B}_n^{\beta}(s) K(s,t) \mathbf{B}_n^{\beta}(t)^T v ds dt \gtrsim \int_0^1 (v^T \mathbf{B}_n^{\beta}(s))^2 ds = \|v\|_2^2 O(n^{\nu_1}).$$

Therefore, the largest and smallest eigenvalue of J_n, denoted by $\lambda_{\max}(J_n)$ and $\lambda_{\min}(J_n)$ respectively, satisfies

$$c_1 n^{\nu_1} \leq \lambda_{\min}(J_n) \leq \lambda_{\max}(J_n) \leq c_2 n^{\nu_1} \tag{S9.29}$$

for some positive constant c_1 and c_2. Next we derive the asymptotic confidence band for β_0. Denote

$$\Gamma(\theta, W) = l'_{\beta}(\theta, W)[\mathbf{B}_n] - l'_{\alpha}(\theta, W)[\mathbf{v}_1^*] - l'_{\gamma}(\theta, W)[\mathbf{v}_2^*] - l'_{\zeta}(\theta, W)[\tilde{\mathbf{v}}_3^*],$$

with $\tilde{\mathbf{v}}_3^* = \mathbf{v}_3^*(\Lambda_0(\cdot)) + g_0'(\Lambda_0(\cdot))\Lambda'_{0g}(\cdot)[\mathbf{v}_3^*]$. Follow a similar argument as proving Theorem 4, we have

$$\sqrt{n}\mathbb{P}_n \left\{\Gamma(\theta_0, W)\right\} = -\sqrt{n}\mathbb{P}\left\{\Gamma'_{\beta}(\theta_0, W)\right\}[\hat{\beta}_n - \beta_0] + R_n, \tag{S9.30}$$

where the remainder term satisfies $\|R_n\|_2 = o_P(n^{\frac{\nu_1}{2} - \min\{c_1, c_2\}})$ with $c_1 = \frac{1}{2} - 2c + \nu_3$ and $c_2 = \frac{\nu_{\max}}{2} - \min\{p_1\nu_1, p_2\nu_2, (p_3-1)\nu_3, \frac{1-\nu_{\max}}{2}\}$. It follows that

$$
\begin{aligned}
&\frac{\sqrt{n}(\mathbf{B}_n^\beta(s))^T J_n^{-1}\mathbb{P}_n\left\{\Gamma(\theta_0, W)\right\}}{\sqrt{(\mathbf{B}_n^\beta(s))^T J_n^{-1}\mathbf{B}_n^\beta(s)}} \\
= \quad & -\frac{\sqrt{n}\mathbb{P}\left\{\Gamma'_\beta(\theta_0, W)\right\}[\hat{\beta}_n - \beta_0]}{\sqrt{(\mathbf{B}_n^\beta(s))^T J_n^{-1}\mathbf{B}_n^\beta(s)}} + \frac{\mathbf{B}_n^\beta(s))^T J_n^{-1} R_n}{\sqrt{(\mathbf{B}_n^\beta(s))^T J_n^{-1}\mathbf{B}_n^\beta(s)}}. \quad \text{(S9.31)}
\end{aligned}
$$

Combining (S9.29) and $\frac{1}{2p_1+2} < \nu_1 < -1/2 + 2c - \nu_3$, we have uniformly over $s \in (0, 1)$,

$$
\frac{\mathbf{B}_n^\beta(s))^T J_n^{-1} R_n}{\sqrt{(\mathbf{B}_n^\beta(s))^T J_n^{-1}\mathbf{B}_n^\beta(s)}} = o_P(1).
$$

Recall that Lemma 1.2 indicates there exists $\beta_{0n} \in \mathcal{F}_n^{p_1}$ such that $\|\beta_{0n} - \beta_0\|_\infty = O(n^{-\nu_1 p_1})$, we assume $\beta_{0n} = b_0^T \mathbf{B}_n^\beta$ and $\hat{\beta}_n = b_n^T \mathbf{B}_n^\beta$. Therefore,

$$
\begin{aligned}
&\sqrt{n}\mathbb{P}\left\{\Gamma'_\beta(\theta_0, W)\right\}[\hat{\beta}_n - \beta_0] \\
=&\sqrt{n}\mathbb{P}\left\{\Gamma'_\beta(\theta_0, W)\right\}[\hat{\beta}_n - \beta_{0n}] + \sqrt{n}\mathbb{P}\left\{\Gamma'_\beta(\theta_0, W)\right\}[\beta_{0n} - \beta_0] \\
=&\sqrt{n}J_n(b_n - b_0) + Q_n
\end{aligned}
$$

with $\|Q_n\|_2 = O(n^{\frac{\nu_1}{2} - \nu_1 p_1})$, where the last equality is obtained by (S9.27). Multiplying $(\mathbf{B}_n^\beta(s))^T J_n^{-1}$ on both sides, we have

