REGRESSION ANALYSIS OF PANEL COUNT DATA WITH BOTH TIME-DEPENDENT COVARIATES AND TIME-VARYING EFFECTS

Yuanyuan Guo[1*], Dayu Sun[2*] and Jianguo Sun[1]

[1] *University of Missouri and* [2] *Emory University*

Supplementary Material

This Supplementary Material contains three sections. First we will give some additional results from the simulation study and the real data analysis in the paper and then sketch the proof of the asymptotic results given in the paper.

S1 Additional Simulation Results

This section contains some additional simulation results on $\hat{\boldsymbol{\beta}}(t)$.

*Equal contribution.

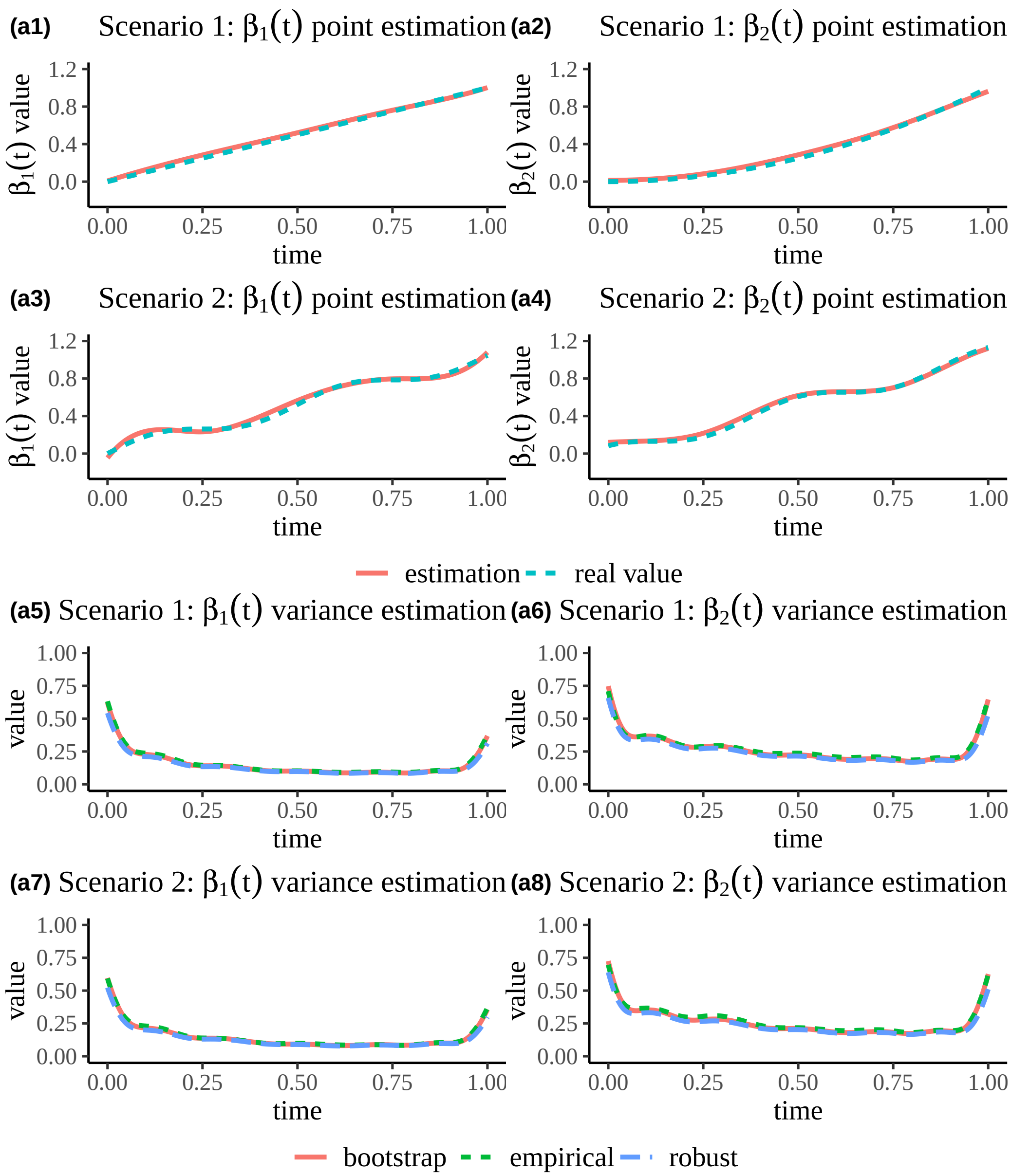

Figure 1: Simulation results on estimation of $\beta_1(t)$ and $\beta_2(t)$ under the Poisson process with $\Lambda_0(t) = 2t + 3$ and dependent covariates. (a1) - (a4): on estimation of $\beta_1(t)$ and $\beta_2(t)$; (a5) - (a8): on variance estimation of $\hat{\beta}_1(t)$ and $\hat{\beta}_2(t)$.

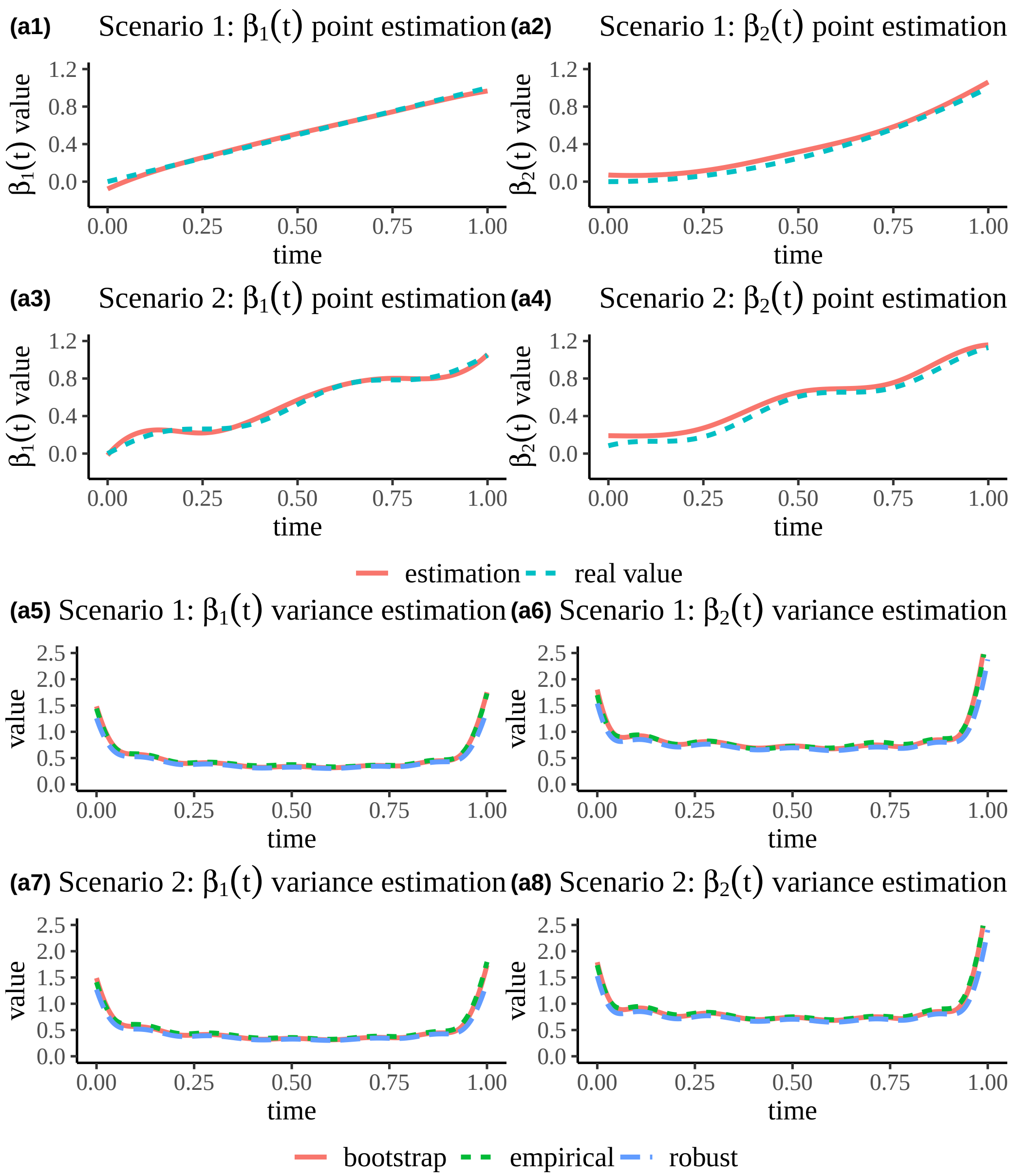

Figure 2: Simulation results on estimation of $\beta_1(t)$ and $\beta_2(t)$ under the Non-Poisson process with $\Lambda_0(t) = 2t + 3$ and dependent covariates. (a1) - (a4): on estimation of $\beta_1(t)$ and $\beta_2(t)$; (a5) - (a8): on variance estimation of $\hat{\beta}_1(t)$ and $\hat{\beta}_2(t)$.

