Supplementary Material for the Article
"Efficient Estimation of a Varying-Coefficient Partially Linear Proportional Hazards Model with Current Status Data"

Jun-Qiang Yang, Yuan Dong, Radhey Singh, Cheng Dong, Xuewen Lu

Construction of Sieve Spaces

To apply the sieve method, we first use B-splines to construct a sieve space for estimating $H(t)$ and $\phi_j(w)$'s, respectively, then followed by the maximum likelihood estimation for all the resulting parameters in the log-likelihood function (9). We consider $\phi_j(W)$ first. Assume that W takes values in $[L_w, U_w]$, where L_w and U_w are finite numbers. Let $L_w = \zeta_0 < \zeta_1 < \ldots < \zeta_K < \zeta_{K+1} = U_w$ be a set of knots that partition $[L_w, U_w]$ into $(K+1)$ sub-intervals, with $I_{K_i} = [\zeta_i, \zeta_{i+1})$, $i = 0, \ldots, K-1$ and $I_{K_K} = [\zeta_K, \zeta_{K+1}]$, where $K \geq 1$ is a positive integer satisfying $\max_{1 \leq i \leq K+1}(\zeta_i - \zeta_{i-1}) = O(n^{-\nu})$ and $K = O(n^{\nu})$ for some ν in $(0, 0.5)$. Let $\mathbb{B}_1$ be the space of polynomial splines of order $\rho_w \geq 2$, in which a functional element s satisfies that (i) s is a ρ_w-th order polynomial on interval I_{K_i} for $0 \leq i \leq K$, and (ii) for $\rho_w \geq 2$ and $0 \leq r \leq \rho_w - 2$, s is r times continuously differential on $[L_w, U_w]$. According to Corollary 4.10 of Schumaker (1981), there exists a local basis $\mathbf{B}_1 = \{B_{1k}, 1 \leq k \leq \mathrm{df}_w\}^\top$ for $\mathbb{B}_1$, where B_{1k}'s are normalized B-splines basis functions and $\mathrm{df}_w = K + \rho_w$. Thus for functions $\phi_j(w)$, we can use the following B-spline functions to model them:

$$\phi_j(w) = \sum_{k=1}^{\mathrm{df_w}} \eta_{jk} B_{1k}(w) = \eta_j^\top \mathbf{B}_1(w), \tag{25}$$

where $\phi_j(\cdot) \in \mathbb{B}_1$ and $\eta_j = (\eta_{j1}, \ldots, \eta_{j\mathrm{df}_w})^\top$ is a vector of coefficients.

Now we consider the baseline function, $H(t)$. We also use B-splines to model it and construct a sieve space to carry out the B-splines smoothing method. We notice that $H(t)$ is differentiable and increasing over $t \in [0, +\infty)$, and takes values in $(-\infty, \infty)$. Suppose that the support of variable C is a finite interval, say $[L_c, U_c]$. Similar to the partition of $[L_w, U_w]$, a partition of this interval can be made such that $\max_{1 \leq r \leq R+1}(t_r - t_{r-1}) = O(n^{-\nu})$, where t_r are knots, $r = 1, \ldots, R+1$, with $t_0 = L_c$ and $t_{R+1} = U_c$, $R \equiv R_n = O(n^{\nu})$ is a sample size dependent positive integer for some ν in $(0, 0.5)$. The reason of postulating an upper bound of 0.5 on ν in both $K = O(n^{\nu})$ and $R = O(n^{\nu})$ is that, according to Stone (1985), the optimal rate of convergence of a nonparametric estimator in an L_2-norm is typically of form $n^{-\rho/(2\rho+1)}$ $(\rho > 0.5)$, which is achieved at $\nu = 1/(2\rho+1)$. This implies that $0 < \nu \leq 1/(2\rho+1) < 0.5$, namely the upper bound of ν will not exceed 0.5. Let $\mathbb{B}_0$

be the space of monotone increasing polynomial splines of order $\rho_c \geq 2$ based on the knots $t_0, \ldots, t_R$ and sub-intervals $I_{R_r} = [t_r, t_{r+1})$, $r = 0, \ldots, R-1$ and $I_{R_R} = [t_R, t_{R+1}]$. We use the same B-splines approximation to $H(\cdot)$ as for $\phi(\cdot)$ explained above. There exists a local normalized B-splines basis of $\mathbb{B}_0$: $\mathbf{B}_0 = \{B_{0r}, 1 \leq r \leq \mathrm{df}_c\}^\top$ and $\mathrm{df}_c = R + \rho_c$. Then function $H(t)$ can be modled by B-spline functions, which is a linear combination of $(B_{01}, \ldots, B_{0\mathrm{df}_c})^\top$, i.e.

$$H(t) = \sum_{r=1}^{\mathrm{df}_c} \vartheta_r B_{0r}(t) = \vartheta^\top \mathbf{B}_0(t), \tag{26}$$

where $\vartheta = (\vartheta_1, \ldots, \vartheta_{\mathrm{df}_c})^\top$ is a vector of coefficients subject to the constraint $\vartheta_1 \leq \cdots \leq \vartheta_{\mathrm{df}_c}$ so that $H(t)$ is increasing in t. By Theorem 5.9 of Schumaker (1981), the monotonicity of $H(t)$ is guaranteed by such a constraint on the coefficients.