$$
\sqrt{n}\left(\hat{\beta}_n(s) - \beta_{0n}(s)\right) = -\sqrt{n}(\mathbf{B}_n^\beta(s))^T J_n^{-1}\mathbb{P}\left\{\Gamma'_\beta(\theta_0, W)\right\}[\hat{\beta}_n - \beta_0] + o(n^{-\nu/2})
$$

holds uniformly over $s \in (0, 1)$. Together with (S9.31) indicates

$$
\sup_{s\in(0,1)}\left|\frac{\sqrt{n}\left(\hat{\beta}_n(s) - \beta_0(s)\right)}{\sqrt{(\mathbf{B}_n^\beta(s))^T J_n^{-1}\mathbf{B}_n^\beta(s)}} - \frac{\sqrt{n}(\mathbf{B}_n^\beta(s))^T J_n^{-1}\mathbb{P}_n\left\{\Gamma(\theta_0, W)\right\}}{\sqrt{(\mathbf{B}_n^\beta(s))^T J_n^{-1}\mathbf{B}_n^\beta(s)}}\right| = o_P(1). \tag{S9.32}
$$

Define $U_n = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} J_n^{-1/2}\Gamma(\theta_0, W_i) = \sqrt{n}J_n^{-1/2}\mathbb{P}_n\left\{\Gamma(\theta_0, W)\right\}$. Noted that for any $i \in \{1, \dots, n\}$, $J_n^{-1/2}\Gamma(\theta_0, W_i)$ is a $q_{n,1}$ dimensional random vector with zero mean and identity covariance. Moreover,

$$\mathbb{E}\left|J_n^{-1/2}\Gamma(\theta_0, W_i)\right|^3 = O(n^{2\nu_1}).$$

From Yurinskii's coupling (Yurinskii, 1978) for sums of random vectors, there exists $V_n \sim \mathcal{N}(0, I)$, such that

$$P\left(|U_n - V_n| > 3/\log n\right) = O\left((\log n)^3 n^{3\nu - 1/2}\right),$$

which indicates $U_n \xrightarrow{P} V_n$. Denote

$$G_n(s) = \frac{(\mathbf{B}_n^{\beta}(s))^T J_n^{-1/2} V_n}{\sqrt{(\mathbf{B}_n^{\beta}(s))^T J_n^{-1}\mathbf{B}_n^{\beta}(s)}}, s \in [0, 1].$$

It follows from (S9.32) that

$$\sup_{s\in(0,1)}\left|\frac{\sqrt{n}\left(\hat{\beta}_n(s) - \beta_0(s)\right)}{\sqrt{(\mathbf{B}_n^{\beta}(s))^T J_n^{-1}\mathbf{B}_n^{\beta}(s)}} - G_n\right| = o_P(1).$$

Noted that G_n is a Gaussian process on $[0, 1]$ with $E[G_n(s)] = 0$ and $\mathrm{Var}[G_n(s)] = 1$ and covariance

$$G_n(s,t) = E[G_n(s)G_n(t)] = \frac{(\mathbf{B}_n^{\beta}(s))^T J_n^{-1}\mathbf{B}_n^{\beta}(t)}{\sqrt{(\mathbf{B}_n^{\beta}(s))^T J_n^{-1}\mathbf{B}_n^{\beta}(s)}\sqrt{(\mathbf{B}_n^{\beta}(t))^T J_n^{-1}\mathbf{B}_n^{\beta}(t)}},$$

$\forall s, t \in [0, 1]$, which completes the proof.

Bibliography

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.

Horowitz, J. L. (1996). Semiparametric Estimation of a Regression Model with an Unknown Transformation of the Dependent Variable. *Econometrica*, 64:103–117.

Liu, C., Su, W., Liu, K.-Y., Yin, G., and Zhao, X. (2024). Efficient estimation for functional accelerated failure time model.

Qu, S., Wang, J. L., and Wang, X. (2016). Optimal Estimation for the Functional Cox Model. *The Annals of Statistics*, 44:1708–1738.

Ritter, K., Wasilkowski, G. W., and Wozniakowski, H. (1995). Multivariate Integration and Approximation for Random Fields Satisfying Sacks-Ylvisaker Conditions. *The Annals of Applied Probability*, 5:518–540.

Schumaker, L. (2007). *Spline Functions: Basic Theory (3rd Edition)*. Cambridge University Press, Cambridge Mathematical Library.

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

Tang, W., He, K., Xu, G., and Zhu, J. (2022). Survival Analysis via Ordinary Differential Equations. *Journal of the American Statistical Association*, 118:2406–2421.

Van der Vaart, A. W. and Wellner, J. A. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, New York.

Walter, W. (1998). *Ordinary Differential Equations*, volume 182 of *Graduate Texts in Mathematics*. Springer New York, New York.

Yurinskii, V. V. (1978). On the Error of the Gaussian Approximation for Convolutions. *Theory of Probability & Its Applications*, 22:236–247.