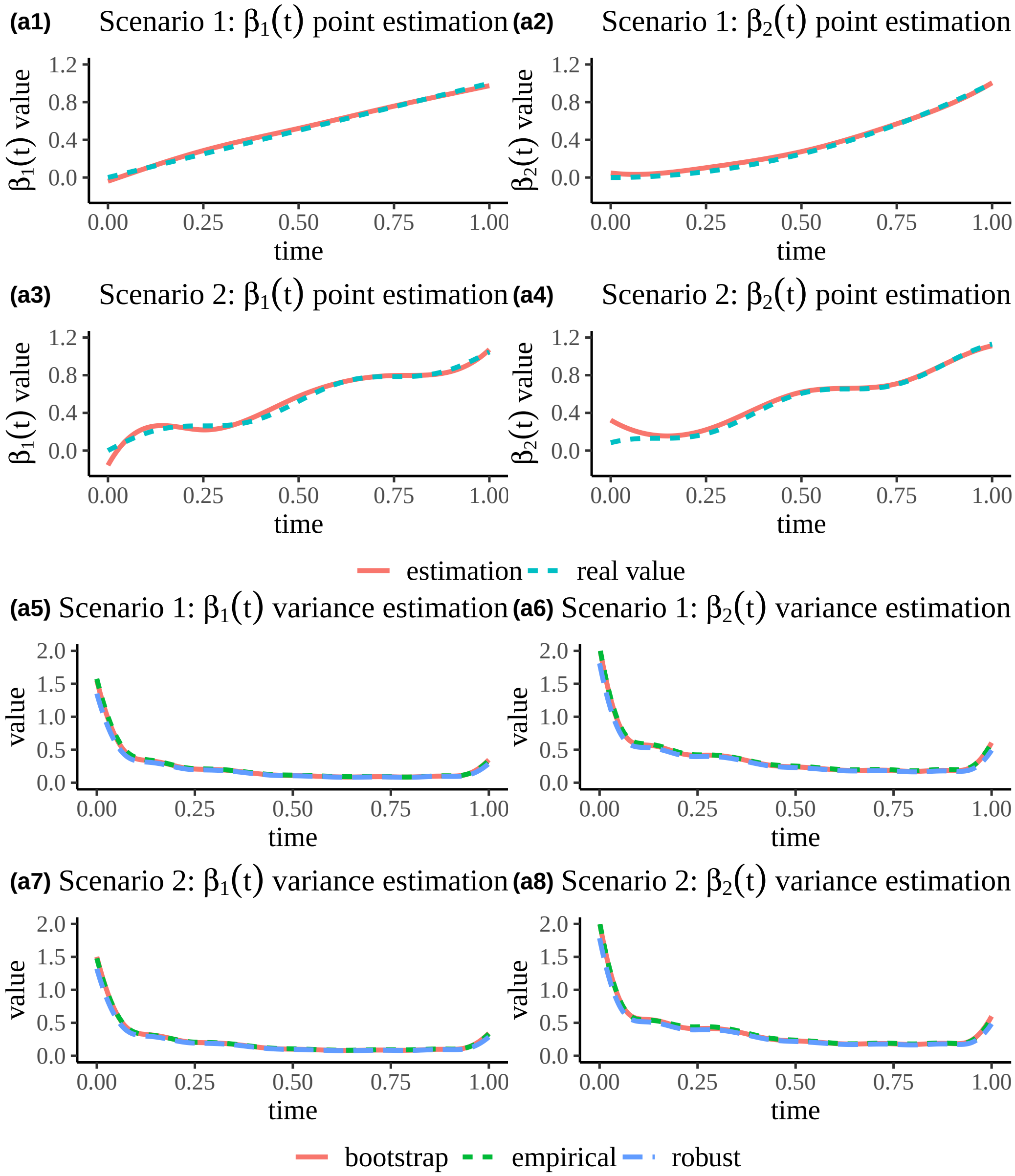

Figure 3: Simulation results on estimation of $\beta_1(t)$ and $\beta_2(t)$ under the Poisson process with $\Lambda_0(t) = (\sin(4\pi t) + 4\pi t)/2$ and dependent covariates. (a1) - (a4): on estimation of $\beta_1(t)$ and $\beta_2(t)$; (a5) - (a8): on variance estimation of $\hat{\beta}_1(t)$ and $\hat{\beta}_2(t)$.

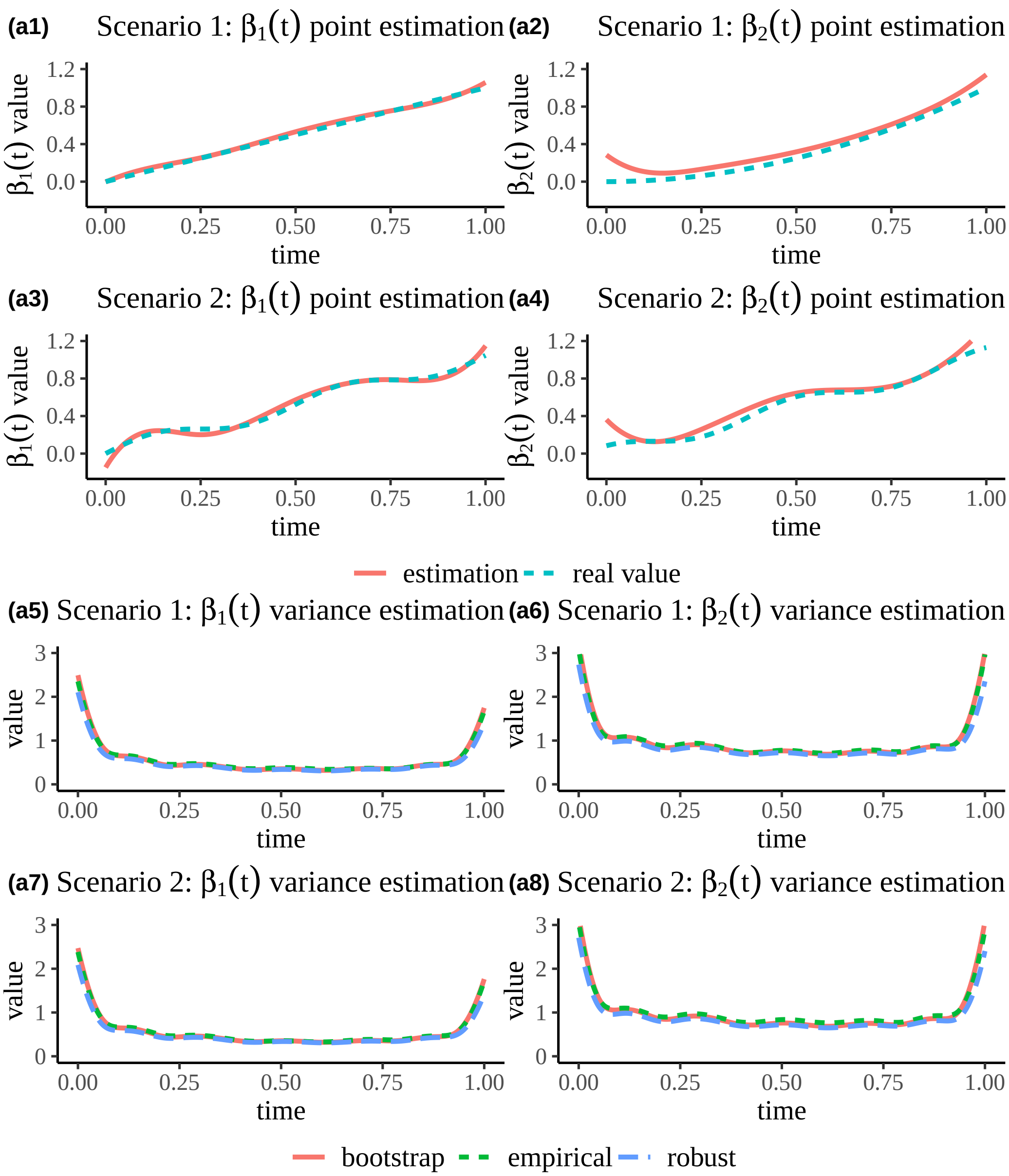

Figure 4: Simulation results on estimation of $\beta_1(t)$ and $\beta_2(t)$ under the Non-Poisson process with $\Lambda_0(t) = (\sin(4\pi t) + 4\pi t)/2$ and dependent covariates. (a1) - (a4): on estimation of $\beta_1(t)$ and $\beta_2(t)$; (a5) - (a8): on variance estimation of $\hat{\beta}_1(t)$ and $\hat{\beta}_2(t)$.

S2 Additional Real Data Analysis Results

This section includes some additional results for the real data analysis with the number of interior knots being 5 and 7.