Appendix A: Proofs of Theorem 2 and Theorem 3

A1: Proof of Theorem 2

To prove Theorem 2, we need Lemmas 1-6, the details of which are included in Appendix B. For l_0 defined in Lemma 6 and g satisfying $Pl_0(\cdot, g_n) - Pl_0(\cdot, g) \geq 0$ (see Lemma 6, this holds with high probability), we define a distance d_n to be

$$d_n^2(g, g_n) = Pl_0(\cdot, g_n) - Pl_0(\cdot, g).$$

Let $M_0(g) = Pl_0(\cdot, g), \zeta > 0$, by Lemma 3 and Lemma 3.4.2 of van der Vaart and Wellner (1996), we have

$$\begin{aligned} & \mathrm{E}\sup_{\zeta/2 \leq d_n(g, g_n) \leq \zeta} |\mathbb{P}_n(l_0(\cdot, g) - \mathbb{P}_n(l_0(\cdot, g_n)) - (M_0(g) - M_0(g_n))| \\ = \; & \mathrm{E}\sup_{\zeta/2 \leq d_n(g, g_n) \leq \zeta} \left|(\mathbb{P}_n - P)\{l_0(\cdot, g) - l_0(\cdot, g_n)\}\right| \\ \leq \; & n^{-1/2} \zeta q_n^{1/2}. \end{aligned}$$

Since $\hat{g}_n$ is a consistent estimator of g_n shown by Lemma 5, we obtain by Theorem 3.4.1 of van der Vaart and Wellner (1996) that

$$r_n^2 d_n^2(\hat{g}_n, g_n) = O_p(1),$$

where r_n satisfies

$$r_n^2 (r_n^{-1} q_n^{1/2}) = O(n^{1/2}).$$

It follows that $r_n = q_n^{-1/2} n^{1/2} = n^{(1-v)/2}$. Therefore, assuming $g = \hat{g}_n$ in Lemma 6, we obtain by Lemma 5 that

$$\|\hat{g}_n - g_n\|_2^2 = O_p(n^{-(1-v)} + n^{-2vp}).$$

Because $\|g_n - g_0\|_2^2 = O_p(n^{-(1-v)} + n^{-2vp})$ by Lemma 4, we have

$$\|\hat{g}_n - g_0\|_2^2 = O_p(n^{-(1-v)} + n^{-2vp}).$$

Observe that

$$\begin{aligned}\|\hat{g}_n - g_0\|_2^2 &= E[\{\hat{H}(C) + \sum_{j=1}^{J} \hat{\phi}_j(W)X_j + \hat{\beta}^\top V\} \\ &\quad -\{H_0(C) + \sum_{j=1}^{J} \phi_{j0}(W)X_j + \beta_0^\top V\}]^2,\end{aligned}$$

where the expectation is taken with respect to $\omega = (C, Z) = (C, X, V, W)$. Let $q(\omega) = p(\omega)(1 - p(\omega))$ and assume all the parameters in $q(\omega)$ take the true parameter values, then there exist $0 < m_1 < m_2 < \infty$ such that $m_1 \leq q(\omega) \leq m_2$. Then we have

$$\begin{aligned}&E[\{\hat{H}(C) + \sum_{j=1}^{J} \hat{\phi}_j(W)X_j + \hat{\beta}^\top V - (H_0(C) + \sum_{j=1}^{J} \phi_{j0}(W)X_j + \beta_0^\top V)\}^2 q(\omega)] \\ &= E[\{(\hat{H}(C) - H_0(C)) + \{\sum_{j=1}^{J}(\hat{\phi}_j(W) - \phi_{j0}(W))X_j + (\hat{\beta} - \beta_0)^\top V\}\}^2 q(\omega))] \\ &= E[\{(\hat{H}(C) - H_0(C)) + \{\sum_{j=1}^{J}(\hat{\phi}_j(W) - \phi_{j0}(W))X_j \\ &\quad +(\hat{\beta} - \beta_0)^\top(V - \boldsymbol{a}^*(C) - \sum_{j=1}^{J} X_j \boldsymbol{h}_j^*)\} + (\hat{\beta} - \beta_0)^\top(\boldsymbol{a}^*(C) + \sum_{j=1}^{J} X_j \boldsymbol{h}_j^*)\}^2 q(\omega))] \\ &= E[\{(\hat{\beta} - \beta_0)^\top(V - \boldsymbol{a}^*(C) - \sum_{j=1}^{J} X_j \boldsymbol{h}_j^*)\}^2 q(\omega)] \\ &\quad +E[\{(\hat{H}(C) - H_0(C)) + \sum_{j=1}^{J}(\hat{\phi}_j(W) - \phi_{j0}(W))X_j \\ &\quad +(\hat{\beta} - \beta_0)^\top(\boldsymbol{a}^*(C) + \sum_{j=1}^{J} X_j \boldsymbol{h}_j^*)\}^2 q(\omega)].\end{aligned}$$

