Efficient Estimation of the Partly Linear Additive Hazards Model with Current Status Data (Additional Supporting Information: Supplementary Material – Technical Lemmas and Their Proofs)

Xuewen Lu

Department of Mathematics and Statistics

University of Calgary

2500 University Drive NW

Calgary, AB T2N 1N4, Canada

email: lux@math.ucalgary.ca

Peter X.-K. Song

Department of Biostatistics

University of Michigan

1420 Washington Heights

Ann Arbor, MI 48109-2029

email: pxsong@umich.edu

1. LEMMAS L.1-L.5 AND THEIR PROOFS

Lemmas L.1-L.5 are used to prove Theorem 2, which addresses the consistency and rate of convergence of all the estimators for the nonparametric functions. We follow the route of Huang (1999) in the partly linear additive Cox model with right censored data. We first establish a sub-optimal convergence rate by taking advantage of concavity of the likelihood function. Then we focus our attention on a sufficiently small neighborhood of the parameters to establish Theorem 2.

For any probability measure Q and any function f, define $L_2(Q) = \{f : \int f^2 dQ < \infty\}$ and $\|f\|_2 = (\int f^2 dQ)^{1/2}$. For any subclass $\mathcal{F}$ of $L_2(Q)$, define the bracketing number $N_{[]}(\varepsilon, \mathcal{F}, L_2(Q))$ =min$\{m$: there exist $f_1^L, f_1^U, \cdots, f_m^L, f_m^U$ such that for each $f \in \mathcal{F}$, $f_i^L \leq f \leq f_i^U$ for some i, and $\|f_i^U - f_i^L\|_2 \leq \varepsilon\}$. For any $\delta > 0$, denote

$$J_{[]}(\delta, \mathcal{F}, L_2(Q)) = \int_0^\delta \sqrt{1 + \log N_{[]}(\varepsilon, \mathcal{F}, L_2(Q))}\, d\varepsilon.$$

For $\mathbf{V}_i = (C_i, \mathbf{Z}_i(C_i), \mathbf{W}_i)$, let $\mathbb{P}_n$ be the empirical measure of $(\Delta_i, \mathbf{V}_i)$, $1 \leq i \leq n$ and let P be the probability measure of $(\Delta, \mathbf{V})$. Using linear functional notation, for any measurable function f, we can write $\mathbb{P}_n f = \int f d\mathbb{P}_n = n^{-1} \sum_{i=1}^n f(\Delta_i, \mathbf{V}_i)$.

Lemma L.1. *Without loss of generality, assume $r_n = q_n$. For any $\eta > 0$, let*

$$\begin{aligned}\Theta_n &= \{\Lambda(c) + \boldsymbol{\beta}'\mathbf{z}(c) + c\phi(\mathbf{w}) : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \leq \eta, \|\Lambda - \Lambda_0\|_2 \leq \eta, \|\phi - \phi_0\|_2 \leq \eta,\\ &\quad\ \Lambda \in \mathbb{L}_n, \phi(\mathbf{w}) \in \Phi_n\}.\end{aligned}$$

Then, for any $0 < \varepsilon < \eta$, there exists a constant $m > 0$, such that,

$$\log N_{[]}(\varepsilon, \Theta_n, L_2(P)) \leq m\{q_n \log(\eta/\varepsilon)\}.$$

Proof. Hereafter, we use m or m_i or m_z for generic positive constants, wherever applicable. Following the calculation of Shen and Wong (1994, page 597), we have $\log N_{[]}(\varepsilon, \mathbb{L}_n, L_2(P)) \leq m_1\{q_n \log(\eta/\varepsilon)\}$ and $\log N_{[]}(\varepsilon, \Phi_n, L_2(P)) \leq m_2\{q_n \log(\eta/\varepsilon)\}$. Therefore, the logarithm of the bracketing number of the class

$$\Psi_n = \{\Lambda(c) + c\phi(\mathbf{w}) : \|\Lambda - \Lambda_0\|_2 \leq \eta, \|\phi - \phi_0\|_2 \leq \eta, \Lambda(c) \in \mathbb{L}_n, \phi(\mathbf{w}) \in \Phi_n\}$$