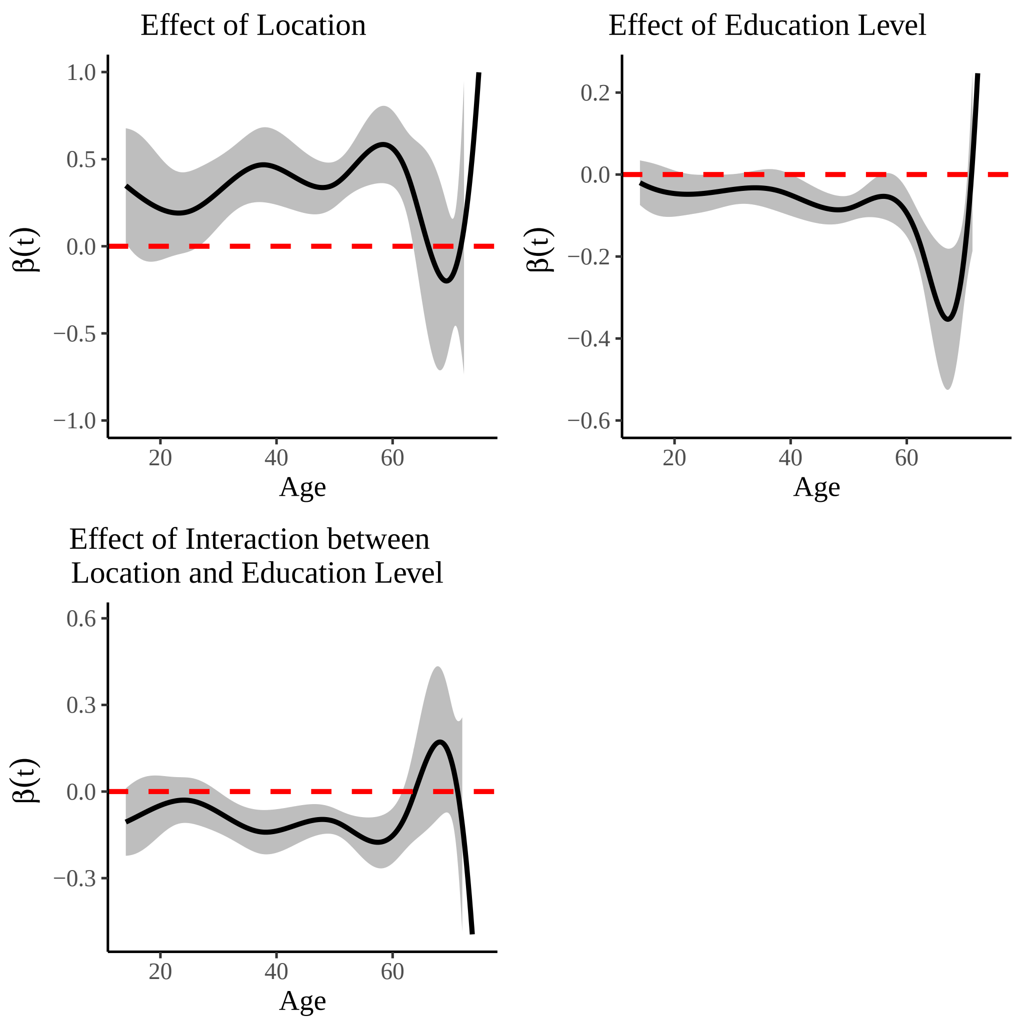

Figure 5: Estimated time-varying effects of the location, education level and the interaction between them (solid curves) and corresponding pointwise 95% confidence intervals (ribbons) with 5 interior knots.

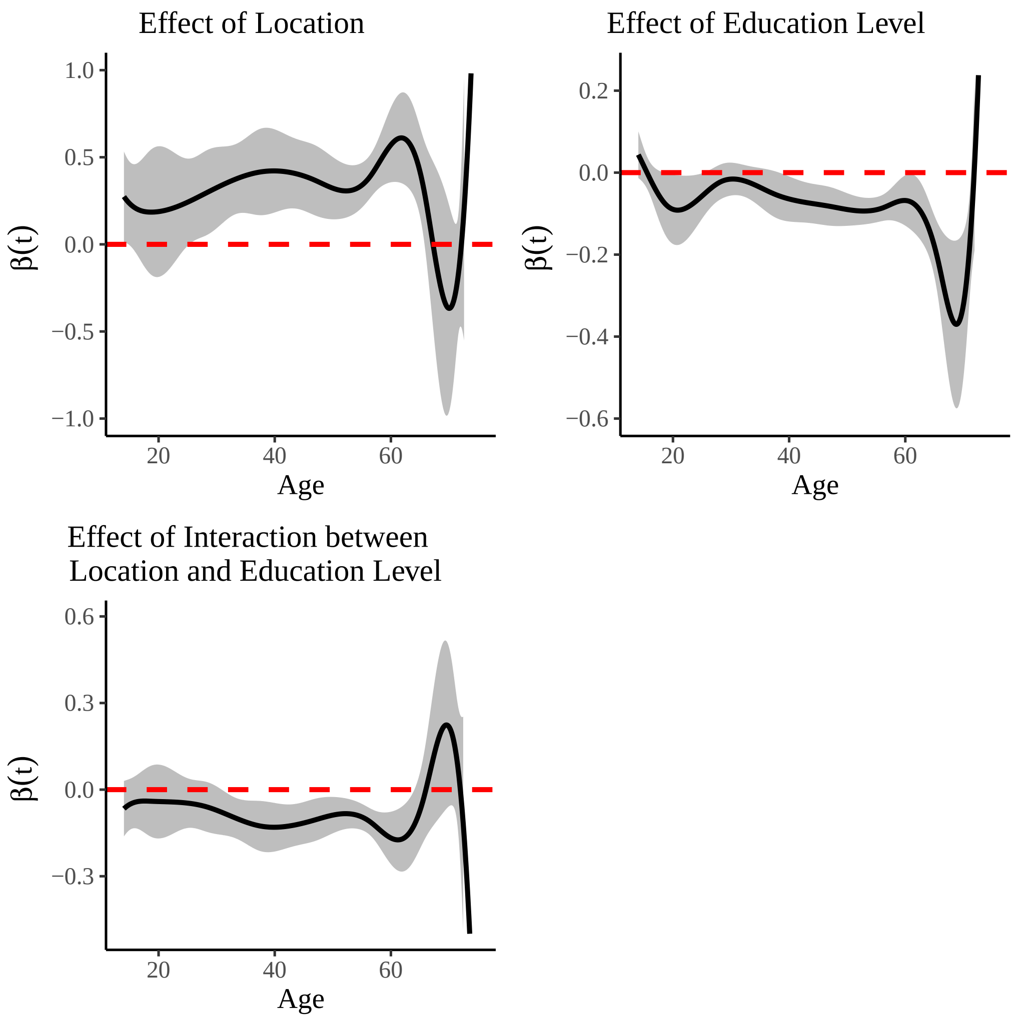

Figure 6: Estimated time-varying effects of the location, education level and the interaction between them (solid curves) and corresponding pointwise 95% confidence intervals (ribbons) with 7 interior knots.

S3 The proof of the asymptotic properties

In this section, we will first introduce some additional notation and preliminary results needed for the proof and then sketch the proofs of the consistency, the rate of convergence and the asymptotic normality discussed in Sections S3.2, S3.3 and S3.4, respectively.

S3.1 The preliminary results

Note that it is not straightforward to study the asymptotic properties of $\hat{\boldsymbol{\vartheta}}_n$ based on the sieve estimating equation (2.3). In this part, we will introduce some notation and show that solving (2.3) is equivalent to maximize pseudo-likelihood function,

$$l_n(\boldsymbol{\vartheta}) = \frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)\left(N_i(t)\log(\Lambda(t)) + N_i(t)\boldsymbol{\theta}^T(t)\mathbf{V}_i(t) - \exp\left(\boldsymbol{\theta}^T(t)\mathbf{V}_i(t)\right)\Lambda(t)\right)dH_i(t) \tag{S3.1}$$

with respect to $\boldsymbol{\vartheta}$ over $\Theta_n = \mathcal{A}\times\mathcal{M}_n\times\mathcal{F}$. Here, with slight abuse of notation, $\boldsymbol{\theta}^T(t)\mathbf{V}_i(t) = \boldsymbol{\gamma}^T\mathbf{W}_i(t) + \boldsymbol{\beta}^T(t)\mathbf{Z}_i(t)$ and $\boldsymbol{\theta}(t) = (\boldsymbol{\gamma}, \boldsymbol{\beta}(t))$. Let $\mathbb{P}_n$ be the empirical measure and $\mathbf{P}$ be the true probability measure. Let $\mathbb{M}_n(\boldsymbol{\vartheta}) = l_n(\boldsymbol{\vartheta}) = \mathbb{P}_n m_{\boldsymbol{\vartheta}}(\mathbf{V})$ and $\mathbf{M}(\boldsymbol{\vartheta}) = \mathbf{P}m_{\boldsymbol{\vartheta}}(\mathbf{V})$, where

$$m_{\boldsymbol{\vartheta}}(O) = \int_0^{\tau} Y(t)\left(N(t)\log(\Lambda(t)) + N(t)\boldsymbol{\theta}^T(t)\mathbf{V}(t) - \exp\left(\boldsymbol{\theta}^T(t)\mathbf{V}(t)\right)\Lambda(t)\right)dH(t)$$

and $O = (\mathbf{V}, H, Y)$.