The last equality follows from the orthogonality given in the proof of Theorem 1. Therefore, the first term on the right-hand-side of the last equality equals

$$
\begin{aligned}
& E[\{(\hat\beta-\beta_0)^\top(V-\boldsymbol{a}^*(C)-\sum_{j=1}^J X_j\boldsymbol{h}_j^*)\}^2 q(\omega)] \\
= \;& (\hat\beta-\beta_0)^\top E[\{V-\boldsymbol{a}^*(C)-\sum_{j=1}^J X_j\boldsymbol{h}_j^*\}^{\bigotimes 2} q(\omega)](\hat\beta-\beta_0) \\
= \;& (\hat\beta-\beta_0)^\top I(\beta_0)(\hat\beta-\beta_0) \\
\le \;& E[\{\hat H(C)+\sum_{j=1}^J \hat\phi_j(W)X_j+\hat\beta^\top V \\
& -(H_0(C)+\sum_{j=1}^J \phi_{j0}(W)X_j+\beta_0^\top V)\}^2 q(\omega)] \\
\le \;& m_2 E[\{\hat H(C)+\sum_{j=1}^J \hat\phi_j(W)X_j+\hat\beta^\top V\} \\
& -\{H_0(C)+\sum_{j=1}^J \phi_{j0}(W)X_j+\beta_0^\top V\}]^2 \\
= \;& m_2\|\hat g-g_0\|_2^2 \\
\le \;& O_p(n^{-(1-v)}+n^{-2vp}).
\end{aligned}
$$

Because the information matrix $I(\beta_0)$ is assumed to be nonsingular, it follows that

$$\left\|\hat\beta-\beta_0\right\|^2 = O_p(n^{-(1-v)}+n^{-2vp}).$$

This in turn implies that

$$
\begin{aligned}
E\left\|\hat H(C)-H_0(C)\right\|_2^2 &= O_p(n^{-(1-v)}+n^{-2vp}), \\
E\left\|\hat\phi_j(W)-\phi_{j0}(W)\right\|_2^2 &= O_p(n^{-(1-v)}+n^{-2vp}), \quad 1\le j\le J.
\end{aligned}
$$

A2: Proof of Theorem 3

To approve Theorem 3, we need Lemma 7, the detail of which is included in the Appendix B. Because $\mathbb{P}_n s(\cdot,\hat g_n)[V]=0$ by (C1) and (C2) of Lemma 7, we have

$$P\{s(\cdot,\hat g_n)[V-\boldsymbol{U}^*]-s(\cdot,g_0)[V-\boldsymbol{U}^*]\} \;=\; \mathbb{P}_n s(\cdot,g_0)[V-\boldsymbol{U}^*]+o_p(n^{-1/2}),$$

where $\boldsymbol{U}^*=\boldsymbol{a}^*(C)+\sum_{j=1}^J X_j\boldsymbol{h}_j^*(W)$. Hence, by (C3) of Lemma 7,

$$
\begin{aligned}
I(\beta_0)(\hat\beta_n-\beta_0) &= \mathbb{P}_n s(\cdot,g_0)[V-\boldsymbol{U}^*]+o_p(n^{-1/2}) \\
&= \mathbb{P}_n l^*_{\beta_0}(\delta,\omega)+o_p(n^{-1/2}).
\end{aligned}
$$

Finally, by the Central Limit Theorem, $\sqrt{n}(\hat{\beta}_n - \beta_0)$ has an asymptotically normal distribution with the asymptotic covariance matrix $I^{-1}(\beta_0)$; hence, $\hat{\beta}_n$ is semiparametric efficient. The proof of Theorem 3 is completed.

Appendix B: Proofs of Lemmas 1-7

Appendix B provides proofs of Lemmas 1-7, which are needed in proving Theorems 2-3.

Lemmas 1-6 are used to prove Theorem 2, which addresses the consistency and rate of convergence of all the estimators of the nonparametric functions and finite dimensional parameters. We follow the route of Huang (1999) for 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 on a sufficiently small neighbourhood of the parameters to establish Theorem 2. The proof of Theorem 3 is based on Theorem 6.1 of Huang (1996), which provides a set of sufficient conditions for the MLE of the finite-dimensional parameter in a class of semiparametric models to satisfy the Central Limit Theorem. Lemma 7 proves those sufficient conditions under the current setting of the proposed model.

