Submitted to the Annals of Statistics
DOI: 10.1214/11-AOS934SUPP

A SIEVE M-THEOREM FOR BUNDLED PARAMETERS IN SEMIPARAMETRIC MODELS, WITH APPLICATION TO THE EFFICIENT ESTIMATION IN A LINEAR MODEL FOR CENSORED DATA

(Supplementary Material)

By Ying Ding [*] and Bin Nan [*,†]

University of Michigan

1. Proofs of technical lemmas. We first prove the lemmas that are needed for the proofs of Theorems 4.1, 4.2 and 4.3.

1.1. *Proof of Lemma 7.1.* This result follows by direct calculation:

$$
\begin{aligned}
&\dot{l}_\beta(\beta,\zeta(\cdot,\beta);Z)\\
&\quad= -X\bigg\{\Delta\dot{g}(\epsilon_0 - X'(\beta-\beta_0))\\
&\qquad\quad - \int_a^b 1(\epsilon_0\ge t)\exp\{g(t-X'(\beta-\beta_0))\}\dot{g}(t-X'(\beta-\beta_0))\,dt\bigg\},\\
&\dot{l}_\zeta(\beta,\zeta(\cdot,\beta);Z)[h(\cdot,\beta)] = \frac{\partial}{\partial\eta} l(\beta,(\zeta+\eta h)(\cdot,\beta);Z)|_{\eta=0}\\
&\quad= \Delta w(\epsilon_0 - X'(\beta-\beta_0))\\
&\qquad\quad - \int_a^b 1(\epsilon_0\ge t)\exp\{g(t-X'(\beta-\beta_0))\}w(t-X'(\beta-\beta_0))\,dt,\\
&\ddot{l}_{\beta\beta}(\beta,\zeta(\cdot,\beta);Z)\\
&\quad= XX'\bigg\{\Delta\ddot{g}(\epsilon_0-X'(\beta-\beta_0)) - \int_a^b 1(\epsilon_0\ge t)\exp\{g(t-X'(\beta-\beta_0))\}\\
&\qquad\quad\cdot\left[\ddot{g}(t-X'(\beta-\beta_0)) + \dot{g}^2(t-X'(\beta-\beta_0))\right]dt\bigg\},
\end{aligned}
$$

[*]Supported in part by NSF Grant DMS-07-06700.
[†]Supported in part by NSF grant DMS-10-07590 and NIH grant R01-AG036802.
AMS 2000 subject classifications: Primary 62E20, 62N01; secondary 62D05
Keywords and phrases: Accelerated failure time model; B-spline; bundled parameters; efficient score function; semiparametric efficiency; sieve maximum likelihood estimation.

$$
\begin{aligned}
&\ddot{l}_{\beta\zeta}(\beta,\zeta(\cdot,\beta);Z)[h(\cdot,\beta)] = \ddot{l}'_{\zeta\beta}(\beta,\zeta(\cdot,\beta);Z)[h(\cdot,\beta)] \\
&\quad = -X\bigg\{\Delta\dot{w}(\epsilon_0 - X'(\beta-\beta_0)) - \int_a^b 1(\epsilon_0 \geq t)\exp\{g(t - X'(\beta-\beta_0))\} \\
&\qquad\quad \cdot \big[\dot{w}(t - X'(\beta-\beta_0)) + \dot{g}(t - X'(\beta-\beta_0))w(t - X'(\beta-\beta_0))\big]\,dt\bigg\}, \\
&\ddot{l}_{\zeta\zeta}(\beta,\zeta(\cdot,\beta);Z)[h_1(\cdot,\beta),h_2(\cdot,\beta)] \\
&\quad = -\int_a^b 1(\epsilon_0 \geq t)\exp\{g(t - X'(\beta-\beta_0))\} \\
&\qquad\quad \cdot w_1(t - X'(\beta-\beta_0))w_2(t - X'(\beta-\beta_0))\,dt,
\end{aligned}
$$

where $h \in \mathbb{H} = \{h : h(\cdot,\beta) = \frac{\partial\zeta_\eta(\cdot,\beta)}{\partial\eta}|_{\eta=0} = w(\psi(\cdot,\beta)), \zeta_\eta \in \mathcal{H}^p\}$. All the above derivatives are continuous and bounded by Conditions (C.1)-(C.3) and (C.6).

1.2. *Proof of Lemma 7.2.* This is a direct result of Corollary 6.21 in [2], that is, there exists a $g_{0,n} \in \mathcal{G}_n^p$ such that $\zeta_{0,n}(t,x,\beta_0) = g_{0,n}(t)$ and

$$\|\zeta_{0,n}(\cdot,\beta_0) - \zeta_0(\cdot,\beta_0)\|_\infty = \|g_{0,n} - g_0\|_\infty = O(q_n^{-p}) = O(n^{-p\nu}).$$

1.3. *Proof of Lemma 7.3.* By the calculation in [3] on page 597, denote the ceiling of x by $\lceil x \rceil$, then for any $\varepsilon > 0$, there exists a set of brackets $\{[g_i^L, g_i^U] : i = 1, 2, \cdots, \lceil (1/\varepsilon)^{c_1 q_n} \rceil\}$ such that for any $g \in \mathcal{G}_n^p$, $g_i^L(t) \leq g(t) \leq g_i^U(t)$ for some $1 \leq i \leq \lceil (1/\varepsilon)^{c_1 q_n} \rceil$ and all $t \in [a,b]$, where $\|g_i^U - g_i^L\|_\infty \leq \varepsilon$. Since $\mathcal{B} \subseteq \mathbb{R}^d$ is compact, $\mathcal{B}$ can be covered by $\lceil c_2(1/\varepsilon)^d \rceil$ balls with radius ε; that is, for any $\beta \in \mathcal{B}$, there exist β_s, $1 \leq s \leq \lceil c_2(1/\varepsilon)^d \rceil$, such that $|\beta - \beta_s| \leq \varepsilon$, i.e., $|(\beta-\beta_0) - (\beta_s - \beta_0)| \leq \varepsilon$, and hence $|x'(\beta-\beta_0) - x'(\beta_s-\beta_0)| \leq C\varepsilon$ for any $x \in \mathcal{X}$ because of Condition (C.2)(a), where $C > 0$ is a constant. This indicates that $t - x'(\beta-\beta_0) \in [t - x'(\beta_s-\beta_0) - C\varepsilon, t - x'(\beta_s-\beta_0) + C\varepsilon]$ for any x and t. Assume $g_i^L(t - x'(\beta_s-\beta_0) + c_1^{i,t}\varepsilon)$ and $g_i^U(t - x'(\beta_s-\beta_0) + c_2^{i,t}\varepsilon)$ are the minimum and maximum values of g_i^L and g_i^U within the interval $[t - x'(\beta_s-\beta_0) - C\varepsilon, t - x'(\beta_s-\beta_0) + C\varepsilon]$, where $c_1^{i,t}$ and $c_2^{i,t}$ are two constants that only depend on g_i^L, g_i^U and t with $|c_1^{i,t}|, |c_2^{i,t}| \leq C$. So we have