Also, let

$$\hat{\Lambda}(t, \boldsymbol{\theta}(t)) = \frac{\sum_{j=1}^{n} Y_j(t)N_j(t)dH_j(t)}{\sum_{j=1}^{n} Y_j(t)\exp\left(\boldsymbol{\theta}^T(t)\mathbf{V}_j(t)\right)dH_j(t)}.$$

We first show $\hat{\Lambda}(t,\boldsymbol{\theta}(t))$ maximizes (S3.1) for a fixed $\boldsymbol{\theta}(t)$. After some algebra,

$$
\begin{aligned}
&\mathbb{M}_n(\boldsymbol{\theta},\Lambda)-\mathbb{M}_n\left(\boldsymbol{\theta},\hat{\Lambda}\right)\\
=&\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}Y_i(t)\hat{\Lambda}(t,\boldsymbol{\theta}(t))\exp\left(\boldsymbol{\theta}^T(t)\mathbf{V}_i(t)\right)\\
&\times\left(\frac{N_i(t)}{\hat{\Lambda}(t,\boldsymbol{\theta}(t))\exp\left(\boldsymbol{\theta}^T(t)\mathbf{V}_i(t)\right)}\log\left(\frac{\Lambda(t)}{\hat{\Lambda}(t,\boldsymbol{\theta}(t))}\right)-\frac{\Lambda(t)}{\hat{\Lambda}(t,\boldsymbol{\theta}(t))}+1\right)dH_i(t).
\end{aligned}
$$

Plug in $\hat{\Lambda}(t,\boldsymbol{\theta}(t))$, and by Fubini's theorem, we have

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}Y_i(t)\hat{\Lambda}(t,\boldsymbol{\theta}(t))\exp\left(\boldsymbol{\theta}^T(t)\mathbf{V}_i(t)\right)\left(\frac{\Lambda(t)}{\hat{\Lambda}(t,\boldsymbol{\theta}(t))}-1\right)dH_i(t)\\
=&\frac{1}{n}\int_0^{\tau}\left(1-\frac{\Lambda(t)}{\hat{\Lambda}(t,\boldsymbol{\theta}(t))}\right)\left(\sum_{j=1}^{n}Y_j(t)N_j(t)dH_j(t)\right)\sum_{i=1}^{n}\frac{Y_i(t)\exp\left(\boldsymbol{\theta}^T(t)\mathbf{V}_i(t)\right)dH_i(t)}{\sum_{j=1}^{n}Y_j(t)\exp\left(\boldsymbol{\theta}^T(t)\mathbf{V}_i(t)\right)dH_j(t)}\\
=&\frac{1}{n}\int_0^{\tau}\sum_{j=1}^{n}Y_j(t)N_j(t)\left(1-\frac{\Lambda(t)}{\hat{\Lambda}(t,\boldsymbol{\theta}(t))}\right)dH_j(t).
\end{aligned}
$$

Therefore, since $\log(x)-x+1\leq-(x-1)^2$ for all positive x and the equality holds iff $x=1$,

$$
\begin{aligned}
&\mathbb{M}_n(\boldsymbol{\theta},\Lambda)-\mathbb{M}_n\left(\boldsymbol{\theta},\hat{\Lambda}\right)\\
=&\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}Y_i(t)N_i(t)\left\{\log\left(\frac{\Lambda(t)}{\hat{\Lambda}(t,\boldsymbol{\theta}(t))}\right)-\frac{\Lambda(t)}{\hat{\Lambda}(t)}+1\right\}dH_i(t)\\
\leq&-\frac{1}{n}\sum_{i=1}^{n}\int_0^{\tau}Y_i(t)N_i(t)\left(\frac{\Lambda(t)}{\hat{\Lambda}(t,\boldsymbol{\theta}(t))}-1\right)^2dH_i(t)\\
\leq&0.
\end{aligned}
$$

This implies $\mathbb{M}_n(\boldsymbol{\theta}, \Lambda) \leq \mathbb{M}_n\left(\boldsymbol{\theta}, \hat{\Lambda}\right)$ for any $\boldsymbol{\theta}$. The equality holds iff $\Lambda(t) = \hat{\Lambda}(t, \boldsymbol{\theta}(t))$ at points where $\sum_{i=1}^n Y_i(t) H_i(t)$ jumps. Since $l_n(\boldsymbol{\vartheta})$ is only determined by the value of $\hat{\Lambda}(t, \boldsymbol{\theta}(t))$ at points where $\sum_{i=1}^n Y_i(t) H_i(t)$ jumps, $\hat{\Lambda}(t, \boldsymbol{\theta}(t))$ is the unique maximizer of $l_n(\boldsymbol{\theta}, \Lambda)$ with respect to Λ. Then, to maximize $l_n\left(\boldsymbol{\theta}, \hat{\Lambda}(t, \boldsymbol{\theta})\right)$ with respect to $\boldsymbol{\theta}$ over $\mathcal{A} \times \mathcal{M}_n$, by the idea of profile likelihood in Wellner and Zhang (2007), we need to maximize $l_n\left(\boldsymbol{\theta}_n, \hat{\Lambda}(t, \boldsymbol{\theta}_n)\right)$ with respect to $\boldsymbol{\theta}_n$. After some algebra, $\partial l_n\left(\boldsymbol{\theta}_n, \hat{\Lambda}(t, \boldsymbol{\theta}_n)\right) / \partial \boldsymbol{\theta}_n$ equals the left hand side of (2.3). Obviously, $l_n\left(\boldsymbol{\theta}_n, \hat{\Lambda}(t, \boldsymbol{\theta}_n)\right)$ is convex with respect to $\boldsymbol{\theta}_n$, implying maximizing $l_n(\boldsymbol{\vartheta})$ over $\mathcal{A} \times \mathcal{M}_n \times \mathcal{F}$ is equivalent to solving (2.3).

After showing the equivalence of solving the estimating equation and maximizing $l_n(\boldsymbol{\vartheta})$, the estimator is actually an M-estimator and its (asymptotic) behavior can be investigated through $m_\theta(\mathbf{O})$ with the modern empirical process theory. Moreover, (S3.1) coincides with the pseudo-likelihood function for panel count data proposed in Wellner and Zhang (2007) which has been extensively investigated, similar to He et al. (2017). We can then use many conclusions in the existing literature to facilitate our theoretical justification.

S3.2 Proof of the consistency

To prove the consistency of $\hat{\boldsymbol{\theta}}_n$, we will employ Theorem 3.1 and Remark 3.1 in Chen (2007). First we will show that $\boldsymbol{\vartheta}_0$ is the unique maximizer of $\mathbf{M}(\boldsymbol{\vartheta})$. After some calculation based on the conditional expectation on $\mathbf{V}(t)$, we have

$$\mathbf{M}(\boldsymbol{\vartheta}) = \int \exp\left(\boldsymbol{\theta}^T(t)\, v(t)\right) \Lambda(t) \left\{\log(\Lambda(t)) + \boldsymbol{\theta}^T(t)\, v(t) - 1\right\} dv_1(t, v),$$

and therefore

$$\mathbf{M}(\boldsymbol{\vartheta}_0) - \mathbf{M}(\boldsymbol{\vartheta}) = \int \Lambda(u) \exp\left(\boldsymbol{\theta}^T(u)\, v(u)\right) h \left\{ \frac{\Lambda_0(u) \exp\left(\boldsymbol{\theta}_0^T(u)\, v(u)\right)}{\Lambda(u) \exp\left(\boldsymbol{\theta}^T(u)\, v(u)\right)} \right\} dv_1(t, v)\ , \tag{S3.2}$$

where $h(x) = x \log(x) - x + 1$. Note that $h(x) \geq 0$ for all $x > 0$ and the equality holds only when $x = 1$. Therefore, by similar argument in Wellner and Zhang (2007), under conditions (C2) and (C8), $\mathbf{M}(\boldsymbol{\vartheta}_0) = \mathbf{M}(\boldsymbol{\vartheta})$ if and only if $\boldsymbol{\theta}(t) = \theta_0(t)$ and $\Lambda(t) = \Lambda_0(t)$ a.e. with respect to μ_1. In this manner, $\boldsymbol{\vartheta}_0$ is the unique maximizer of $\mathbf{M}(\boldsymbol{\vartheta})$ on μ_1.

By the similar arguments used in Wellner and Zhang (2007) and conditions (C1)-(C5), we can show that $\hat{\Lambda}(t)$ is uniformly bounded in probability for $t \in [0, \tau]$ by $\mu_1(\{\tau\}) > 0$ in (C2).