For any probability measure Q and any function f, define $L_2(Q) = \{f : \int f^2 \mathrm{d}Q < \infty\}$ and $\|f\|_2 = (\int f^2 \mathrm{d}Q)^{1/2}$. For any subclass $\mathscr{F}$ of $L_2(Q)$, define the bracketing number $N_{[]}(\varepsilon, \mathscr{F}, L_2(Q)) =$
$\min\{m :$ there exist $f_1^L, f_1^U, \ldots, f_m^L, f_m^U$ such that for each $f \in \mathscr{F}, f_i^L \leq f \leq f_i^U$ for some i, and $\left\|f_i^U - f_i^L\right\|_2 \leq \varepsilon\}$. For any $\delta_0 > 0$, denote

$$J_{[]}(\delta_0, \mathscr{F}, L_2(Q)) = \int_0^{\delta_0} \sqrt{1 + \ln N_{[]}(\varepsilon, \mathscr{F}, L_2(Q))}\ \mathrm{d}\varepsilon.$$

For $\omega_i = (C_i, Z_i) = (C_i, X_i, V_i, W_i)$, let $\mathbb{P}_n$ be the empirical measure of $(\delta_i, \omega_i), i \leq i \leq n$ and P be the probability measure of (δ, ω). Using linear functional notation, for any measurable function f, we can write $\mathbb{P}_n f = \int f \mathrm{d}\mathbb{P}_n = n^{-1} \sum_{i=1}^n f(\delta_i, \omega_i)$.

The following Lemma 1 is Lemma 3.4.2 of van der Vaart and Wellner (1996), which is also used in Huang (1999). Let $X_1, \ldots, X_n$ be i.i.d. random variables with distribution Q, and Q_n be the empirical measure of these random variables. Denote $G_n = \sqrt{n}(Q_n - Q)$, and $\|G_n\|_{\mathscr{F}} = \sup_{f\in\mathscr{F}} |G_n f|$ for any measurable class of functions $\mathscr{F}$.

Lemma 1. *Let M_0 be a finite positive constant. Let $\mathscr{F}$ be a uniformly bounded class of measurable functions such that $\|f\|_2 < \zeta$ and $\|f\|_\infty \leq M_0$. Then*

$$\mathrm{E}_Q^* \|G_n\|_{\mathscr{F}} \leq C_0 J_{[]}(\delta_0, \mathscr{F}, L_2(Q)) \left(1 + \frac{J_{[]}(\delta_0, \mathscr{F}, L_2(Q))}{\zeta^2 \sqrt{n}} M_0\right),$$

where C_0 is a finite constant independent of n.

Lemma 2. *Without loss of generality, assume* $\mathrm{df}_c = \mathrm{df}_w = q_n$ *(the number of basis functions in constructing H and ϕ_j, respectively, see* (6) *and* (7)*). For any $\zeta > 0$, let*

$$\begin{aligned}\Theta_n = \quad & \{H(C) + \sum_{j=1}^{J} \phi_j X_j + \beta^\top V : \|H - H_0\|_2 \leq \zeta, \\ & \left\|\phi_j - \phi_{j0}\right\|_2 \leq \zeta, \|\beta - \beta_0\| \leq \zeta, H \in \mathscr{A}_n, \phi_j \in \mathscr{L}_n, 1 \leq j \leq J\}.\end{aligned}$$

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

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

Proof. Hereafter, we use m or m_i for generic positive constants, wherever applicable. Following the calculation of Shen and Wong (1994), we have

$$\ln N_{[]}(\varepsilon, \mathscr{A}_n, L_2(P)) \leq m_1\{q_n \ln(\zeta/\varepsilon)\},$$

and

$$\ln N_{[]}(\varepsilon, \mathscr{L}_n, L_2(P)) \leq m_2\{q_n \ln(\zeta/\varepsilon)\}.$$

Therefore, the logarithm of the bracketing number of the class

$$\begin{aligned}\Psi_n \quad = \quad & \{H(C) + \sum_{j=1}^{J} \phi_j(W) X_j : \|H - H_0\|_2 \leq \zeta, \\ & \left\|\phi_j - \phi_{j0}\right\|_2 \leq \zeta, H \in \mathscr{A}_n, \phi_j \in \mathscr{L}_n, 1 \leq j \leq J\}\end{aligned}$$

is bounded by $m_3\{q_n \ln(\zeta/\varepsilon)\}$. Since the neighbourhood $B(\zeta) = \{\beta : \|\beta - \beta_0\| \leq \zeta\}$ can be covered in $\mathbb{R}^d$ by $m_4(\zeta/\varepsilon)^d$ balls with radius ε, and $\left\|\beta^\top V - \beta_0^\top V\right\| \leq m_V \zeta$ on $B(\zeta)$ by condition (B3). So $B_V(\zeta) = \{\beta^\top V : \|\beta - \beta_0\| \leq \zeta\}$ can be covered by $m_5(\zeta/\varepsilon)^d$ balls with radius ε. Therefore, the logarithm of the bracketing number Θ_n is bounded by

$$m_3 q_n \ln(\zeta/\varepsilon) + d \cdot m_5(\zeta/\varepsilon) \leq m\{q_n \ln(\zeta/\varepsilon)\}$$

for $m = m_3 + d m_5$, since $q_n \geq 3 > 1$.