$$
\begin{aligned}
g_i^L(t - x'(\beta_s-\beta_0) + c_1^{i,t}\varepsilon) &\leq g_i^L(t - x'(\beta-\beta_0)) \\
&\leq g(t - x'(\beta-\beta_0)) \leq g_i^U(t - x'(\beta-\beta_0)) \\
&\leq g_i^U(t - x'(\beta_s-\beta_0) + c_2^{i,t}\varepsilon).
\end{aligned}
$$

Hence we can construct a set of brackets

$$\big\{[m_{i,s}^L(Z), m_{i,s}^U(Z)] : i = 1, \cdots, \lceil (1/\varepsilon)^{c_1 q_n} \rceil;\; s = 1, \cdots, \lceil c_2(1/\varepsilon)^d \rceil\big\}$$

such that for any $m(\theta; Z) \in \mathcal{F}_n$, there exists a pair (i, s) such that for any sample point Z, $m(\theta; Z) \in [m_{i,s}^L(Z), m_{i,s}^U(Z)]$, where

$$\begin{aligned} m_{i,s}^L(Z) &= \Big\{ \Delta g_i^L(\epsilon_0 - X'(\beta_s - \beta_0) + c_1^{i,\epsilon_0}\varepsilon) \\ &\quad - \int_a^b 1(\epsilon_0 \geq t) \exp\{g_i^U(t - x'(\beta_s - \beta_0) + c_2^{i,t}\varepsilon)\}\, dt \Big\} \\ &\quad - l(\theta_{0,n}; Z), \end{aligned}$$

and

$$\begin{aligned} m_{i,s}^U(Z) &= \Big\{ \Delta g_i^U(\epsilon_0 - X'(\beta_s - \beta_0) + c_2^{i,\epsilon_0}\varepsilon) \\ &\quad - \int_a^b 1(\epsilon_0 \geq t) \exp\{g_i^L(t - x'(\beta_s - \beta_0) + c_1^{i,t}\varepsilon)\}\, dt \Big\} \\ &\quad - l(\theta_{0,n}; Z). \end{aligned}$$

It then follows that

$$\begin{aligned} &|m_{i,s}^U(Z) - m_{i,s}^L(Z)| \\ &\quad \leq |g_i^U(\epsilon_0 - X'(\beta_s - \beta_0) + c_2^{i,\epsilon_0}\varepsilon) - g_i^L(\epsilon_0 - X'(\beta_s - \beta_0) + c_1^{i,\epsilon_0}\varepsilon)| \\ &\qquad + \int_a^b |\exp\{g_i^U(t - X'(\beta_s - \beta_0) + c_2^{i,t}\varepsilon)\} \\ &\qquad\qquad - \exp\{g_i^L(t - x'(\beta_s - \beta_0) + c_1^{i,t}\varepsilon)\}|\, dt \\ &\quad = A_1 + A_2. \end{aligned}$$

For A_1, by subtracting and adding the terms $g(\epsilon_0 - X'(\beta_s - \beta_0) + c_2^{i,\epsilon_0}\varepsilon)$ and $g(\epsilon_0 - X'(\beta_s - \beta_0) + c_1^{i,\epsilon_0}\varepsilon)$ and applying the Taylor expansion to g at $\epsilon_0 - X'(\beta_s - \beta_0) + c_1^{i,\epsilon_0}\varepsilon$, we have

$$\begin{aligned} A_1 &\leq |g_i^U(\epsilon_0 - X'(\beta_s - \beta_0) + c_2^{i,\epsilon_0}\varepsilon) - g(\epsilon_0 - X'(\beta_s - \beta_0) + c_2^{i,\epsilon_0}\varepsilon)| \\ &\quad + |g(\epsilon_0 - X'(\beta_s - \beta_0) + c_2^{i,\epsilon_0}\varepsilon) - g(\epsilon_0 - X'(\beta_s - \beta_0) + c_1^{i,\epsilon_0}\varepsilon)| \\ &\quad + |g(\epsilon_0 - X'(\beta_s - \beta_0) + c_1^{i,\epsilon_0}\varepsilon) - g_i^L(\epsilon_0 - X'(\beta_s - \beta_0) + c_1^{i,\epsilon_0}\varepsilon)| \\ &\leq \|g_i^U - g\|_\infty + |\dot{g}(\epsilon_0 - X'(\beta_s - \beta_0) + \tilde{c}\varepsilon)(c_2^{i,\epsilon_0} - c_1^{i,\epsilon_0})\varepsilon| + \|g - g_i^L\|_\infty \\ &\leq \|g_i^U - g_i^L\|_\infty + C_1|(c_2^{i,\epsilon_0} - c_1^{i,\epsilon_0})|\varepsilon + \|g_i^U - g_i^L\|_\infty \\ &\leq 2\varepsilon + 2C_1C_2\varepsilon \lesssim \varepsilon, \end{aligned}$$