By Helly-Selection Theorem and compactness of Θ_n, it follows that $\hat{\boldsymbol{\vartheta}}_n = \left(\hat{\gamma}, \hat{\boldsymbol{\beta}}_n, \hat{\Lambda}\right)$ has a subsequence $\hat{\boldsymbol{\vartheta}}_{n_k}$ converging to $\boldsymbol{\vartheta}^+ = \left(\boldsymbol{\gamma}^+, \boldsymbol{\beta}^+, \Lambda^+\right)$ with $\boldsymbol{\vartheta}^+ \in \Theta$. Obviously, under (C3)-(C6) we have the compactness of Θ_n as well as the fact that $m_{\boldsymbol{\vartheta}}(O)$ is upper semicontinuous in $\boldsymbol{\vartheta}$ for almost all O. Furthermore, $m_{\boldsymbol{\vartheta}} \leq M_0 < \infty$ with $\mathbf{P}M_0(\mathbf{V}) < \infty$ by (C9). Thus, by Theorem A.1 of Wellner and Zhang (2007), we have

$$\limsup_{n\to\infty}\sup_{\boldsymbol{\vartheta}\in\Theta_n}\left(\mathbb{P}_n-\mathbf{P}\right)m_{\boldsymbol{\vartheta}}\left(\mathbf{V}\right)\leq 0 \tag{S3.3}$$

almost surely. By the Dominated Convergence Theorem and (C9), $\mathbf{M}\left(\boldsymbol{\vartheta}\right)$ is continuous in $\boldsymbol{\vartheta}$. By the Corollary 6.21 of Schumaker (2007), there exists a spline approximation $\beta_{n0j}\left(t\right)\in\mathcal{M}_{nj}$ to β_{0j} such that

$$\sup_{t\in[0,\tau]}\left|\beta_{0j}\left(t\right)-\beta_{n0j}\left(t\right)\right|=O\left(K_n^{-r}\right)=O\left(n^{-vr}\right) \tag{S3.4}$$

for $j=1,\ldots,p_2$. Therefore, for any $\epsilon>0$, there exists $\boldsymbol{\beta}_0^*\in\mathcal{M}_n$ such that

$$\mathbf{M}\left(\boldsymbol{\vartheta}_0\right)-\epsilon\leq\mathbf{M}\left(\boldsymbol{\gamma}_0,\boldsymbol{\beta}_0^*,\Lambda_0\right)$$

with $\max_{j=1,\ldots,p_2}\left\|\beta_{0j}\left(t\right)-\beta_{0j}^*\left(t\right)\right\|_\infty=o\left(1\right)$. Also, by the similar argument in Lu et al. (2009), we have

$$\mathbb{M}_n\left(\boldsymbol{\gamma}_0,\boldsymbol{\beta}_0^*,\Lambda_0\right)-\mathbf{M}\left(\boldsymbol{\gamma}_0,\boldsymbol{\beta}_0^*,\Lambda_0\right)=o_p\left(1\right)$$

and

$$\mathbb{M}_n\left(\boldsymbol{\gamma}_0,\boldsymbol{\beta}_0^*,\Lambda_0\right)\leq\mathbb{M}_n\left(\hat{\boldsymbol{\gamma}},\hat{\boldsymbol{\beta}}_n,\hat{\Lambda}\right).$$

Then by (S3.3) and the arguments similar to those used in Lu et al. (2009), we can show that $\mathbf{M}\left(\boldsymbol{\vartheta}_0\right)=\mathbf{M}\left(\boldsymbol{\vartheta}^+\right)$, implying $\boldsymbol{\beta}^+=\boldsymbol{\beta}_0$ and $\Lambda^+=\Lambda_0$ a.e. with respect to

μ_1. Since this holds for any convergent subsequence, we conclude that all the limits of subsequence of $\hat{\boldsymbol{\vartheta}}_{n_k}$ are $\boldsymbol{\vartheta}_0$. Therefore, due to the uniform boundedness of $\hat{\Lambda}(t)$ and Dominated Convergence Theorem, we obtain the weak consistency of $\hat{\boldsymbol{\vartheta}}_n$ in the metric d.

S3.3 Proof of the rate of convergence

In (S3.2), since $h(x) \geq (1/4)(x-1)^2$ for $0 \leq x \leq 5$, for $\boldsymbol{\theta}$ in a sufficiently small neighborhood of θ_0 and some constant c,

$$\begin{aligned}\mathbf{M}(\boldsymbol{\vartheta}_0) - \mathbf{M}(\boldsymbol{\vartheta}) &\geq \frac{1}{4}\int \Lambda(u)\exp\left(\boldsymbol{\theta}^T(u)\,v(u)\right)\left\{\frac{\Lambda_0(u)\exp\left(\boldsymbol{\theta}_0^T(u)\,v(u)\right)}{\Lambda(u)\exp\left(\boldsymbol{\theta}^T(u)\,v(u)\right)} - 1\right\}^2 dv_1(u,v) \\ &\geq c\int \left\{\Lambda(u)\exp\left(\boldsymbol{\theta}^T(u)\,v(u)\right) - \Lambda_0(u)\exp\left(\boldsymbol{\theta}_0^T(u)\,v(u)\right)\right\}^2 dv_1(u,v)\ .\end{aligned} \tag{S3.1}$$

Note that c used through the paper represents some constant and could be different in different contexts. Let $\rho(u,z) = \Lambda(u)\exp\left(\boldsymbol{\beta}^T(u)\,z(u)\right)$ and $\rho_0(u,z) = \Lambda_0(u)\exp\left(\boldsymbol{\beta}_0^T(u)\,z(u)\right)$. We also define $\rho_s = t\rho + (1-t)\rho_0$, $\Lambda_s = s\Lambda + (1-s)\Lambda_0$, $\boldsymbol{\gamma}_s = s\boldsymbol{\gamma} + (1-s)\boldsymbol{\gamma}_s$, $\boldsymbol{\beta}_s = s\boldsymbol{\beta} + (1-s)\boldsymbol{\beta}_s$, $\boldsymbol{\theta}_s = s\boldsymbol{\theta} + (1-s)\boldsymbol{\theta}_s$ for $s \in (0,1)$. Also let

$$g(s) = \rho_s(U, \mathbf{Z})\exp\left(\boldsymbol{\gamma}_s^T \mathbf{W}(U)\right).$$

It is easy to see that

$$\Lambda(U)\exp\left(\boldsymbol{\theta}^T(U)\,\mathbf{V}(U)\right) - \Lambda_0(U)\exp\left(\boldsymbol{\theta}_0^T(U)\,\mathbf{V}(U)\right) = g(1) - g(0).$$

By the mean value theorem, there exists $0 \le \xi \le 1$ such that $g(1) - g(0) = g'(\xi)$ where

$$\begin{aligned} g'(\xi) &= \exp\left(\boldsymbol{\gamma}_\xi^T \mathbf{W}(U)\right)\left\{(\rho-\rho_0)(U,\mathbf{Z}) + (\xi\rho + (1-\xi)\rho_0)(U,\mathbf{Z})(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)^T\mathbf{W}(U)\right\} \\ &= \exp\left(\boldsymbol{\gamma}_\xi^T \mathbf{W}(U)\right)\left\{(\rho-\rho_0)(U,\mathbf{Z})\left\{1+\xi(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)^T\mathbf{W}(U)\right\} + (\xi\rho + (1-\xi)\rho_0)(U,\mathbf{Z})\right\}. \end{aligned}$$

From (S3.1), we have that, for some constant $c > 0$,

$$\begin{aligned} &\mathbf{M}(\boldsymbol{\vartheta}_0) - \mathbf{M}(\boldsymbol{\vartheta}) \\ \ge& c\int\left\{(\rho-\rho_0)(u,z)\left\{1+\xi(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)^T w(U)\right\} + (\xi\rho + (1-\xi)\rho_0)(u,z)\right\}^2 d\nu_1(u,z,w) \\ =& c\nu_1\{g_1 h + g_2\}^2, \end{aligned}$$

where $g_1(U,\mathbf{V}) = \left((\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)^T\mathbf{W}(U)\right)\rho_0(U,\mathbf{Z})$, $g_2(U,\mathbf{Z}) = (\rho-\rho_0)(U,\mathbf{Z})$ and $h(U,\mathbf{Z}) = 1+\xi(\rho-\rho_0)(U,\mathbf{Z})/\rho_0(U,\mathbf{Z})$. By the similar method in Wellner and Zhang (2007) and He et al. (2017), under condition (C12), for some constants $c, c_1 > 0$,

$$\begin{aligned}\mathbf{M}(\boldsymbol{\vartheta}_0) - \mathbf{M}(\boldsymbol{\vartheta}) &\geq c\nu_1 \{g_1 h + g_2\}^2 \\ &\geq c_1 \left\{\nu_1\left(g_1^2\right) + \nu_1\left(g_2^2\right)\right\}.\end{aligned}$$

Similarly, by the mean value theorem and condition (C12), $\exists c_2 > 0$,

$$\begin{aligned}\nu_1\left(g_2^2\right) &= \nu_1\left((h_2 g_3 + g_4)^2\right) \\ &\geq c_2 \left\{\nu_1\left(g_3^2\right) + \nu_1\left(g_4^2\right)\right\},\end{aligned}$$

where $g_3(U, \mathbf{Z}) = \left((\boldsymbol{\beta} - \boldsymbol{\beta}_0)^T(U)\,\mathbf{Z}(U)\right)\Lambda_0(U)$, $g_4(U) = (\Lambda - \Lambda_0)(U)$ and $h_2(U) = 1 + \zeta(\Lambda - \Lambda_0)(U)/\Lambda_0(U)$ for some $\zeta \in (0, 1)$. Therefore, under (C1) and (C8), there exist some positive constants c_3 and c_3^*,

$$\begin{aligned}\mathbf{M}(\boldsymbol{\vartheta}_0) - \mathbf{M}(\boldsymbol{\vartheta}) &\geq c_3 \left\{\nu_1\left(g_1^2\right) + \nu_1\left(g_3^2\right) + \nu_1\left(g_4^2\right)\right\} \\ &\geq c_3^* \left(\|\boldsymbol{\gamma} - \gamma_0\|_2^2 + \int \|\boldsymbol{\beta}(u) - \boldsymbol{\beta}_0(u)\|_2^2\, d\mu_1(u) + \|\Lambda - \Lambda_0\|_{L_2(\mu_1)}^2\right) \\ &\gtrsim d(\boldsymbol{\vartheta}_0, \boldsymbol{\vartheta}).\end{aligned}$$