□

Lemma 3. *Let* $l_0(\delta, C, Z, H, \phi_1, \ldots, \phi_J, \beta) = \delta \ln[\exp\{-\exp(H(C) + \sum_{j=1}^{J} \phi_j(W) X_j + \beta^\top V)\}] + (1-\delta) \ln[1 - \exp\{-\exp(H(C) + \sum_{j=1}^{J} \phi_j(W) X_j + \beta^\top V)\}]$. *Define a class of functions*

$$\begin{aligned}\Gamma_0(\zeta) \quad = \quad & \{l_0 : \|H - H_0\|_2 \leq \zeta, \left\|\phi_j - \phi_{j0}\right\|_2 \leq \zeta, \\ & \|\beta - \beta_0\| \leq \zeta, H \in \mathscr{A}_n, \phi_j \in \mathscr{L}_n, 1 \leq j \leq J\}.\end{aligned}$$

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

$$\ln N_{[]}(\varepsilon, \Gamma_0(\zeta), L_2(P)) \leq m_0\{q_n \ln(\zeta/\varepsilon)\}.$$

Consequently, by Lemme 3.4.2 of van der Vaart and Wellner (1996),

$$J_{[]}(\zeta, \Gamma_0(\zeta), L_2(P)) \leq m_0 q_n^{1/2}\zeta.$$

Proof. Since the ln and exp functions are both monotone, by Lemma 2, the entropy of the class consisting of functions $\exp(H(C) + \sum_{j=1}^{J} \phi_j(W)X_j + \beta^\top V)$ and $\ln[1 - \exp\{-\exp(H(C) + \sum_{j=1}^{J} \phi_j(W)X_j + \beta^\top V)\}]$ for $H(C) + \sum_{j=1}^{J} \phi_j(W)X_j + \beta^\top V \in \Theta_n$ is bounded by $(m_0/2)\{q_n \ln(\zeta/\varepsilon)\}$. Therefore, the bracketing entropy of the class $\Gamma_0(\zeta)$ is bounded by $2(m_0/2)\{q_n \ln(\zeta/\varepsilon)\} = m_0\{q_n \ln(\zeta/\varepsilon)\}$, since $\Gamma_0(\zeta)$ is the sum of the two classes. □

Lemma 4. *Suppose that* $g = H(C) + \sum_{j=1}^{J} \phi_j(W)X_j + \beta^\top V, H \in \mathscr{A}$. *Then there exists a function* $g_n = H_n(C) + \sum_{j=1}^{J} \phi_{jn}(W)X_j + \beta^\top V, H_n(C) \in \tilde{\mathscr{A}_n}, \phi_{jn} \in \mathscr{L}_n$ *such that*

$$\|g_n - g\|_2 = O_p(n^{-vp}).$$

Proof. According to Lu (2007), there exists $H_n \in \tilde{\mathscr{A}}$ such that $\|H_n - H\|_2 = O_p(n^{-vp})$. Also, by Corollary 6.21 of Schumaker (1981), $\|\phi_{jn} - \phi_j\|_2 = O_p(n^{-\nu\rho})$. Let $g_n = H_n(C) + \sum_{j=1}^{J} \phi_{jn}(W)X_j + \beta^\top V$, then $\|g_n - g\|_2 = O_p(n^{-vp})$. □

Lemma 5. *For* $\omega = (C, Z) = (C, X, V, W)$, *let* $g_n(\omega) = H_n(C) + \sum_{j=1}^{J} \phi_{jn}(W)X_j + \beta^\top V$, *which satisfies* $\|g_n - g_0\|_2 = O_p(n^{-vp} + n^{-(1-v)/2})$ *by taking* $g = g_0$ *in Lemma 4. Denote the estimator of* $g_0(\omega)$ *by* $\hat{g}(\omega) = \hat{H}(C) + \sum_{j=1}^{J} \hat{\phi}_j(W)X_j + \hat{\beta}^\top V$. *Let* q_n *be the number of polynomial splines basis functions defined in Section 2. Then we have*

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

Furthermore, $\|\hat{g} - g_n\|_\infty = o_p(1)$ *by Lemma 7 of Stone (1985).*

Proof. Choose $\tau_n \in \tilde{\mathscr{A}_n}, \psi_{jn} \in \mathscr{L}_n, b \in \mathbb{R}^d$ such that $\left\|\tau_n(C) + \sum_{j=1}^{J} \psi_{jn}(W)X_j + b^\top V\right\|_2^2 = O(q_n^{-1})$. This is possible because X and V are bounded. Denote $h_n = \tau_n(C) + \sum_{j=1}^{J} \psi_{jn}(W)X_j + b^\top V$. Let $b_n(\omega, s) = g_n(\omega) + sh_n$ and

$H_n(s) = \mathbb{P}_n(l_0(\delta, b_n(\cdot, s))) = \mathbb{P}_n(l_0(\delta, g_n + sh_n))$, then

$$\begin{aligned}
H_n(s) &= \frac{1}{n}\sum_{i=1}^{n}[\delta_i\{-\exp(b_n(\omega_i, s))\} + (1-\delta_i)\ln\{1-\exp(-\exp(b_n(\omega_i, s)))\}],\\
H_n^{'}(s) &= \frac{1}{n}\sum_{i=1}^{n} h_n \exp(b_n(\omega_i, s))\left\{\frac{\exp(-\exp(b_n(\omega_i, s))) - \delta_i}{1-\exp(-\exp(b_n(\omega_i, s)))}\right\},\\
H_n^{''}(s) &= -\frac{1}{n}\sum_{i=1}^{n} h_n^2 \exp(2b_n(\omega_i, s))\left\{\frac{(1-\delta_i)\exp(-\exp(b_n(\omega_i, s)))}{1-\exp(-\exp(b_n(\omega_i, s)))}\right\}^2\\
&\quad \frac{1}{n}\sum_{i=1}^{n} h_n^2 \exp(b_n(\omega_i, s))\left\{\frac{\exp(-\exp(b_n(\omega_i, s))) - \delta_i}{1-\exp(-\exp(b_n(\omega_i, s)))}\right\}.
\end{aligned}$$