is bounded by $m_3\{q_n \log(\eta/\varepsilon)\}$. Since the neighborhood $B(\eta) = \{\boldsymbol{\beta} : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \le \eta\}$ can be covered in $\mathbb{R}^d$ by $m_4(\eta/\varepsilon)^d$ balls with radius ε, and $|\boldsymbol{\beta}'\mathbf{z}(c) - \boldsymbol{\beta}_0'\mathbf{z}(c)| \le m_z\eta$ on $B(\eta)$ because of condition (B3), $B_z(\eta) = \{\boldsymbol{\beta}'\mathbf{z}(c) : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \le \eta\}$ can be covered by $m_5(\eta/\varepsilon)^d$ balls with radius ε. Therefore, the logarithm of the bracketing number of Θ_n is bounded by $m_3\{q_n \log(\eta/\varepsilon)\} + m_5 d \log(\eta/\varepsilon) \le m\{q_n \log(\eta/\varepsilon)\}$ for $m = m_3 + m_5 d$, $q_n \ge 4$.

Lemma L.2. *Let* $l_0(\Delta, c, \mathbf{z}(c), \mathbf{w}; \boldsymbol{\beta}, \phi, \Lambda) = \Delta \log[\exp\{-\Lambda(c) - \boldsymbol{\beta}'\mathbf{z}(c) - c\phi(\mathbf{w})\}] + (1 - \Delta)\log[1 - \exp\{-\Lambda(c) - \boldsymbol{\beta}'\mathbf{z}(c) - c\phi(\mathbf{w})\}]$. *Define a class of functions*

$$\mathcal{L}_0(\eta) = \{l_0 : \|\boldsymbol{\beta} - \boldsymbol{\beta}_0\| \le \eta, \|\Lambda - \Lambda_0\|_2 \le \eta, \|\phi - \phi_0\|_2 \le \eta, \Lambda \in \mathbb{L}_n, \phi(\mathbf{w}) \in \Phi_n\}.$$

Then for any $0 < \varepsilon < \eta$ *and some positive constant* m_0,

$$\log N_{[]}(\varepsilon, \mathcal{L}_0(\eta), L_2(P)) \le m_0\{q_n \log(\eta/\varepsilon)\}.$$

Consequently, by Lemma 3.4.2 of Van der Vaart and Wellner (1996),

$$J_{[]}(\eta, \mathcal{L}_0(\eta), L_2(P)) \le m_0 q_n^{1/2}\eta.$$

Proof. Since the exponential function $\exp(\cdot)$ is monotone, by Lemma L.1, the entropy of the class consisting of functions $\exp\{-\Lambda(c) - \boldsymbol{\beta}'\mathbf{z}(c) - c\phi(\mathbf{w})\}$ for $\Lambda(c) + \boldsymbol{\beta}'\mathbf{z}(c) + c\phi(\mathbf{w}) \in \Theta_n$ is bounded by $m_0\{q_n \log(\eta/\varepsilon)\}$. Therefore, the bracketing entropy of the class $\mathcal{L}_0(\eta)$ is bounded by $m_0\{q_n \log(\eta/\varepsilon)\}$ as well.

Lemma L.3. *Suppose that* $g = \Lambda(C) + \boldsymbol{\beta}'\mathbf{Z}(C) + C\phi(\mathbf{W})$, $\Lambda \in \mathcal{L}$, $\phi \in \mathcal{A}$. *Then, there exists a function* $g_n = \Lambda_n(C) + \boldsymbol{\beta}'\mathbf{Z}(C) + C\phi_n(\mathbf{W})$, $\Lambda_n \in \mathbb{L}_n$, $\phi_n \in \Phi_n$ *with* $\mathbb{P}_n\phi_n = 0$ *such that*