Next, we need to find $\varphi_n(\delta)$ such that

$$E\left[\sup_{d(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_0) < \delta} \sqrt{n}\,\left|(\mathbb{P}_n - \mathbf{P})\left(m_{\boldsymbol{\vartheta}}(O) - m_{\boldsymbol{\vartheta}_0}(O)\right)\right|\right] \leq c\varphi_n(\delta).$$

Let

$$\mathcal{F}_\delta = \left\{ m_{\boldsymbol{\vartheta}}(O) - m_{\boldsymbol{\vartheta}_0}(O) : d\left(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_0\right) \leq \delta \right\} .$$

From the result of Theorem 2.7.5 of Vaart and Wellner (1996) and Lemma A.2 of Lu et al. (2009), for any $\epsilon \leq \delta$, we have

$$\log N_{[]}\left(\epsilon, \mathcal{F}_\delta, \|\cdot\|_{\mathbf{P},B}\right) \leq c\left(\frac{1}{\epsilon} + (p_1 + p_2 q_n)\log\left(\frac{\delta}{\epsilon}\right)\right),$$

where $\|\cdot\|_{P,B}$ is the Bernstein norm defined as $\|f\|_{P,B} = \left\{2\mathbf{P}\left(e^{|f|} - 1 - |f|\right)\right\}^{1/2}$ by Vaart and Wellner (1996, page 324). Similar to the argument in Wellner and Zhang (2007); Lu et al. (2009), under conditions (C6) and (C10), we have

$$\|m_{\boldsymbol{\vartheta}}(O) - m_{\boldsymbol{\vartheta}_0}(O)\|^2_{\mathbf{P},B} \leq c\delta^2,$$

for any $m_{\boldsymbol{\vartheta}}(O) - m_{\boldsymbol{\vartheta}_0}(O) \in \mathcal{F}_\delta$. Therefore, by Lemma 3.4.3 in Vaart and Wellner (1996), we can show a maximal inequality

$$E\left\|\sqrt{n}\left(\mathbb{P}_n - \mathbf{P}\right)\right\|_{\mathcal{F}_\delta} \leq cJ_{[]}\left(\delta, \mathcal{F}_\delta, \|\cdot\|_{\mathbf{P},B}\right)\left\{1 + \frac{J_{[]}\left(\delta, \mathcal{F}_\delta, \|\cdot\|_{\mathbf{P},B}\right)}{\delta^2 n^{1/2}}\right\}$$

where

$$
\begin{aligned}
J_{[]}\left(\delta, \mathcal{F}_\delta, \|\cdot\|_{\mathbf{P},B}\right) &= \int_0^\delta \left\{1 + \log N_{[]}\left(\epsilon, \mathcal{F}_\delta, \|\cdot\|_{\mathbf{P},B}\right)\right\}^{1/2} d\epsilon \\
&\leq c_1 q_n^{1/2} \int_0^\delta \left\{1 + \frac{1}{\epsilon} + \log\left(\frac{\delta}{\epsilon}\right)\right\}^{1/2} d\epsilon \\
&\leq q_n^{1/2}\delta^{1/2}
\end{aligned}
$$

Thus,

$$
\varphi_n(\delta) = q_n^{\frac{1}{2}}\delta^{\frac{1}{2}}\left(1 + \frac{q_n^{1/2}\delta^{1/2}}{\delta^2 n^{1/2}}\right) = q_n^{\frac{1}{2}}\delta^{\frac{1}{2}} + \frac{q_n}{\delta n^{1/2}}.
$$

It is not hard to show that $\varphi_n(\delta)/\delta$ is decreasing in δ and therefore

$$
a_n^2\varphi_n\left(\frac{1}{a_n}\right) = a_n^{3/2}q_n^{1/2} + a_n^3 q_n n^{-1/2} \lesssim n^{1/2}
$$

if $a_n = \min\left\{n^{\frac{1-\nu}{3}}, n^{r\nu}\right\}$ and $0 < \nu < 1/2$.

Moreover, using the similar argument in Lu et al. (2009), we can show $\mathbb{M}_n\left(\hat{\boldsymbol{\vartheta}}_n\right) - \mathbb{M}_n\left(\boldsymbol{\vartheta}_0\right) > -O_p\left(n^{-2r\nu}\right) \geq O_p\left(a_n^2\right)$. Then, by Theorem 3.2.5 of Wellner and Zhang (2007), we have $a_n d\left(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_0\right) = O_p(1)$. If ν is chosen as $1/\left(3r+1\right)$, we obtain the optimal rate $n^{r/(3r+1)}$ because $(1-\nu)/3 = r\nu$.

S3.4 Proof of the asymptotic normality

For the proof, we will mainly use the method in He et al. (2017) to derive the asymptotic normality. Define a sequence of maps S_n mapping a neighborhood of $\boldsymbol{\vartheta}_0$, denoted by $\mathcal{U}$, in the parameter space for $\boldsymbol{\vartheta}$ into $l^\infty\left(\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3\right)$ as

$$
\begin{aligned}
S_n\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right] =& \frac{d}{d\epsilon}\, l_n\left(\boldsymbol{\gamma} + \epsilon \boldsymbol{h}_1, \boldsymbol{\beta} + \epsilon \boldsymbol{h}_2, \Lambda + \epsilon h_3\right)\Big|_{\epsilon=0} \\
=& n^{-1} \sum_{i=1}^{n} \int_0^\tau Y_i(t) \left\{ \frac{N_i(t)}{\Lambda\left(t\right)} h_3\left(t\right) + N_i(t)\left(\boldsymbol{h}_1^T \mathbf{W}_i(t) + \boldsymbol{h}_2^T\left(t\right) \mathbf{Z}_i(t)\right) \right. \\
& - \left(\boldsymbol{h}_1^T \mathbf{W}_i(t) + \boldsymbol{h}_2^T\left(t\right) \mathbf{Z}_i(t)\right) \exp\left(\boldsymbol{\gamma}^T \mathbf{W}_i(t) + \boldsymbol{\beta}^T\left(t\right) \mathbf{Z}_i(t)\right) \Lambda\left(t\right) \\
& \left. - \exp\left(\boldsymbol{\gamma}^T \mathbf{W}_i(t) + \boldsymbol{\beta}^T \mathbf{Z}_i(t)\right) h_3\left(t\right) \right\} dH_i(t) \\
=& A_{n1}\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_1\right] + A_{n2}\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_2\right] + A_{n3}\left(\boldsymbol{\vartheta}\right)\left[h_3\right] \\
=& \mathbb{P}_n \psi\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right],
\end{aligned}
$$

where

$$
A_{n1}\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_1\right] = n^{-1} \sum_{i=1}^{n} \int_0^\tau Y_i(t) \boldsymbol{h}_1^T \mathbf{W}_i(t) \left\{ N_i(t) - \exp\left(\boldsymbol{\gamma}^T \mathbf{W}_i(t) + \boldsymbol{\beta}^T\left(t\right) \mathbf{Z}_i(t)\right) \Lambda\left(t\right) \right\} dH_i(t),
$$

$$
A_{n2}\left(\boldsymbol{\vartheta}\right)\left[\boldsymbol{h}_2\right] = n^{-1} \sum_{i=1}^{n} \int_0^\tau Y_i(t) \boldsymbol{h}_2^T\left(t\right) \mathbf{Z}_i(t) \left\{ N_i(t) - \exp\left(\boldsymbol{\gamma}^T \mathbf{W}_i(t) + \boldsymbol{\beta}^T\left(t\right) \mathbf{Z}_i(t)\right) \Lambda\left(t\right) \right\} dH_i(t),
$$

and

$$A_{n3}(\boldsymbol{\vartheta})[h_3] = n^{-1}\sum_{i=1}^{n}\int_0^{\tau} Y_i(t)h_3(t)\left\{\frac{N_i(t)}{\Lambda(t)} - \exp\left(\boldsymbol{\gamma}^T\mathbf{W}_i(t) + \boldsymbol{\beta}^T\mathbf{Z}_i(t)\right)\right\} dH_i(t).$$

Correspondingly, we define the limit map $S : \mathcal{U} \to l^{\infty}(\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3)$ as

$$S(\boldsymbol{\vartheta})[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] = A_1(\boldsymbol{\vartheta})[\boldsymbol{h}_1] + A_2(\boldsymbol{\vartheta})[\boldsymbol{h}_2] + A_3(\boldsymbol{\vartheta})[h_3]\ ,$$

where

$$A_1(\boldsymbol{\vartheta})[\boldsymbol{h}_1] = \mathbf{P}\int_0^{\tau} Y(t)\boldsymbol{h}_1^T\mathbf{W}(t)\left\{N(t) - \exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T(t)\mathbf{Z}(t)\right)\Lambda(t)\right\} dH(t),$$

$$A_2(\boldsymbol{\vartheta})[\boldsymbol{h}_2] = \mathbf{P}\int_0^{\tau} Y(t)\boldsymbol{h}_2^T(t)\mathbf{Z}(t)\left\{N(t) - \exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T(t)\mathbf{Z}(t)\right)\Lambda(t)\right\} dH(t),$$

and

$$A_3(\boldsymbol{\vartheta})[h_3] = \mathbf{P}\int_0^{\tau} Y(t)h_3(t)\left\{\frac{N(t)}{\Lambda(t)} - \exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T\mathbf{Z}(t)\right)\right\} dH(t).$$

To derive the asymptotic normality of $\hat{\boldsymbol{\vartheta}}_n$, we need to verify the following five conditions in He et al. (2017).