where the third inequality holds because $\|g_i^U - g\|_\infty, \|g - g_i^L\|_\infty \leq \|g_i^U - g_i^L\|_\infty$ and $\dot{g}$ is bounded by C_1. The Constant C_1 may be proportional to c_n that is allowed to grow with n slowly enough, but it does not affect the later calculations on convergence rate (see [3], page 591, for their constant l_n), thus we drop c_n for simplicity. For A_2, by using the similar arguments as for A_1 and denote $t - X'(\beta_s - \beta_0) = t_s$ for notational simplicity, we have

$$\begin{aligned}
A_2 &\leq \int_a^b \{|\exp\{g_i^U(t_s + c_2^{i,t}\varepsilon)\} - \exp\{g(t_s + c_2^{i,t}\varepsilon)\}| \\
&\quad + |\exp\{g(t_s + c_2^{i,t}\varepsilon)\} - \exp\{g(t_s + c_1^{i,t}\varepsilon)\}| \\
&\quad + |\exp\{g(t_s + c_1^{i,t}\varepsilon)\} - \exp\{g_i^L(t_s + c_1^{i,t}\varepsilon)\}|\}\, dt \\
&= \int_a^b \{|\exp\{\tilde{g}_i^U(t_s + c_2^{i,t}\varepsilon)\}(g_i^U - g)(t_s + c_2^{i,t}\varepsilon)| \\
&\quad + |\exp\{g(t_s + \tilde{c}\varepsilon)\}(c_2^{i,t} - c_1^{i,t})\varepsilon| \\
&\quad + |\exp\{\tilde{g}_i^L(t_s + c_1^{i,t}\varepsilon)\}(g_i^L - g)(t_s + c_1^{i,t}\varepsilon)|\}\, dt \\
&\lesssim \|g_i^U - g\|_\infty + |(c_2^{i,t} - c_1^{i,t})\varepsilon| + \|g - g_i^L\|_\infty \lesssim \varepsilon.
\end{aligned}$$

The above equality is from Taylor expansion, where $\tilde{g}_i^U = g + \xi(g_i^U - g)$ for some $0 < \xi < 1$ and thus $|\tilde{g}_i^U(\cdot)| \leq |g(\cdot)| + \varepsilon$, which is bounded in $[a, b]$; similarly $|\tilde{g}_i^L|$ is also bounded in $[a, b]$. Hence $\|m_i^U - m_i^L\|_\infty \lesssim \varepsilon$ and the ε-bracketing number associated with $\|\cdot\|_\infty$ norm for the class $\mathcal{F}_n$ follows

$$N_{[]}(\varepsilon, \mathcal{F}_n, \|\cdot\|_\infty) \leq (1/\varepsilon)^{c_1 q_n} c_2 (1/\varepsilon)^d \lesssim (1/\varepsilon)^{c_1 q_n + d}.$$

1.4. *Proof of Lemma 7.4.* In the proof of Theorem 4.2 in the main text of the paper we show that such defined $(h_1^*, \ldots, h_d^*)$ determines the least favorable submodel for β. Now, by Conditions (C.4)-(C.5), the following conditional density of ϵ_0 given X

$$f_{\epsilon_0|X}(t|X = x) = f(t)\bar{G}_{C|X}(t + x'\beta_0|X = x) + g_{C|X}(t + x'\beta_0|X = x)\bar{F}(t)$$

is uniformly bounded for all $x \in \mathcal{X}$, and its derivative with respect to t

$$\begin{aligned}
&\dot{f}_{\varepsilon_0|X}(t|X = x) \\
&\quad = \dot{f}(t)\bar{G}_{C|X}(t + x'\beta_0|X = x) - f(t)g_{C|X}(t + x'\beta_0|X = x) \\
&\qquad + \dot{g}_{C|X}(t + x'\beta_0|X = x)\bar{F}(t) - g_{C|X}(t + x'\beta_0|X = x)f(t)
\end{aligned}$$

is also uniformly bounded. Hence the density of ϵ_0

$$f_{\epsilon_0}(t) = \int_{\mathcal{X}} f_{\epsilon_0|X}(t|X = x)f_X(x)\, dx$$

and its derivative

$$\dot{f}_{\epsilon_0}(t) = \int_{\mathcal{X}} \dot{f}_{\epsilon_0|X}(t|X=x) f_X(x)\ dx$$

are bounded. Thus the first and second derivatives of $P(\epsilon_0 \geq t)$, i.e., $-f_{\epsilon_0}(t)$ and $-\dot{f}_{\epsilon_0}(t)$, are both bounded. In addition, under Condition (C.2)(a), the first and second derivatives of $P[X1(\epsilon_0 \geq t)]$ with respect to t

$$\frac{dP[X1(\epsilon_0 \geq t)]}{dt} = -\int_{\mathcal{X}} x f_X(x) f_{\epsilon_0|X}(t|X=x)\ dx$$

and

$$\frac{d^2 P[X1(\epsilon_0 \geq t)]}{dt^2} = -\int_{\mathcal{X}} x f_X(x) \dot{f}_{\epsilon_0|X}(t|X=x)\ dx$$

are also bounded. Therefore, $P[X|\epsilon_0 \geq t] = P[X1(\epsilon_0 \geq t)]/P(\epsilon_0 \geq t)$ has a bounded second derivative with respect to t for $t \leq \tau$, where τ is the truncation time defined in Condition (C.3). Thus $P[X|\epsilon_0 \geq t] \in \mathcal{G}^2$. Moreover, since $g_0 \in \mathcal{G}^p$ for $p \geq 3$, we have $\dot{g}_0 \in \mathcal{G}^{p-1}$ with $p-1 \geq 2$. Thus according to Corollary 6.21 of [2], there exists an $h^*_{j,n} \in \mathcal{H}_n^{\min(p-1,2)} = \mathcal{H}_n^2$ such that $h^*_{j,n}(t,x,\beta_0) = w^*_{j,n}(\psi(t,x,\beta_0)) = w^*_{j,n}(t)$ and $\|h^*_{j,n}(\cdot,\beta_0) - h^*_j(\cdot,\beta_0)\|_\infty = \|w^*_{j,n} - w^*_j\|_\infty = O(q_n^{-2}) = O(n^{-2\nu})$.