$$\|g_n - g\|_2 = O_p(n^{-\nu p} + n^{-(1-\nu)/2}).$$

Proof. According to Lu (2007), there exists $\Lambda_n \in \mathbb{L}_n$ such that $\|\Lambda_n - \Lambda\|_2 = O_p(n^{-\nu p})$. By Corollary 6.21 of Schumaker (1981, page 227), for ϕ, there exists a $\phi_n^* \in \Phi_n$ such that $\|\phi_n^* - \phi\|_\infty = O(n^{-\nu p})$. Let $\phi_n = \phi_n^* - n^{-1}\sum_{i=1}^n \phi_n^*(W_i) = \phi_n^* - \mathbb{P}_n\phi_n^*$. Then $\mathbb{P}_n\phi_n = 0$. Noticing $|\phi_n - \phi| \le |\phi_n^* - \phi| + |\mathbb{P}_n\phi_n^*|$, we consider

$$\mathbb{P}_n\phi_n^* = (\mathbb{P}_n - P)\phi_n^* + P(\phi_n^* - \phi).$$

By Lemma L.2 and Lemma 3.4.2 of Van der Vaart and Wellner (1996), $(\mathbb{P}_n - P)\phi_n^* = O_p(n^{-1/2}n^{\nu/2})$, and $|P(\phi_n^* - \phi)| \le \|\phi_n^* - \phi\|_\infty = O(n^{-\nu p})$. Therefore, $\|\phi_n - \phi\|_\infty \le O_p(n^{-\nu p} + n^{-(1-\nu)/2})$ and $\|\phi_n - \phi\|_2 = O_p(n^{-\nu p} + n^{-(1-\nu)/2})$. Finally, let $g_n = \Lambda_n(C) + \boldsymbol{\beta}'\mathbf{Z}(C) + C\phi_n(\mathbf{W})$, the lemma follows from the triangle inequality.

Lemma L.4. *Denote* $l_0(\Delta, g) = \Delta\{-g\} + (1-\Delta)\log\{1 - \exp(-g)\}$. *For any* g *with* $\|g - g_n\|_\infty \le \eta$, *constant* $\eta > 0$, *there exist constants* $0 < m_1, m_2 < \infty$ *such that*

$$\begin{aligned} & -m_1\|g - g_n\|_2^2 + O_p(n^{-2\nu p} + n^{-(1-\nu)}) \\ \le\ & Pl_0(\Delta, g) - Pl_0(\Delta, g_n) \\ \le\ & -m_2\|g - g_n\|_2^2 + O_p(n^{-2\nu p} + n^{-(1-\nu)}). \end{aligned}$$

Proof. Let $h = g - g_0$, where g_0 is the true value of g. Let

$$L_1(s) = Pl_0(\Delta, g_0 + sh) - Pl_0(\Delta, g_0).$$

The first and the second derivatives of $L_1(s)$ are given by

$$\begin{aligned} \dot{L}_1(s) &= P\left[(1+C)h(1-\Delta)\frac{\exp(-(g_0+sh)) - \exp(-g_0)}{\{1-\exp(-(g_0+sh))\}\{1-\exp(-g_0)\}}\right], \\ \ddot{L}_1(s) &= -P\left[\frac{(1+C)(1-\Delta)\exp(-(g_0+sh))}{\{1-\exp(-(g_0+sh))\}^2}h^2\right]. \end{aligned}$$

Since $L_1(0) = \dot{L}_1(0) = 0$, by Taylor expansion, we have

$$Pl_0(\Delta, g) - Pl_0(\Delta, g_0) = L_1(1) = \ddot{L}_1(\xi)/2,$$

where ξ is a value between 0 and 1. By the same arguments as those made in the proof of Lemma L.5, there exit constants $m_1 > m_2 > 0$ such that

$$-(m_1/2)\|g - g_0\|_2^2 \le Pl_0(\Delta, g) - Pl_0(\Delta, g_0) \le -(2m_2)\|g - g_0\|_2^2.$$

Likewise, it can be shown that

$$|Pl_0(\Delta, g_n) - Pl_0(\Delta, g_0)| = O_p(\|g_n - g_0\|_2^2).$$

Finally, using the following inequality,

$$(1/2)\|g - g_n\|_2^2 - \|g_n - g_0\|_2^2 \leq \|g - g_0\|_2^2 \leq 2\|g - g_n\|_2^2 + 2\|g_n - g_0\|_2^2,$$

we obtain

$$\begin{aligned} -m_1\|g - g_n\|_2^2 + O_p(1)\|g_n - g_0\|_2^2 &\leq Pl_0(\Delta, g) - Pl_0(\Delta, g_n) \\ &\leq -m_2\|g - g_n\|_2^2 + O_p(1)\|g_n - g_0\|_2^2. \end{aligned}$$

Combining this inequality and Lemma L.3, we complete the proof.