(a1) $\sqrt{n}(S_n - S)\left(\hat{\boldsymbol{\vartheta}}_n\right) - \sqrt{n}(S_n - S)(\boldsymbol{\vartheta}_0) = o_p(1)$.

(a2) $S(\boldsymbol{\vartheta}_0) = 0$ and $S_n\left(\hat{\boldsymbol{\vartheta}}_n\right) = o_p\left(n^{-1/2}\right)$.

(a3) $\sqrt{n}\left(S_n - S\right)(\theta_0)$ converges in distribution to a tight Gaussian process on $l^\infty\left(\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3\right)$.

(a4) $S(\boldsymbol{\vartheta})$ is Fréchet-differentiable at $\boldsymbol{\vartheta}_0$ denoted by $\dot{S}(\boldsymbol{\vartheta}_0)$.

(a5) $S\left(\hat{\boldsymbol{\vartheta}}_n\right) - S(\boldsymbol{\vartheta}_0) - \dot{S}(\boldsymbol{\vartheta}_0)\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) = o_p\left(n^{-1/2}\right)$.

By using similar argument in Lu et al. (2009), it is not hard to show

$$\left\{\psi(\boldsymbol{\vartheta})[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] - \psi(\boldsymbol{\vartheta}_0)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] : d(\boldsymbol{\vartheta}, \boldsymbol{\vartheta}_0) < \delta, (\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3\right\}$$

is a Donkser class for some δ. Therefore,

$$\sup_{(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) \in \mathcal{A} \times \mathcal{M} \times \mathcal{F}} \mathbf{P}\left\{\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] - \psi(\boldsymbol{\vartheta}_0)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3]\right\}^2 \to 0$$

as $d\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}_0\right) \to 0$ in probability and thus (a1) holds.

For (a2), clearly, $S(\boldsymbol{\vartheta}_0) = 0$. For $\boldsymbol{h}_2 \in \mathcal{H}_2$, let $\boldsymbol{h}_{2n}$ be the B-spline function approximation of $\boldsymbol{h}_2$ with $\max_{j=1,\ldots,p_2} \|h_{2j} - h_{2nj}\|_\infty = O\left(n^{-\nu r}\right)$ by (S3.4). Then we have $S_n\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_{3n}] = 0$. Thus,

$$\begin{aligned} S_n\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] &= \sqrt{n}\mathbb{P}_n\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] - \sqrt{n}\mathbf{P}\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_{3n}] \\ &= I_{n1} - I_{n2} + I_{n3} + I_{n4}\,, \end{aligned}$$

where

$$I_{n1} = \sqrt{n}\left(\mathbb{P}_n - \mathbf{P}\right)\left\{\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] - \psi(\boldsymbol{\vartheta}_0)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3]\right\}$$

$$I_{n2} = \sqrt{n}\left(\mathbb{P}_n - \mathbf{P}\right)\left\{\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_{2n}, h_3\right] - \psi\left(\boldsymbol{\vartheta}_0\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_{2n}, h_3\right]\right\}$$

$$I_{n3} = \sqrt{n}\mathbb{P}_n\left\{\psi\left(\boldsymbol{\vartheta}_0\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right] - \psi\left(\boldsymbol{\vartheta}_0\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_{2n}, h_3\right]\right\}$$

and

$$I_{n4} = \sqrt{n}\mathbf{P}\left\{\psi\left(\hat{\boldsymbol{\vartheta}}_n\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right] - \psi\left(\hat{\boldsymbol{\vartheta}}_n\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_{2n}, h_3\right]\right\}.$$

From (a1), we have $I_{n1} = o_p(1)$ and $I_{n2} = o_p(1)$. Next we need to show $I_{n3} = o_p(1)$ and $I_{n4} = o_p(1)$. Note that

$$\begin{aligned}
I_{n3} =& \sqrt{n}\left(\mathbb{P}_n - \mathbf{P}\right)\left\{\psi\left(\boldsymbol{\vartheta}_0\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right] - \psi\left(\boldsymbol{\vartheta}_0\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_{2n}, h_3\right]\right\} \\
&+ \sqrt{n}\mathbf{P}\left\{\psi\left(\boldsymbol{\vartheta}_0\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3\right] - \psi\left(\boldsymbol{\vartheta}_0\right)\left[\boldsymbol{h}_1, \boldsymbol{h}_{2n}, h_3\right]\right\} \\
=& I_{n31} + I_{n32}.
\end{aligned}$$

Similarly to proving (a1), $I_{n31} = o_p(1)$ and $I_{n32} = 0$ since $S(\boldsymbol{\vartheta}_0) = 0$ for any $h_{2,}$ and $h_{2n} \in \mathcal{H}_2$. For I_{n4},

$$\begin{aligned}
|I_{n4}| &\leq \sqrt{n} d\left(\hat{\boldsymbol{\vartheta}}_n, \theta_0\right)\left(\max_{j=1,\ldots,p_2} \|h_{2j} - h_{2nj}\|_\infty\right) \\
&= O_p\left(\max\left\{n^{-(1-\nu)/3}, n^{-r\nu}\right\} n^{-rv+1/2}\right) \\
&= o_p(1)
\end{aligned}$$

if $1/(4r) < \nu < 1/2$. Thus (a2) holds.

Condition (a3) holds because $\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$ is a Donsker class and the functionals $A_1(\boldsymbol{\vartheta})[\boldsymbol{h}_1]$, $A_2(\boldsymbol{\vartheta})[\boldsymbol{h}_2]$ and $A_3(\boldsymbol{\vartheta})[h_3]$ are bounded Lipschitz functions with respect to $\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$ due to the compactness of $\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$.

For (a4), by the smoothness of $S(\boldsymbol{\vartheta})$ the Fréchet differentiability holds and the derivative of $S(\boldsymbol{\vartheta})$ at $\boldsymbol{\vartheta}_0$, denoted by $\dot{S}(\boldsymbol{\vartheta}_0)$ is a map from the space $\{\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0 : \boldsymbol{\vartheta} \in \mathcal{U}\}$ to $l^\infty(\mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3)$. Now we calculate $\dot{S}(\boldsymbol{\vartheta}_0)$ as

$$
\begin{aligned}
&\dot{S}(\boldsymbol{\vartheta}_0)(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] \\
=&\frac{d}{d\epsilon}\left\{A_1(\boldsymbol{\vartheta}_0 + \epsilon(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0))[\boldsymbol{h}_1]\right\}\big|_{\epsilon=0} \\
&+ \frac{d}{d\epsilon}\left\{A_2(\boldsymbol{\vartheta}_0 + \epsilon(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0))[\boldsymbol{h}_1]\right\}\big|_{\epsilon=0} \\
&+ \frac{d}{d\epsilon}\left\{A_3(\boldsymbol{\vartheta}_0 + \epsilon(\boldsymbol{\vartheta} - \boldsymbol{\vartheta}_0))[\boldsymbol{h}_1]\right\}\big|_{\epsilon=0} \\
=& - \mathbf{P}\int_0^\tau Y(t)\boldsymbol{h}_1^T\mathbf{W}(t)\exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T(t)\,\mathbf{Z}(t)\right) \\
&\times\left\{\left((\boldsymbol{\gamma} - \gamma_0)^T\mathbf{W}(t) + (\boldsymbol{\beta}(t) - \boldsymbol{\beta}_0(t))^T\mathbf{Z}(t)\right)\Lambda_0(t) + (\Lambda(t) - \Lambda_0(t))\right\}dH(t) \\
&- \mathbf{P}\int_0^\tau Y(t)\boldsymbol{h}_2^T(t)\,\mathbf{Z}(t)\exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T(t)\,\mathbf{Z}(t)\right) \\
&\times\left\{\left((\boldsymbol{\gamma} - \gamma_0)^T\mathbf{W}(t) + (\boldsymbol{\beta}(t) - \boldsymbol{\beta}_0(t))^T\mathbf{Z}(t)\right)\Lambda_0(t) + (\Lambda(t) - \Lambda_0(t))\right\}dH(t) \\
&- \mathbf{P}\int_0^\tau Y(t)h_3(t)\exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t) + \boldsymbol{\beta}^T\mathbf{Z}(t)\right) \\
&\times\left\{\frac{\Lambda(t) - \Lambda_0(t)}{\Lambda_0(t)} + (\boldsymbol{\gamma} - \gamma_0)^T\mathbf{W}(t) + (\boldsymbol{\beta}(t) - \boldsymbol{\beta}_0(t))^T\mathbf{Z}(t)\right\}dH(t).
\end{aligned}
$$