1.5. *Proof of Lemma 7.5.* The proof is similar to the bracketing number calculation in Lemma 7.3, thus omitted. We refer all the details to [1].

1.6. *Proof of Lemma 7.6.* The proof is also similar to the bracketing number calculation in Lemma 7.3, thus omitted. We again refer all the details to [1].

2. Proof of Theorem 4.1. We shall apply Theorem 1 of [3] to derive the convergence rate. We proceed by verifying their conditions C1-C3. Since $Pl(\beta,\zeta(\cdot,\beta);Z)$ is maximized at $(\beta_0,\zeta_0(\cdot,\beta_0))$, its first derivatives at $(\beta_0,\zeta_0(\cdot,\beta_0))$ are equal to 0. By Lemma 7.1 that all the second derivatives of $l(\beta,\zeta(\cdot,\beta);Z)$ are continuous and bounded, the Taylor expansion yields

$$\begin{aligned}
(2.1)\quad & Pl(\beta,\zeta(\cdot,\beta);Z) - Pl(\beta_0,\zeta_0(\cdot,\beta_0);Z) \\
&= \frac{1}{2} P\big\{(\beta-\beta_0)'\ddot{l}_{\beta\beta}(\beta_0,\zeta_0(\cdot,\beta_0);Z)(\beta-\beta_0) \\
&\quad + 2(\beta-\beta_0)'\ddot{l}_{\beta\zeta}(\beta_0,\zeta_0(\cdot,\beta_0);Z)[\zeta(\cdot,\beta)-\zeta_0(\cdot,\beta_0)] \\
&\quad + \ddot{l}_{\zeta\zeta}(\beta_0,\zeta_0(\cdot,\beta_0);Z)[\zeta(\cdot,\beta)-\zeta_0(\cdot,\beta_0),\zeta(\cdot,\beta)-\zeta_0(\cdot,\beta_0)]\big\} \\
&\quad + o(d^2(\theta,\theta_0)) \\
&= A + o(d^2(\theta,\theta_0)),
\end{aligned}$$

where $\theta = (\beta, \zeta(\cdot,\beta)) \in \Theta_n^p$. By the model assumption, we have that the conditional expectation $P\{\dot{l}_\zeta(\beta_0, \zeta_0(\cdot,\beta_0); Z)[h(\cdot,\beta_0)]|X\} = 0$ for all $h \in \mathbb{H}$. Taking h to be $\ddot{g}_0$ and $\dot{g}(\psi(\cdot,\beta)) - \dot{g}_0(\psi(\cdot,\beta_0))$ respectively, we have

$$P\left\{\Delta\ddot{g}_0(\epsilon_0) - \int_a^b 1(\epsilon_0 \geq t)\exp\{g_0(t)\}\ddot{g}_0(t)\,dt \middle| X\right\} = 0$$

and

$$\begin{aligned} P\Bigg\{ & \Delta[\dot{g}(\epsilon_0 - X'(\beta-\beta_0)) - \dot{g}_0(\epsilon_0)] \\ & - \int_a^b 1(\epsilon_0 \geq t)\exp\{g_0(t)\}[\dot{g}(t - X'(\beta-\beta_0)) - \dot{g}_0(t)]\,dt \Bigg| X\Bigg\} = 0. \end{aligned}$$

Then it follows

$$\begin{aligned} A &= P\Bigg\{\frac{1}{2}\int_a^b 1(\epsilon_0 \geq t)\exp\{g_0(t)\}\big\{-[\dot{g}_0(t)X'(\beta-\beta_0)]^2 \\ &\qquad + 2\dot{g}_0(t)X'(\beta-\beta_0)[g(t - X'(\beta-\beta_0)) - g_0(t)] \\ &\qquad - [g(t - X'(\beta-\beta_0)) - g_0(t)]^2\big\}\,dt\Bigg\} \\ &= -\frac{1}{2}\int_a^b \exp\{g_0(t)\}P\big\{1(\epsilon_0 \geq t)\big[\dot{g}_0(t)X'(\beta-\beta_0) \\ &\qquad - (g(t_\beta) - g_0(t))\big]^2\big\}\,dt, \end{aligned} \tag{2.2}$$

where $t_\beta = t - X'(\beta-\beta_0)$. The integrand is from a to $\tau = b$ because of Condition (C.3) and (C.6). Denote $s_0(t) = -\dot{g}_0(t)$, $s_1(t;\epsilon_0,X) = 1(\epsilon_0 \geq t)X'(\beta-\beta_0)$, and $s_2(t;\epsilon_0,X) = 1(\epsilon_0 \geq t)[g(t_\beta) - g_0(t)]$, then

$$\begin{aligned} & P\big\{1(\epsilon_0 \geq t)[\dot{g}_0(t)X'(\beta-\beta_0) - (g(t_\beta) - g_0(t))]^2\big\} \\ &\quad = P\big\{[s_0(t)s_1(t;\epsilon_0,X) + s_2(t;\epsilon_0,X)]^2\big\} \\ &\quad \geq s_0^2(t)P[s_1^2(t;\epsilon_0,X)] + P[s_2^2(t;\epsilon_0,X)] \\ &\qquad\quad - 2|s_0(t)P[s_1(t;\epsilon_0,X)s_2(t;\epsilon_0,X)]| \\ &\quad \geq s_0^2(t)P[s_1^2(t;\epsilon_0,X)] + P[s_2^2(t;\epsilon_0,X)] \\ &\qquad\quad - (1-\eta)^{\frac{1}{2}} \cdot 2|s_0(t)[P(s_1^2(t;\epsilon_0,X))]^{\frac{1}{2}}| \cdot |[P(s_2^2(t;\epsilon_0,X))]^{\frac{1}{2}}| \\ &\quad \geq \{1 - (1-\eta)^{\frac{1}{2}}\}\{s_0^2(t)P(s_1^2(t;\epsilon_0,X)) + P(s_2^2(t;\epsilon_0,X))\} \\ &\quad \gtrsim \dot{g}_0^2(t)(\beta-\beta_0)'P[1(\epsilon_0 \geq t)XX'](\beta-\beta_0) \\ &\qquad\quad + P[1(\epsilon_0 \geq t)(g(t_\beta) - g_0(t))^2], \end{aligned} \tag{2.3}$$