By the concavity of $l_0(\delta, g)$, $H_n(s)$ is a concave function of s and $H_n^{'}(s)$ is a non-increasing function. Therefore, to prove this lemma, it suffices to show that there exists $s = s_0 > 0$, $H_n^{'}(s_0) < 0$ and $H_n^{'}(-s_0) > 0$ except on events with probability tending to zero. Note that if this property holds, the $\hat{g}$ must be between $g_0 - s_0 h_n$ and $g_0 + s_0 h_n$, so that $\|\hat{g} - g_n\|_2 < s_0\|h_n\|_2$. Without loss of generality, assume $s_0 = 1$, with the identity

$$P\left[h_n \exp(g_0(\omega))\left\{\frac{\exp(-\exp(g_0(\omega))) - \delta}{1-\exp(-g_0(\omega))}\right\}\right] = 0,$$

by some algebraic operations we have

$$\begin{aligned}
H_n^{'}(1) = (\mathbb{P} - P)&\left[h_n \exp(b_n(\omega, 1))\left\{\frac{\exp(-\exp(b_n(\omega, 1))) - \delta}{1-\exp(b_n(\omega, 1)))}\right\}\right]\\
&+\left[P\left[h_n \exp(b_n(\omega, 1))\left\{\frac{\exp(-\exp(b_n(\omega, 1)) - \delta}{1-\exp(-\exp(b_n(\omega, 1)))}\right\}\right]\right.\\
&\left.-P\left[h_n \exp(g_n(\omega))\left\{\frac{\exp(-\exp(g_n(\omega))) - \delta}{1-\exp(-\exp(g_n(\omega)))}\right\}\right]\right]\\
&+\left[P\left[h_n \exp(g_n(\omega))\left\{\frac{\exp(-\exp(g_n(\omega))) - \delta}{1-\exp(-\exp(g_n(\omega)))}\right\}\right]\right.\\
&\left.-P\left[h_n \exp(g_0(\omega))\left\{\frac{\exp(-\exp(g_0(\omega))) - \delta}{1-\exp(-\exp(g_0(\omega)))}\right\}\right]\right]\\
\overset{\text{def.}}{=}\;& I_{1n} + I_{2n} + I_{3n}.
\end{aligned}$$

Since $\inf_W\{1-\exp(-\exp(b_n(\omega, 1)))\} > 1/m_1$ and $0 < m_{w_1} \leq \sup_W \exp(b_n(\omega, 1)) \leq m_{w_2} < \infty$ for some constant $m_1, m_{w_1}, m_{w_2} > 0$, the first term is of order $n^{-1/2}$. In fact, by Lemma

7 of Stone (1986), $\|h_n\|_\infty \le m_h df_{L_n}^{1/2}\|h_n\|_2 = O(1)$ for some constant $m_h > 0$; by Lemma 2 and Lemma 3 on the bracket number for $\mathcal{L}_0(\eta)$, taking $\eta = df_{L_n}^{-1/2}$ leads to

$$
\begin{aligned}
|I_{1n}| &\le m_1 \sup_{(\delta,\mathbf{W})} |(\mathbb{P}-P)\left[h_n\{\exp(-\exp(b_n(\omega,1))) - \exp(-\exp(g_0(\omega)))\right. \\
&\quad \left. + \exp(-\exp(g_0(\omega))) - \delta\}\right] \\
&\le O_p(1)n^{-1/2}\{df_{L_n}^{-1/2}(df_{L_n}^{1/2}) + O_p(1)\} \\
&= O_p(n^{-1/2}).
\end{aligned}
$$

In a similar way, we can show

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

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

Now, we evaluate I_{2n}. Let

$$
\begin{aligned}
S(s) &= P\left[h_n \exp(b_n(\omega,s))\left\{\frac{\exp(-\exp(b_n(\omega,s))) - \delta}{1-\exp(-\exp(b_n(\omega,s)))}\right\}\right] \\
&\quad -P\left[h_n \exp(g_n(\omega))\left\{\frac{\exp(-\exp(g_n(\omega))) - \delta}{1-\exp(-\exp(g_n(\omega)))}\right\}\right].
\end{aligned}
$$

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

$$
\begin{aligned}
\dot{S}(s) &= -P\left[h_n^2 \exp(2b_n(\omega,s))\frac{(1-\delta)\exp(-\exp(b_n(\omega,s)))}{\{1-\exp(-\exp(b_n(\omega,s)))\}^2}\right] \\
&\quad +P\left[h_n^2 \exp(b_n(\omega,s))\frac{\exp(-\exp(b_n(\omega,s))) - \delta}{1-\exp(-\exp(b_n(\omega,s)))}\right] \\
&= R_{n1}(s) + R_{n2}(s).
\end{aligned}
$$