Lemma L.5. *For* $\mathbf{v} = (c, \mathbf{z}(c), \mathbf{w})$, *let* $g_n(\mathbf{v}) = \Lambda_n(c) + \boldsymbol{\beta}'\mathbf{z}(c) + c\phi_n(\mathbf{w})$. *Denote the estimator of* $g_0(\mathbf{v})$ *by* $\hat{g}_n(\mathbf{v}) = \hat{\Lambda}_n(c) + \hat{\boldsymbol{\beta}}_n'\mathbf{z}(c) + c\hat{\phi}_n(\mathbf{w})$. *Let* $q_n = K_n + \rho$ *be the number of polynomial splines basis functions defined in Section 2, we have*

$$\|\hat{g}_n - g_n\|_2^2 = O_p(q_n^{-1}).$$

Furthermore, by Lemma 7 of Stone (1986), $\|\hat{g}_n - g_n\|_\infty = o_p(1)$.

Proof. Choose $\mathbf{b} \in \mathbb{R}^d$, $\psi_n \in \Phi_n$ and $\tau_n \in \mathbb{L}_n$ such that $\|\tau_n(C) + \mathbf{b}'\mathbf{Z}(C) + C\psi_n(\mathbf{W})\|_2^2 = O(q_n^{-1})$. This is possible because both c and $\mathbf{z}(c)$ are bounded. Denote $h_n = \tau_n(c) + \mathbf{b}'\mathbf{z}(c) + c\psi_n(\mathbf{w})$. Let $b_n(\mathbf{v}, s) = g_n(\mathbf{v}) + sh_n = \Lambda_n(c) + s\tau_n(c) + (\boldsymbol{\beta} + s\mathbf{b})'\mathbf{z}(c) + c(\phi_n + s\psi_n(\mathbf{w}))$. Let $H_n(s) = \mathbb{P}_n(b_n(\cdot, s)) = \mathbb{P}_n(g_n + sh_n)$. It is easy to obtain

$$\begin{aligned} H_n(s) &= \frac{1}{n}\sum_{i=1}^n \Delta_i\{-b_n(\mathbf{V}_i, s)\} + (1 - \Delta_i)\log\{1 - \exp(-b_n(\mathbf{V}_i, s))\}, \\ \dot{H}_n(s) &= \frac{1}{n}\sum_{i=1}^n (1 + C_i)h_n\left\{-\Delta_i + (1 - \Delta_i)\frac{\exp(-b_n(\mathbf{V}_i, s))}{1 - \exp(-b_n(\mathbf{V}_i, s))}\right\}, \\ \ddot{H}_n(s) &= -\frac{1}{n}\sum_{i=1}^n (1 - \Delta_i)(1 + C_i)^2 h_n^2 \frac{\exp(-b_n(\mathbf{V}_i, s))}{\{1 - \exp(-b_n(\mathbf{V}_i, s))\}^2}. \end{aligned}$$