Thus, we have

$$\begin{aligned}
&\dot{S}(\boldsymbol{\vartheta}_0)(\boldsymbol{\vartheta}-\boldsymbol{\vartheta}_0)[\boldsymbol{h}_1,\boldsymbol{h}_2,h_3] \\
=&(\boldsymbol{\gamma}-\gamma_0)^T Q_1(\boldsymbol{h}_1,\boldsymbol{h}_2,h_3) \\
&+\int_0^\tau (\boldsymbol{\beta}(t)-\boldsymbol{\beta}_0(t))^T dQ_2(\boldsymbol{h}_1,\boldsymbol{h}_2,h_3)(t) \\
&+\int_0^\tau (\Lambda(t)-\Lambda_0(t))\, dQ_3(\boldsymbol{h}_1,\boldsymbol{h}_2,h_3)(t)\ ,
\end{aligned}$$

where

$$\begin{aligned}
Q_1(\boldsymbol{h}_1,\boldsymbol{h}_2,h_3)=-\mathbf{P}\int_0^\tau \mathbf{W}(t)Y(t)&\left(\boldsymbol{h}_1^T\mathbf{W}(t)\Lambda_0(t)+\boldsymbol{h}_2^T(t)\,\mathbf{Z}(t)\Lambda_0(t)+h_3(t)\right)\\
&\times\exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t)+\boldsymbol{\beta}^T(t)\,\mathbf{Z}(t)\right)dH(t)\,,
\end{aligned}$$

$$\begin{aligned}
dQ_2(\boldsymbol{h}_1,\boldsymbol{h}_2,h_3)(t)=-\mathbf{P}\mathbf{Z}(t)Y(t)&\left(\boldsymbol{h}_1^T\mathbf{W}(t)\Lambda_0(t)+\boldsymbol{h}_2^T(t)\,\mathbf{Z}(t)\Lambda_0(t)+h_3(t)\right)\\
&\times\exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t)+\boldsymbol{\beta}^T(t)\,\mathbf{Z}(t)\right)dH(t)\,,
\end{aligned}$$

and

$$\begin{aligned}
dQ_3(\boldsymbol{h}_1,\boldsymbol{h}_2,h_3)(t)=-\mathbf{P}\frac{Y(t)}{\Lambda_0(t)}&\left(\boldsymbol{h}_1^T\mathbf{W}(t)\Lambda_0(t)+\boldsymbol{h}_2^T(t)\,\mathbf{Z}(t)\Lambda_0(t)+h_3(t)\right)\\
&\times\exp\left(\boldsymbol{\gamma}^T\mathbf{W}(t)+\boldsymbol{\beta}^T(t)\,\mathbf{Z}(t)\right)dH(t)\,.
\end{aligned}$$

We can also show $Q = (Q_1, Q_2, Q_3)$ is one-to-one by the similar method in He et al. (2017).

For (a5), we have

$$S\left(\hat{\boldsymbol{\vartheta}}_n\right) - S\left(\boldsymbol{\vartheta}_0\right) - \dot{S}\left(\boldsymbol{\vartheta}_0\right)\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) = D_{n1} + D_{n2} + D_{n3}\,,$$

where

$$D_{n1} = A_1\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_1] - \frac{d}{d\varepsilon}\left\{A_1\left(\boldsymbol{\vartheta}_0 + \varepsilon\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right)\right)[\boldsymbol{h}_1]\right\}\bigg|_{\varepsilon=0},$$

$$D_{n2} = A_2\left(\hat{\boldsymbol{\vartheta}}_n\right)[\boldsymbol{h}_2] - \frac{d}{d\varepsilon}\left\{A_2\left(\boldsymbol{\vartheta}_0 + \varepsilon\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right)\right)[\boldsymbol{h}_2]\right\}\bigg|_{\varepsilon=0},$$

and

$$D_{n3} = A_3\left(\hat{\boldsymbol{\vartheta}}_n\right)[h_3] - \frac{d}{d\varepsilon}\left\{A_3\left(\boldsymbol{\vartheta}_0 + \varepsilon\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right)\right)[h_3]\right\}\bigg|_{\varepsilon=0}.$$

It is not hard to see

$$\begin{aligned} D_{n1} = \mathbf{P}\int_0^{\tau} & Y(t)\boldsymbol{h}_1^T\mathbf{W}(t)\exp\left(\boldsymbol{\gamma}_0^T\mathbf{W}(t) + \boldsymbol{\beta}_0^T(t)\,\mathbf{Z}(t)\right)\Lambda_0(t) \\ & \times q_1\left(\left(\hat{\boldsymbol{\gamma}} - \gamma_0\right)^T\mathbf{W}(t) - \left(\hat{\boldsymbol{\beta}}_n(t) - \boldsymbol{\beta}_0(t)\right)^T\mathbf{Z}(t)\right)dH(t), \end{aligned}$$

where $q_1(x) = 1 - \exp(y)(1-y)$ and $q_1(x) \le x^2$ when x is in a neighborhood of 0.

Thus

$$
\begin{aligned}
D_{n1} \leq & \mathbf{P} \int_0^\tau Y(t) \boldsymbol{h}_1^T \mathbf{W}(t) \exp\left(\boldsymbol{\gamma}_0^T \mathbf{W}(t) + \boldsymbol{\beta}_0^T(t) \mathbf{Z}(t)\right) \Lambda_0(t) \\
& \times \left\{ (\hat{\boldsymbol{\gamma}} - \gamma_0)^T \mathbf{W}(t) - \left(\hat{\boldsymbol{\beta}}_n(t) - \boldsymbol{\beta}_0(t)\right)^T \mathbf{Z}(t) \right\}^2 dH(t). \\
= & O\left(d^2\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}\right)\right).
\end{aligned}
$$

Similarly, we can show $D_{n2} \leq O\left(d^2\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}\right)\right)$ and $D_{n3} \leq O\left(d^2\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}\right)\right)$ and hence

$$
S\left(\hat{\boldsymbol{\vartheta}}_n\right) - S(\boldsymbol{\vartheta}_0) - \dot{S}(\boldsymbol{\vartheta}_0)\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) \leq O\left(d^2\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}\right)\right).
$$

Since $n^{1/2} d^2\left(\hat{\boldsymbol{\vartheta}}_n, \boldsymbol{\vartheta}\right) = O_p\left(n^{1/2} \max\left\{n^{-2(1-\nu)/3}, n^{-2r\nu}\right\}\right) = o_p(1)$ if $1/(4r) < v < 1/4$, we can conclude that $S\left(\hat{\boldsymbol{\vartheta}}_n\right) - S(\boldsymbol{\vartheta}_0) - \dot{S}(\boldsymbol{\vartheta}_0)\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right) = o_p\left(n^{-1/2}\right)$ and (a5) holds.

If (a1)-(a5) holds, according to He et al. (2017), we have

$$
-\sqrt{n} \dot{S}(\boldsymbol{\vartheta}_0)\left(\hat{\boldsymbol{\vartheta}}_n - \boldsymbol{\vartheta}_0\right)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] = \sqrt{n}(S_n - S)(\boldsymbol{\vartheta}_0)[\boldsymbol{h}_1, \boldsymbol{h}_2, h_3] + o_p(1)
$$

uniformly in $\boldsymbol{h}_1, \boldsymbol{h}_2, h_3$. For each $(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$, Q is invertible by the similar arguement in He et al. (2017). Then there exists $(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) \in \mathcal{H}_1 \times \mathcal{H}_2 \times \mathcal{H}_3$ such that

$$
Q_1(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) = \boldsymbol{h}_1,\ Q_2(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) = \boldsymbol{h}_2,\ Q_3(\boldsymbol{h}_1, \boldsymbol{h}_2, h_3) = h_3.
$$

Therefore, we have

$$\begin{aligned}&(\hat{\gamma}-\gamma_0)^T \boldsymbol{h}_1+\int_0^\tau\left(\hat{\boldsymbol{\beta}}_n(t)-\boldsymbol{\beta}_0(t)\right)^T d\boldsymbol{h}_2(t)+\int_0^\tau\left(\hat{\Lambda}(t)-\Lambda_0(t)\right)dh_3(t)\\ =&\sqrt{n}\,(S_n-S)(\boldsymbol{\vartheta}_0)[\boldsymbol{h}_1,\boldsymbol{h}_2,h_3]+o_p(1)\rightarrow_d N\left(0,\sigma^2\right),\end{aligned}$$

where $\sigma^2 = E\left[\psi^2(\boldsymbol{\vartheta}_0)[\boldsymbol{h}_1,\boldsymbol{h}_2,h_3]\right]$ because of (a3). To find the asymptotic distribution of γ only, we can find $\boldsymbol{h}_1$, $\boldsymbol{h}_2$ and h_3 as a solution of $Q_2 = 0$ and $Q_3 = 0$. Unfortunately, we cannot find the explicit forms of $\boldsymbol{h}_1$, $\boldsymbol{h}_2$ and h_3 as He et al. (2017). The similar difficulty exists for deriving the asymptotic variance of $\boldsymbol{\beta}(t)$ too. Hence, we adopt the *ad hoc* variance estimation methods in the main body.

References

Chen, X. (2007). Large sample sieve estimation of semi-nonparametric models. In *Handbook of Econometrics*, pp. 5549–5632. Elsevier.

He, X., X. Feng, X. Tong, and X. Zhao (2017). Semiparametric partially linear varying coefficient models with panel count data. *Lifetime Data Analysis 23*(3), 439–466.

Lu, M., Y. Zhang, and J. Huang (2009). Semiparametric estimation methods for panel count data using monotone B-splines. *Journal of the American Statistical Association 104*(487), 1060–1070.

Schumaker, L. (2007). *Spline Functions: Basic Theory.* Cambridge University Press.

Vaart, A. W. V. D. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes.* Springer Series in Statistics. Springer-VerIag New York.

Wellner, J. A. and Y. Zhang (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *The Annals of Statistics 35*(5), 2106–2142.