where (2.3) is obtained by using the same argument in [4] on page 2126, which is, under Condition (C.7), $[P(s_1 s_2)]^2 \leq (1-\eta)P(s_1^2)P(s_2^2)$ for some $\eta \in (0,1)$. Hence from (2.2) we have

$$\begin{aligned} A &\lesssim -\left\{(\beta-\beta_0)'\left[\int_a^b \exp\{g_0(t)\}\dot{g}_0^2(t)P[1(\epsilon_0 \geq t)XX']\,dt\right](\beta-\beta_0)\right. \\ &\quad \left. + \int_a^b \exp\{g_0(t)\}P[1(\epsilon_0 \geq t)(g(t_\beta)-g_0(t))^2]\,dt\right\} \\ &= -(A_1+A_2). \end{aligned}$$

For A_1, Condition (C.3) implies that

$$P[1(\epsilon_0 \geq t)XX'] = P[XX'P(\epsilon_0 \geq t|X)] \geq P[XX'P(\epsilon_0 \geq \tau|X)] \geq \delta P(XX').$$

Then Condition (C.2)(b) yields that $P(XX')$ is positive definite and thus its smallest eigenvalue $\lambda_1 > 0$. In addition, $\int_a^b \exp\{g_0(t)\}\dot{g}_0^2(t)dt$ is bounded away from zero since $\exp\{g_0(t)\}, \dot{g}_0^2(t) \geq 0$ but not a constant zero on $t \in [a,b]$. Hence it follows that

$$A_1 \gtrsim (\beta-\beta_0)'P(XX')(\beta-\beta_0) \geq \lambda_1|\beta-\beta_0|^2 \gtrsim |\beta-\beta_0|^2.$$

For A_2, Condition (C.3) yields

$$\begin{aligned} A_2 &\geq P(\epsilon_0 \geq b)\int_a^b P(g(t-X'(\beta-\beta_0))-g_0(t))^2\, d\Lambda_0(t) \\ &\gtrsim \|\zeta(\cdot,\beta)-\zeta_0(\cdot,\beta_0)\|_2^2. \end{aligned}$$

Therefore

$$A \lesssim -\left\{|\beta-\beta_0|^2 + \|\zeta(\cdot,\beta)-\zeta_0(\cdot,\beta_0)\|_2^2\right\} = -d^2(\theta,\theta_0),$$

and thus from (2.1),

$$Pl(\beta,\zeta(\cdot,\beta);Z) - Pl(\beta_0,\zeta_0(\cdot,\beta_0);Z) \lesssim -d^2(\theta,\theta_0) + o(d^2(\theta,\theta_0)) \lesssim -d^2(\theta,\theta_0),$$

i.e. $P(l(\theta_0;Z)-l(\theta;Z)) \gtrsim d^2(\theta,\theta_0)$ for all $\theta \in \Theta_n^p$, which implies that

$$\inf_{\{d(\theta,\theta_0)\geq\varepsilon,\theta\in\Theta_n^p\}} P(l(\theta_0;Z)-l(\theta;Z)) \gtrsim \varepsilon^2.$$

Hence condition C1 of [3] on page 583 holds with the constant $\alpha = 1$ in their notation.

Next we verify condition C2 of [3]. Denote $\epsilon_\beta = Y - X'\beta$. It follows that

$$\begin{aligned}
&[l(\theta; Z) - l(\theta_0; Z)]^2 \\
&\quad = \left\{ \Delta g(\epsilon_\beta) - \int_a^b 1(\epsilon_0 \geq t) e^{g(t_\beta)}\, dt - \Delta g_0(\epsilon_0) + \int_a^b 1(\epsilon_0 \geq t) e^{g_0(t)}\, dt \right\}^2 \\
&\quad \lesssim \Delta[g(\epsilon_\beta) - g_0(\epsilon_0)]^2 + \left\{ \int_a^b 1(\epsilon_0 \geq t)[e^{g(t_\beta)} - e^{g_0(t)}]\, dt \right\}^2 \\
&\quad \lesssim \Delta[g(\epsilon_\beta) - g_0(\epsilon_0)]^2 + \int_a^b [e^{g(t_\beta)} - e^{g_0(t)}]^2\, dt \\
&\quad \lesssim \Delta[g(\epsilon_\beta) - g(\epsilon_0)]^2 + \Delta[g(\epsilon_0) - g_0(\epsilon_0)]^2 \\
&\qquad\quad + \int_a^b [e^{g(t_\beta)} - e^{g(t)}]^2\, dt + \int_a^b [e^{g(t)} - e^{g_0(t)}]^2\, dt \\
&\quad = I_1 + I_2 + I_3 + I_4,
\end{aligned}$$

where the second inequality holds because of the Cauchy-Schwartz inequality

$$\begin{aligned}
&\left\{ \int_a^b 1(\epsilon_0 \geq t)[e^{g(t_\beta)} - e^{g_0(t)}]\, dt \right\}^2 \\
&\qquad \leq \left\{ \int_a^b 1(\epsilon_0 \geq t)\, dt \right\} \left\{ \int_a^b [e^{g(t_\beta)} - e^{g_0(t)}]^2\, dt \right\} \\
&\qquad \leq (b-a) \int_a^b [e^{g(t_\beta)} - e^{g_0(t)}]^2\, dt,
\end{aligned}$$

and the third inequality holds by subtracting and adding the terms $g(\epsilon_0)$ and $e^{g(t)}$. For I_1, since $\dot{g} \in \mathcal{G}_n^{p-1}$ is bounded, applying the Taylor expansion for g at ϵ_0 we obtain