Because $\ddot{H}_n(s) \leq 0$, $H_n(s)$ is a concave function of s and $\dot{H}_n(s)$ is a non-increasing function. Therefore, to prove the lemma, it suffices to show that for any $s = s_0 > 0$, $\dot{H}_n(s_0) < 0$ and $\dot{H}_n(-s_0) > 0$ except on events with probability tending to zero. Note if this property holds, then $\hat{g}_n$ must be between $g_n - s_0 h_n$ and $g_n + s_0 h_n$, so $\|\hat{g}_n - g_n\|_2 \le s_0 \|h_n\|_2$. Without loss of generality, assume $s_0 = 1$. Using the identity

$$P\left[(1+C)h_n\left\{\frac{\exp(-g_0(\mathbf{V})) - \Delta}{1-\exp(-g_0(\mathbf{V}))}\right\}\right] \quad = \quad 0,$$

by some algebraic operations we have

$$\begin{aligned}
\dot{H}_n(1) = (\mathbb{P}-P)&\left[(1+C)h_n\left\{\frac{\exp(-b_n(\mathbf{V},1)) - \Delta}{1-\exp(-b_n(\mathbf{V},1))}\right\}\right] \\
&+P\left[(1+C)h_n\left\{\frac{\exp(-b_n(\mathbf{V},1)) - \Delta}{1-\exp(-b_n(\mathbf{V},1))}\right\}\right] - P\left[(1+C)h_n\left\{\frac{\exp(-g_n(\mathbf{V})) - \Delta}{1-\exp(-g_n(\mathbf{V}))}\right\}\right] \\
&+P\left[(1+C)h_n\left\{\frac{\exp(-g_n(\mathbf{V})) - \Delta}{1-\exp(-g_n(\mathbf{V}))}\right\}\right] - P\left[(1+C)h_n\left\{\frac{\exp(-g_0(\mathbf{V})) - \Delta}{1-\exp(-g_0(\mathbf{V}))}\right\}\right] \\
\stackrel{\text{def.}}{=} \quad & I_{1n} + I_{2n} + I_{3n}.
\end{aligned}$$

Since $\inf_{\mathbf{V}}\{1-\exp(-b_n(\mathbf{V},1))\} > 1/m_1$ for some constant $m_1 > 0$, the first term is of order $n^{-1/2}$. In fact, by Lemma L.1 and Lemma L.2 on the bracket number for $\mathcal{L}_0(\eta)$, taking $\eta = q_n^{-1/2}$ leads to

$$\begin{aligned}
|I_{1n}| &\le m_1 \sup_{(\Delta,\mathbf{V})} |(\mathbb{P}-P)\left[(1+C)h_n\{\exp(-b_n(\mathbf{V},1)) - \Delta\}\right]| \\
&\le O_p(1) n^{-1/2} q_n^{-1/2}(q_n^{-1/2} + \log^{1/2} q_n) \\
&= O_p(n^{-1/2}).
\end{aligned}$$

In a similar way, we can show

$$\begin{aligned}
|I_{3n}| &\le O(1)\|h_n\|_2\|g_n - g_0\|_2 \\
&= O(1) q_n^{-1/2}(n^{-(1-\nu)/2} + n^{-\nu p}) \\
&= O(n^{-1/2}),
\end{aligned}$$

for $1/(1+2p) < \nu < 1/2$.

Now, we evaluate I_{2n}. Let

$$\begin{aligned}
L(s) &= P\left[(1+C)h_n\left\{\frac{\exp(-b_n(\mathbf{V},s)) - \Delta}{1-\exp(-b_n(\mathbf{V},s))}\right\}\right] - P\left[(1+C)h_n\left\{\frac{\exp(-g_n(\mathbf{V})) - \Delta}{1-\exp(-g_n(\mathbf{V}))}\right\}\right] \\
&= P\left[(1+C)h_n(1-\Delta)\frac{\exp(-b_n(\mathbf{V},s)) - \exp(-g_n(\mathbf{V}))}{\{1-\exp(-b_n(\mathbf{V},s))\}\{1-\exp(-g_n(\mathbf{V}))\}}\right].
\end{aligned}$$

By Taylor expansion, $I_{2n} = L(1) = L(0) + \dot{L}(\xi)$, $\xi \in (0,1)$, where $L(0) = 0$ and

$$\begin{aligned}\dot{L}(s) &= -P\left[\frac{(1+C)(1-\Delta)\exp(-b_n(\mathbf{V},s))}{\{1-\exp(-b_n(\mathbf{V},s))\}^2}h_n^2\right] \\ &= -P\left[\frac{(1+C)\exp(-b_n(\mathbf{V},s))\{1-\exp(-g_0(\mathbf{V}))\}}{\{1-\exp(-b_n(\mathbf{V},s))\}^2}h_n^2\right].\end{aligned}$$