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

$$
\frac{\exp(-\exp(b_n(\omega,s)))}{\{1-\exp(-\exp(b_n(\omega,s)))\}^2} \ge \frac{\exp(-\exp(m_1+m_2))}{\{1-\exp(-\exp(m_1+m_2))\}^2}.
$$

Therefore, we obtain for $0 \le s \le 1$

$$
\begin{aligned}
R_{n1}(s) &\le -\left[\frac{\exp(2m_0)\exp(-\exp(m_1+m_2))\{1-\exp(-m_1)\}}{\{1-\exp(-\exp(m_1+m_2))\}^2}P(h_n^2)\right] \\
&\stackrel{\text{def.}}{=} -m_3\|h_n\|_2^2.
\end{aligned}
$$

By some similar arguments, when n is large enough, there exists $0 < m_4 < m_3$ such that

$$R_{n2}(s) \leq m_4\|h_n\|_2^2.$$

Hence, we have

$$I_{2n} \leq -m_3\|h_n\|_2^2 + m_4\|h_n\|_2^2 = -(m_3 - m_4)O(df_{L_n}^{-1}).$$

In summary, let $m_5 = m_3 - m_4 > 0$, we yield

$$H_n^{'}(1) \leq -m_5 O(df_{L_n}^{-1}) + O(n^{-1/2}) < 0,$$

except on events with probability tending to zero. Using some similar arguments, we can also show that $H_n^{'}(-1) > 0$ with high probability. This completes the proof of the Lemma 5. □

Lemma 6. *Denote $l_0(\delta, g) = \delta\{-\exp(g)\} + (1-\delta)\ln\{1 - \exp(-\exp(g))\}$. Assume g_n satisfies $\|g_n - g_0\|_2 = O_p(n^{-vp} + n^{-(1-v)/2})$ (g_n exists by taking $g = g_0$ in Lemma 4). For any g with $\|g - g_n\|_\infty \leq \zeta$ and a constant $\zeta > 0$, there exist constants $0 < m_1, m_2 < \infty$ such that*

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

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

$$\begin{aligned} L_1^{'}(s) &= P\left[h\exp(g_0 + sh)\frac{\exp(-\exp(g_0+sh)) - \delta}{\{1 - \exp(-\exp(g_0+sh))\}}\right] \\ &= P\left[h\exp(g_0 + sh)\frac{\exp(-\exp(g_0+sh)) - \exp(-\exp(g_0))}{\{1 - \exp(-\exp(g_0+sh))\}}\right], \\ L_1^{''}(s) &= -P\left[h^2\exp(2(g_0+sh))\frac{(1-\delta)\exp(-\exp(g_0+sh))}{\{1 - \exp(-\exp(g_0+sh))\}^2}\right] \\ &\quad +P\left[h^2\exp(g_0+sh)\frac{\exp(-\exp(g_0+sh)) - \delta}{1 - \exp(-\exp(g_0+sh))}\right] \\ &= -P\left[h^2\exp(2(g_0+sh))\frac{(1 - \exp(-\exp(g_0)))\exp(-\exp(g_0+sh))}{\{1 - \exp(-\exp(g_0+sh))\}^2}\right] \\ &\quad +P\left[h^2\exp(g_0+sh)\frac{\exp(-\exp(g_0+sh)) - \exp(-\exp(g_0))}{1 - \exp(-\exp(g_0+sh))}\right]. \end{aligned}$$

Since $L_1(0) = L^{'}(0) = 0$, by Taylor expansion, we have

$$Pl_0(\delta, g) - Pl_0(\delta, g_0) = L_1(1) = L_1^{''}(\xi)/2,$$

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

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

Finally, using the following equation

$$\frac{1}{2}\|g - g_n\|_2^2 - \|g_n - g_0\|_2^2 \le \|g - g_0\|_2^2 \le 2\|g - g_n\|_2^2 + 2\|g_n - g_0\|_2^2 ,$$

we obtain

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

Combining this inequality and $\|g_n - g_0\|_2 = O_p(n^{-vp} + n^{-(1-v)/2})$ completes the proof. $\square$

Lemma 7. *Under the conditions listed in Theorem 3, for l_0 defined in Lemma 3 and g defined in Lemma 4, let $s(\cdot, g) = \partial l_0(\delta, g)/\partial g = \exp(g)[-\delta + (1-\delta)\exp(-\exp(g))/\{1 - \exp(-\exp(g))\}]$. For real-valued vector functions $\boldsymbol{u} = \boldsymbol{a}(c) + \sum_{j=1}^J X_j \boldsymbol{h}_j(w)$ of (c, w), let $\boldsymbol{U} = \boldsymbol{a}(C) + \sum_{j=1}^J X_j \boldsymbol{h}_j(W)$ and $\boldsymbol{U}^* = \boldsymbol{a}^*(C) + \sum_{j=1}^J X_j \boldsymbol{h}_j^*(W)$ and denote*

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

Then, we have the following results:

(C1) $\dot{l}_{nH}(\hat{H}, \hat{\phi}_1, \ldots, \hat{\phi}_J, \hat{\beta})[\boldsymbol{a}^*]$
$+ \sum_{j=1}^J \dot{l}_{n\phi_j}(\hat{H}, \hat{\phi}_1, \ldots, \hat{\phi}_J, \hat{\beta})[X_j \boldsymbol{h}_j^*] = \mathbb{P}_n s(\cdot, \hat{g}_n)[\boldsymbol{U}^*] = o_p(n^{-1/2})$.