$$\begin{aligned}
PI_1 &= P\big\{\Delta[g(\epsilon_0 - X'(\beta - \beta_0)) - g(\epsilon_0)]^2\big\} \\
&\leq P[\dot{g}(\epsilon_0 - X'(\tilde{\beta} - \beta_0))X'(\beta - \beta_0)]^2 \\
&\lesssim P[X'(\beta - \beta_0)]^2 = (\beta - \beta_0)' P(XX')(\beta - \beta_0) \\
&\leq \lambda_d |\beta - \beta_0|^2 \lesssim |\beta - \beta_0|^2,
\end{aligned}$$

where λ_d is the largest eigenvalue of $P(XX')$. For I_2, since the density function for $(Y, \Delta = 1, X)$ is

$$f_{Y,\Delta,X}(y, 1, x) = \lambda_0(y - x'\beta_0) e^{-\Lambda_0(y - x'\beta_0)} \bar{G}_{C|X}(y|X = x) f_X(x),$$

it follows that

$$\begin{aligned}
PI_2 &= P[\Delta(g-g_0)^2(\epsilon_0)] \\
&= \int_{\mathcal{X}}\left\{\int_a^b (g(t)-g_0(t))^2\lambda_0(t)e^{-\Lambda_0(t)}\bar{G}_{C|X}(t+x'\beta_0|X=x)\,dt\right\} \\
&\quad f_X(x)\,dx \\
&\leq \int_{\mathcal{X}}\left\{\int_a^b (g(t)-g_0(t))^2 d\Lambda_0(t)\right\} f_X(x)\,dx \\
&= \|\zeta(\cdot,\beta_0)-\zeta_0(\cdot,\beta_0)\|_2^2.
\end{aligned}$$

Then for I_3, since $g\in\mathcal{G}_n^p$ is bounded, it follows that

$$\begin{aligned}
PI_3 &= P\int_a^b [e^{g(t-X'(\beta-\beta_0))}-e^{g(t)}]^2\,dt \\
&= P\int_a^b e^{2g(t-X'(\tilde{\beta}-\beta_0))}[X'(\beta-\beta_0)]^2\,dt \\
&\lesssim \int_a^b P[X'(\beta-\beta_0)]^2\,dt \\
&\lesssim (\beta-\beta_0)'P[XX'](\beta-\beta_0)\lesssim|\beta-\beta_0|^2.
\end{aligned}$$

Finally for I_4, by the Taylor expansion for $e^{g(t)}$ at g_0, we have

$$\begin{aligned}
PI_4 &= \int_a^b [e^{g(t)}-e^{g_0(t)}]^2\,dt \\
&\leq \int_a^b e^{2\tilde{g}(t)}(g(t)-g_0(t))^2\,dt \\
&= \int_a^b e^{2\tilde{g}(t)-g_0(t)}(g(t)-g_0(t))^2\ d\Lambda_0(t) \\
&\lesssim \int_a^b (g(t)-g_0(t))^2\ d\Lambda_0(t)=\|\zeta(\cdot,\beta_0)-\zeta_0(\cdot,\beta_0)\|_2^2,
\end{aligned}$$

where $\tilde{g}(t)=g_0(t)+\xi(g-g_0)(t)$ for some $0<\xi<1$ and hence is bounded. Since $\|\zeta(\cdot,\beta_0)-\zeta_0(\cdot,\beta_0)\|_2\leq|\zeta(\cdot,\beta)-\zeta(\cdot,\beta_0)\|_2+\|\zeta(\cdot,\beta)-\zeta_0(\cdot,\beta_0)\|_2\lesssim|\beta-\beta_0|+\|\zeta(\cdot,\beta)-\zeta_0(\cdot,\beta_0)\|_2$, we have

$$P(l(\theta;Z)-l(\theta_0;Z))^2\lesssim|\beta-\beta_0|^2+\|\zeta(\cdot,\beta)-\zeta_0(\cdot,\beta_0)\|_2^2=d^2(\theta,\theta_0)$$

for any $\theta\in\Theta_n^p$, which implies that

$$\begin{aligned}
&\sup_{\{d(\theta,\theta_0)\leq\varepsilon,\theta\in\Theta_n^p\}}\mathrm{Var}(l(\theta_0;Z)-l(\theta;Z)) \\
&\qquad\leq\sup_{\{d(\theta,\theta_0)\leq\varepsilon,\theta\in\Theta_n^p\}}P(l(\theta_0;Z)-l(\theta;Z))^2\lesssim\varepsilon^2.
\end{aligned}$$

So condition C2 of [3] on page 583 holds with the constant $\beta = 1$ in their notation.

Finally, we verify condition C3 of [3]. By lemma 7.3, for $\mathcal{F}_n = \{l(\theta; Z) - l(\theta_{0,n}; Z) : \theta \in \Theta_n^p\}$, we have $N_{[]}(\varepsilon, \mathcal{F}_n, \|\cdot\|_\infty) \lesssim (1/\varepsilon)^{cq_n+d}$. Then by the fact that the covering number is bounded by the bracketing number, it follows that

$$H(\varepsilon, \mathcal{F}_n, \|\cdot\|_\infty) = \log N(\varepsilon, \mathcal{F}_n, \|\cdot\|_\infty) \lesssim (cq_n + d)\log(1/\varepsilon) \lesssim n^\nu \log(1/\varepsilon).$$

So condition C3 of [3] on page 583 holds with the constants $2r_0 = \nu$ and $r = 0^+$ in their notation.