By Lemma 7 of Stone (1986), $\|h_n\|_\infty \leq m q_n^{1/2}\|h_n\|_2 = O(1)$ for some constant $m > 0$. Therefore, $m_0 < b_n(\mathbf{v},s) = g_0(\mathbf{v}) + g_n(\mathbf{v}) - g_0(\mathbf{v}) + sh_n \leq g_0(\mathbf{v}) + m_2 \leq m_1 + m_2$ for $0 \leq s \leq 1$ and some constants $m_j > 0$, $j = 0,1,2$. Given that our function $k(x) = \exp(-x)/(1-\exp(-x))^2$ is a non-increasing function on $(0,\infty)$, we have

$$\frac{\exp(-b_n(\mathbf{V},s))}{\{1-\exp(-b_n(\mathbf{V},s))\}^2} \geq \frac{\exp(-m_1-m_2)}{\{1-\exp(-m_1-m_2)\}^2}.$$

Therefore, we obtain

$$\begin{aligned}\dot{L}(s) &\leq -\left[\frac{(1+l_c)\exp(-m_1-m_2)\{1-\exp(-m_1)\}}{\{1-\exp(-m_1-m_2)\}^2}P(h_n^2)\right] \\ &\stackrel{\text{def.}}{=} -m_3\|h_n\|_2^2\end{aligned}$$

and

$$I_{2n} \leq -m_3\|h_n\|_2^2 = -m_3 q_n^{-1}.$$

In summary, we yield

$$\dot{H}_n(1) \leq -m_3 q_n^{-1} + O(n^{-1/2}) < 0,$$

except on events with probability tending to zero. Using similar arguments, we can show that $\dot{H}_n(-1) > 0$ with high probability. This completes the proof of the Lemma L.5.

2. LEMMA L.6 AND ITS PROOF

To prove the asymptotic normality of the estimator of parameter $\boldsymbol{\beta}_0$, we apply a general theorem for semiparametric maximum likelihood estimation given in Huang (1996). The following lemma paves the path to Theorem 3.

Lemma L.6. *Under the given conditions in Theorem 3, for l_0 defined in Lemma L.4, let $s(\cdot,g) = \partial l_0(\cdot,g)/\partial g = -\Delta + (1-\Delta)\exp(-g)/\{1-\exp(-g)\}$. For real-valued vector functions*

$\mathbf{u} = \mathbf{a}_1(c) + c\mathbf{h}(\mathbf{w})$ *of* $(c, \mathbf{w}) \in \mathbb{R}^+ \times \mathbb{R}^J$*, let* $\mathbf{U} = \mathbf{a}_1(C) + C\mathbf{h}(\mathbf{W})$ *and* $\mathbf{U}^* = \mathbf{a}_1^*(C) + C\mathbf{h}^*(\mathbf{W})$*, denote*

$$s(\cdot, g)[\mathbf{Z}] = \frac{\partial s(\cdot, g)}{\partial g}\mathbf{Z}$$

and

$$s(\cdot, g)[\mathbf{U}] = \frac{\partial s(\cdot, g)}{\partial g}\mathbf{U}.$$

Then, we have the following results.

(C1) $\dot{l}_{n\Lambda}(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{\phi}_n)[\mathbf{a}_1^*] + \dot{l}_{n\phi}(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{\phi}_n)[C\mathbf{h}^*] = \mathbb{P}_n s(\cdot, \hat{g}_n)[\mathbf{U}^*] = o_p(n^{-1/2})$.

(C2) $(\mathbb{P}_n - P)\{s(\cdot, \hat{g}_n)[\mathbf{Z}] - s(\cdot, g_0)[\mathbf{Z}]\} = o_p(n^{-1/2})$ *and*
$(\mathbb{P}_n - P)\{s(\cdot, \hat{g}_n)[\mathbf{U}^*] - s(\cdot, g_0)[\mathbf{U}^*]\} = o_p(n^{-1/2})$.