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

(C3) $P\{s(\cdot, \hat{g}_n)(V - \boldsymbol{U}^*) - s(\cdot, g_0)(V - \boldsymbol{U}^*)\} = I(\beta_0)(\hat{\beta} - \beta_0) + o_p(n^{-1/2})$.

Proof of (C1). By Condition (B5) and equations (21) and (22) in the information matrix calculation, we can show that the elements of $\boldsymbol{a}^*$ and $\boldsymbol{h}_j^*$ are κth differentiable and their κth derivatives are bounded. Thus, by similar arguments as those in the proof of Lemma 4, there exist $\boldsymbol{a}_n^*$ and $\boldsymbol{h}_{jn}^*$ with heir elements belonging to $\mathscr{A}_n$ and $\mathscr{L}_n$, respectively, so that

$$\|\boldsymbol{a}_n^* - \boldsymbol{a}^*\|_2 = O(q_n^{-\kappa}) \quad \text{and} \quad \left\|\boldsymbol{h}_{jn}^* - \boldsymbol{h}_j^*\right\|_2 = O(q_n^{-\kappa}), \ \ 1 \le j \le J.$$

By the definition of $(\hat{H}, \hat{\phi}_1, \ldots, \hat{\phi}_J, \hat{\beta})$, for any $\boldsymbol{U}_n = \boldsymbol{a}_n + \sum_{j=1}^J X_j \boldsymbol{h}_{jn}$, $\boldsymbol{a}_n \in \mathscr{A}_n$, $\boldsymbol{h}_{jn} \in \mathscr{L}_n$,

$$\dot{l}_{nH}(\hat{H}, \hat{\phi}_1, \ldots, \hat{\phi}_J, \hat{\beta})[\boldsymbol{a}_n] + \sum_{j=1}^J \dot{l}_{n\phi_j}(\hat{H}, \hat{\phi}_1, \ldots, \hat{\phi}_J, \hat{\beta})[X_j \boldsymbol{h}_j] = \mathbb{P}_n s(\cdot, \hat{g}_n)[\boldsymbol{U}_n] = 0.$$

Also note that $P\{s(\cdot, g_0)[\boldsymbol{U}^* - \boldsymbol{U}_n^*]\} = 0$ for $\boldsymbol{U}_n^* = \boldsymbol{a}_n^* + \sum_{j=1}^J X_j \boldsymbol{h}_{jn}^*$. Hence,

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

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 3, 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

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

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

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

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

This completes the proof of (C2). $\square$

Proof of (C3). By Taylor expansion, for some ε between g_0 and $\hat{g}_n$, we have

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

Note that, for any function $k(\omega) = k(C, X, V, W)$,

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

we obtain

$$
\begin{aligned}
& P\{s(\cdot,\hat{g}_n)[V-\boldsymbol{U}^*]-s(\cdot,g_0)[V-\boldsymbol{U}^*]\} \\
= & -Ps^2(\cdot,g_0)(V-\boldsymbol{U}^*)(V^\top)(\hat{\beta}-\beta_0) \\
& -Ps^2(\cdot,g_0)(V-\boldsymbol{U}^*)\{\hat{H}+\sum_{j=1}^J X_j\hat{\phi}_j-(H_0+\sum_{j=1}^J \phi_{j0})\} \\
+ & O\left(\left\|\hat{\beta}-\beta_0\right\|^2+\left\|\hat{H}-H_0\right\|_2^2+\sum_{j=1}^J X_j^2\left\|\hat{\phi}_j-\phi_{j0}\right\|_2^2\right).
\end{aligned}
$$

By Theorem 1, we see that

$$
Ps^2(\cdot,g_0)(V-\boldsymbol{U}^*)\{\hat{H}+\sum_{j=1}^J X_j\hat{\phi}_j-(H_0+\sum_{j=1}^J X_j\phi_{j0})\}=0
$$

and

$$
Ps^2(\cdot,g_0)(V-\boldsymbol{U}^*)(V^\top)=P\{s^2(\cdot,g_0)(V-\boldsymbol{U}^*)^{\bigotimes 2}\}=I(\beta_0).
$$

By Theorem 2, $\left\|\hat{\beta}-\beta_0\right\|^2=o_p(n^{-1/2})$, $\left\|\hat{H}-H_0\right\|_2^2=o_p(n^{-1/2})$ and $\left\|\hat{\phi}_j-\phi_{j0}\right\|_2^2=o_p(n^{-1/2})$, $1\le j\le J$. Therefore, (C3) is approved. $\square$