Therefore, the constant τ in Theorem 1 of [3] on page 584 is $\frac{1-\nu}{2} - \frac{\log\log n}{2\log n}$. Since $\frac{\log\log n}{2\log n} \to 0$ as $n \to 0$, we can pick a $\tilde{\nu}$ slightly greater than ν such that $\frac{1-\tilde{\nu}}{2} \leq \frac{1-\nu}{2} - \frac{\log\log n}{2\log n}$ for n large. We still denote $\tilde{\nu}$ by ν and then $\tau = \frac{1-\nu}{2}$. Since $\hat{\theta}_n$ maximizes the empirical log-likelihood $\mathbb{P}_n l(\theta; Z)$ over the sieve space Θ_n^p, we have that $\hat{\theta}_n$ satisfies inequality (1.1) in [3] with $\eta_n = 0$. By Lemma 7.2, there exists an $\zeta_{0,n}(\cdot, \beta_0) \in \mathcal{H}_n^p$ such that $\|\zeta_{0,n}(\cdot,\beta_0) - \zeta_0(\cdot,\beta_0)\|_\infty = O(n^{-p\nu})$. Moreover, by the Taylor expansion for $P[l(\beta_0, \zeta_0(\cdot,\beta_0); Z) - l(\beta, \zeta(\cdot,\beta); Z)]$ in (2.1) and plugging in $\theta = \theta_{0,n} = (\beta_0, \zeta_{0,n}(\cdot,\beta_0))$, the Kullback-Leibler distance between $\theta_{0,n} = (\beta_0, \zeta_{0,n}(\cdot,\beta_0))$ and $\theta_0 = (\beta_0, \zeta_0(\cdot,\beta_0))$ is given as

$$\begin{aligned}
& K(\theta_{0,n}, \theta_0) \\
= \;& P[l(\theta_0; Z) - l(\theta_{0,n}; Z)] \\
= \;& -P\{\ddot{l}_{\zeta\zeta}(\beta_0, \zeta_0(\cdot,\beta_0); Z)[\zeta_{0,n}(\cdot,\beta_0) - \zeta_0(\cdot,\beta_0), \zeta_{0,n}(\cdot,\beta_0) - \zeta_0(\cdot,\beta_0)]\} \\
& \quad + o(\|\zeta_{0,n}(\cdot,\beta_0) - \zeta_0(\cdot,\beta_0)\|_2^2) \\
= \;& P\left\{\int_a^b 1(\epsilon_0 \geq t)\exp\{g_0(t)\}(g_{0,n}(t) - g_0(t))^2\, dt\right\} \\
& \quad + o(\|\zeta_{0,n}(\cdot,\beta_0) - \zeta_0(\cdot,\beta_0)\|_2^2) \\
\leq \;& \int_a^b (g_{0,n}(t) - g_0(t))^2\, d\Lambda_0(t) + o(\|\zeta_{0,n}(\cdot,\beta_0) - \zeta_0(\cdot,\beta_0)\|_2^2) \\
= \;& \|\zeta_{0,n}(\cdot,\beta_0) - \zeta_0(\cdot,\beta_0)\|_2^2 + o(\|\zeta_{0,n}(\cdot,\beta_0) - \zeta_0(\cdot,\beta_0)\|_2^2) = O(n^{-2p\nu}),
\end{aligned}$$

where the last equality holds because $\|\zeta_{0,n}(\cdot,\beta_0) - \zeta_0(\cdot,\beta_0)\|_2 \leq \|\zeta_{0,n}(\cdot,\beta_0) - \zeta_0(\cdot,\beta_0)\|_\infty = O(n^{-p\nu})$. Therefore $K^{1/2}(\theta_{0,n}, \theta_0) = O(n^{-p\nu})$. Thus by Theorem 1 of [3], we obtain the convergence rate for $\hat{\theta}_n$ as follows

$$d(\hat{\theta}_n, \theta_0) = O_p\{\max(n^{-(1-\nu)/2}, n^{-p\nu}, n^{-p\nu})\} = O_p\{n^{-\min(p\nu,(1-\nu)/2)}\}.$$

3. Proof of Theorem 4.3. Define $\mathbf{w}_n^*(t) = -\dot{\hat{g}}_n(t)\bar{X}(t;\hat{\beta}_n)$. Then we have

$$l^*_{\hat{\beta}_n}(Y,\Delta,X) = \dot{l}_\beta(\hat{\theta}_n;Z) - \dot{l}_\zeta(\hat{\theta}_n;Z)[\mathbf{h}_n^*].$$

Define

$$\begin{aligned} I^{jk}(\beta_0) &= P\left[\left\{\dot{l}_{\beta_j}(\theta_0;Z) - \dot{l}_\zeta(\theta_0;Z)[h_j^*]\right\}\right. \\ &\qquad \left.\times\left\{\dot{l}_{\beta_k}(\theta_0;Z) - \dot{l}_\zeta(\theta_0;Z)[h_k^*]\right\}\right] \equiv PA^{jk}(\theta_0;Z), \\ \hat{I}_n^{jk}(\hat{\beta}_n) &= \mathbb{P}_n\left[\left\{\dot{l}_{\beta_j}(\hat{\theta}_n;Z) - \dot{l}_\zeta(\hat{\theta}_n;Z)[h_{j,n}^*]\right\}\right. \\ &\qquad \left.\times\left\{\dot{l}_{\beta_k}(\hat{\theta}_n;Z) - \dot{l}_\zeta(\hat{\theta}_n;Z)[h_{k,n}^*]\right\}\right] \equiv \mathbb{P}_n A_n^{jk}(\hat{\theta}_n;Z), \end{aligned}$$

where h_j^* is defined in Lemma 7.4, see also Equation (7.1) in the main text. We will prove $\mathbb{P}_n A_n^{jk}(\hat{\theta}_n;Z) \to PA^{jk}(\theta_0;Z)$ in probability for all $j,k = 1,\ldots,d$. Let

$$\begin{aligned} \mathbb{P}_n A_n^{jk}(\hat{\theta}_n;Z) - PA^{jk}(\theta_0;Z) &= (\mathbb{P}_n - P)A_n^{jk}(\hat{\theta}_n;Z) \\ &\qquad + P\left\{A_n^{jk}(\hat{\theta}_n;Z) - A^{jk}(\theta_0;Z)\right\} \\ &= I_{1n} + I_{2n}. \end{aligned}$$