(C3) $P\{s(\cdot, \hat{g}_n)(\mathbf{Z}(C) - \mathbf{U}^*) - s(\cdot, g_0)(\mathbf{Z}(C) - \mathbf{U}^*)\} = I(\boldsymbol{\beta}_0)(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) + o_p(n^{-1/2})$.

PROOF OF (C1). By condition (B6) and equations (A.2) and (A.3) in the information bound calculation, we can show that the elements of $\mathbf{a}_1^*$ and $\mathbf{h}^*$ are qth differentiable and their qth derivatives are bounded. Thus, by similar arguments as those in the proof of Lemma L.3, there exist an $\mathbf{a}_{1n}^*$ and a $\mathbf{h}_n^*$, their elements belong to $\mathcal{L}_n$ and Φ_n, respectively, such that

$$\|\mathbf{a}_{1n}^* - \mathbf{a}_1^*\|_2 = O(q_n^{-q}) \quad \text{and} \quad \|\mathbf{h}_n^* - \mathbf{h}^*\|_2 = O(q_n^{-q}).$$

By the definition of $(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{\phi}_n)$, for any $\mathbf{U}_n = \mathbf{a}_{1n} + C\mathbf{h}_n$, $\mathbf{a}_{1n} \in \mathcal{L}_n$, $\mathbf{h}_n \in \Phi_n$,

$$\dot{l}_{n\Lambda}(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{\phi}_n)[\mathbf{a}_{1n}] + \dot{l}_{n\phi}(\hat{\boldsymbol{\beta}}_n, \hat{\Lambda}_n, \hat{\phi}_n)[C\mathbf{h}_n] = \mathbb{P}_n s(\cdot, \hat{g}_n)[U_n] = 0.$$

Also notice that

$$P\{s(\cdot, g_0)[\mathbf{U}^* - \mathbf{U}_n^*]\} = 0$$

for $\mathbf{U}_n^* = \mathbf{a}_{1n}^* + C\mathbf{h}_n^*$. Hence,

$$\begin{aligned}
\mathbb{P}_n s(\cdot, \hat{g}_n)[\mathbf{U}^*] &= \mathbb{P}_n s(\cdot, \hat{g}_n)[\mathbf{U}^* - \mathbf{U}_n^*] \\
&= (\mathbb{P}_n - P)s(\cdot, \hat{g}_n)[\mathbf{U}^* - \mathbf{U}_n^*] + P\{(s(\cdot, \hat{g}_n) - s(\cdot, g_0))[\mathbf{U}^* - \mathbf{U}_n^*]\} \\
&= I_{1n} + I_{2n}.
\end{aligned}$$

By the maximal inequality in Lemma 3.4.2 of Van der Vaart and Wellner (1996) and some entropy calculations similar to those in Lemma L.2, it can be shown that $I_{1n} = o_p(n^{-1/2})$. By Taylor expansion and the given boundary conditions, there exists a constant $m > 0$ such that

$$|I_{2n}| \leq m\|\mathbf{U}^* - \mathbf{U}_n^*\|_2\|\hat{g}_n - g_0\|_2.$$

Therefore, $I_{2n} = n^{-q\nu}O_p(n^{-\nu p} + n^{-(1-\nu)/2}) = o_p(n^{-1/2})$ under the conditions in Theorem 3.

PROOF OF (C2). For $\mathbf{U} = \mathbf{Z}$ or $\mathbf{U}^*$, we have $P\{s(\cdot, \hat{g}_n)[\mathbf{U}] - s(\cdot, g_0)[\mathbf{U}]\}^2 \leq O(\|\hat{g}_n - g_0\|_2^2)$, and the ε-bracketing number of the class functions $S(\eta) = \{s(\cdot, \hat{g}_n)[\mathbf{U}] - s(\cdot, g_0)[\mathbf{U}] : \|g - g_0\|_2 \leq \eta\}$ is $q_n \log(\eta/\varepsilon)$. The corresponding entropy integral $J_{[]}(\eta, S(\eta), L_2(P))$ is $\eta q_n^{1/2} + q_n n^{-1/2}$. Therefore, by Lemma 3.4.2 of Van der Vaart and Wellner (1996) and Theorem 2, for $\eta = r_n = n^{(1-\nu)/2} + n^{\nu p}$, we have

$$E|(\mathbb{P}_n - P)\{s(\cdot, \hat{g}_n)[\mathbf{U}] - s(\cdot, g_0)[\mathbf{U}]\}| \leq O(1)n^{-1/2}(r_n^{-1}q_n^{1/2} + q_n n^{-1/2}) = o(n^{-1/2}).$$

This completes the proof of (C2).

PROOF OF (C3). By Taylor expansion, for some ξ between g_0 and $\hat{g}_n$, we have

$$s(\cdot, \hat{g}_n) = s(\cdot, g_0) + \left.\frac{\partial s(\cdot, g)}{\partial g}\right|_{g=g_0}(\hat{g}_n - g_0) + \frac{1}{2}\left.\frac{\partial^2 s(\cdot, g)}{\partial g}\right|_{g=\xi}(\hat{g}_n - g_0)^2.$$

Noticing that, for any function $k(\mathbf{v}) = k(c, \mathbf{z}, \mathbf{w})$ and $\mathbf{V} = (C, \mathbf{Z}(C), \mathbf{W})$,

$$-P\left\{\left.\frac{\partial s(\cdot, g)}{\partial g}\right|_{g=g_0} k(\mathbf{V})\right\} = P\{s^2(\cdot, g_0)k(\mathbf{V})\},$$

we obtain

$$\begin{aligned} &P\{s(\cdot, \hat{g}_n)[\mathbf{Z} - \mathbf{U}^*] - s(\cdot, g_0)[\mathbf{Z} - \mathbf{U}^*]\} \\ &= -Ps^2(\cdot, g_0)(\mathbf{Z} - \mathbf{U}^*)(\mathbf{Z}')(\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0) - Ps^2(\cdot, g_0)(\mathbf{Z} - \mathbf{U}^*)\{\hat{\Lambda}_n + C\hat{\phi}_n - (\Lambda_0 + C\phi_0)\} \\ &\quad +O(\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|^2 + \|\hat{\Lambda}_n - \Lambda_0\|_2^2 + \|\hat{\phi}_n - \phi_0\|_2^2). \end{aligned}$$

By (A.1) and Theorem 1, we see that

$$Ps^2(\cdot, g_0)(\mathbf{Z} - \mathbf{U}^*)\{\hat{\Lambda}_n + C\hat{\phi}_n - (\Lambda_0 + C\phi_0)\} = 0$$

and

$$Ps^2(\cdot, g_0)(\mathbf{Z} - \mathbf{U}^*)(\mathbf{Z}') = P\{s^2(\cdot, g_0)(\mathbf{Z} - \mathbf{U}^*)^{\otimes 2}\} = I(\boldsymbol{\beta}_0).$$

By Theorem 2, $\|\hat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0\|^2 = o_p(n^{-1/2})$, $\|\hat{\Lambda}_n - \Lambda_0\|_2^2 = o_p(n^{-1/2})$ and $\|\hat{\phi}_n - \phi_0\|_2^2 = o_p(n^{-1/2})$, therefore, (C3) is approved.

REFERENCES

Huang, J. (1996), "Efficient estimation for the proportional hazards model with interval censoring," *The Annals of Statistics*, 24, 540–568.

Huang, J. (1999), "Efficient estimation of the partly linear additive Cox model," *The Annals of Statistics*, 27, 1536–1563.

Lu, M. (2007), *Analysis of Panel Count Data Using Monotone Polynomial Splines*, University of Iowa: Doctoral Dissertation.

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

Shen, X., and Wong, W. H. (1994), "Convergence rate of sieve estimates," *The Annals of Statistics*, 22, 580–615.

Stone, C. J. (1986), "The dimensionality reduction principle for generalized additive models," *The Annals of Statistics*, 14, 590–606.

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