For I_{1n}, we first define the class of functions

$$\begin{aligned} \mathcal{F}_{n,j}^{\beta,\zeta}(\eta) &= \{\dot{l}_{\beta_j}(\theta;z) - \dot{l}_\zeta(\theta;z)[h_j] : \theta \in \Theta_n^p, d(\theta,\theta_0) \le \eta, h_j \in \mathcal{H}_n^2, \\ &\qquad \|\dot{g}(\psi(\cdot,\beta)) - \dot{g}_0(\psi(\cdot,\beta_0))\|_2 \le \eta, \|h_j - h_j^*\|_\infty \le \eta\}. \end{aligned}$$

Then by Lemmas 7.5 and 7.6 we have

$$N_{[]}(\varepsilon, \mathcal{F}_{n,j}^{\beta,\zeta}, \|\cdot\|_\infty) \lesssim (1/\varepsilon)^{cq_n+d}$$

for some constant $c > 0$. This is because for any function $\dot{l}_{\beta_j}(\theta;z) - \dot{l}_\zeta(\theta;z)[h_j] \in \mathcal{F}_{n,j}^{\beta,\zeta}(\eta)$, it can be written as

$$\begin{aligned} &\dot{l}_{\beta_j}(\theta;z) - \dot{l}_\zeta(\theta;z)[h_j] \\ &\quad = \{\dot{l}_{\beta_j}(\theta;z) - \dot{l}_{\beta_j}(\theta_0;z)\} - \{\dot{l}_\zeta(\theta;z)[h_j^*] - \dot{l}_\zeta(\theta_0;z)[h_j^*]\} \\ &\qquad + \{\dot{l}_\zeta(\theta;z)[h_j^*] - \dot{l}_\zeta(\theta;z)[h_j]\} + \{\dot{l}_{\beta_j}(\theta_0;z) - \dot{l}_\zeta(\theta_0;z)[h_j^*]\} \\ &\quad = A_1 + A_2 + A_3 + A_4, \end{aligned}$$

where $A_1 \in \mathcal{F}_{n,j}^{\beta}(\eta)$ and $A_2 \in \mathcal{F}_{n,j}^{\zeta}(\eta)$ defined in Lemma 7.6, $A_3 \in \mathcal{F}_n^j(\eta)$ defined in Lemma 7.5, and A_4 is a fixed function (the efficient score function).

Assume $A_m^L \le A_m \le A_m^U$ with $\|A_m^U - A_m^L\|_\infty \lesssim \varepsilon$, $m = 1, 2, 3$. Then $A_m^L + A_{m'}^L \le A_m + A_{m'} \le A_m^U + A_{m'}^U$ with $\|(A_m^U + A_{m'}^U) - (A_m^L + A_{m'}^L)\|_\infty \le \|A_m^U - A_m^L\|_\infty + \|A_{m'}^U - A_{m'}^L\|_\infty \lesssim \varepsilon$. Therefore the ε-bracketing number associated with $\|\cdot\|_\infty$ for $\mathcal{F}_{n,j}^{\beta,\zeta}(\eta)$ is also bounded by $(\eta/\varepsilon)^{cq_n+\alpha}$.

We next define the class of functions

$$\begin{aligned}\mathcal{F}_{n,jk}^{\beta,\zeta}(\eta) \;=\; \{(\dot{l}_{\beta_j}(\theta;z) - \dot{l}_\zeta(\theta;z)[h_j])(\dot{l}_{\beta_k}(\theta;z) - \dot{l}_\zeta(\theta;z)[h_k]) : \theta \in \Theta_n^p,\\ h_j, h_k \in \mathcal{H}_n^{p-1}, d(\theta,\theta_0) \le \eta, \|\dot{g}(\psi(\cdot,\beta)) - \dot{g}_0(\psi(\cdot,\beta_0))\|_2 \le \eta,\\ \|h_j - h_j^*\|_\infty \le \eta, \|h_k - h_k^*\|_\infty \le \eta\}.\end{aligned}$$

Then if $B_j^L \le \dot{l}_{\beta_j}(\theta;z) - \dot{l}_\zeta(\theta;z)[h_j] \le B_j^U$ and $B_k^L \le \dot{l}_{\beta_j}(\theta;z) - \dot{l}_\zeta(\theta;z)[h_k] \le B_k^U$ with $\|B_j^U - B_j^L\|_\infty \le \varepsilon$ and $\|B_k^U - B_k^L\|_\infty \le \varepsilon$, we have $B_j^* B_k^* \le (\dot{l}_{\beta_j}(\theta;z) - \dot{l}_\zeta(\theta;z)[h_j])(\dot{l}_{\beta_k}(\theta;z) - \dot{l}_\zeta(\theta;z)[h_k]) \le B_j^{**} B_k^{**}$, where B_j^*, B_j^{**} take values of either B_j^L or B_j^U, and the same for B_k^*, B_k^{**}. Thus

$$\begin{aligned}\|B_j^{**}B_k^{**} - B_j^*B_k^*\|_\infty &= \|(B_j^{**} - B_j^*)B_k^{**} + (B_k^{**} - B_k^*)B_j^*\|_\infty\\ &= \|B_j^{**} - B_j^*\|_\infty \|B_k^{**}\|_\infty + \|B_k^{**} - B_k^*\|_\infty \|B_j^*\|_\infty\\ &\lesssim \|B_j^U - B_j^L\|_\infty + \|B_k^U - B_k^L\|_\infty\\ &\lesssim \varepsilon,\end{aligned}$$

which yields

$$N_{[\,]}(\varepsilon, \mathcal{F}_{n,jk}^{\beta,\zeta}, \|\cdot\|_\infty) \lesssim (1/\varepsilon)^{cq_n+d}$$

for some constant $c > 0$.

Finally, similar to the verification of Assumption (A4) in the proof of Theorem 4.2 and together with the following fact:

$$\begin{aligned}&\|h_j^*(\cdot,\beta_0) - h_{j,n}^*(\cdot,\hat{\beta}_n)\|_\infty\\ &\quad = \|\dot{g}_0(t)P(X|\epsilon_0 \ge t) - \dot{\hat{g}}_n(t_{\hat{\beta}_n})\bar{X}(t_{\hat{\beta}_n};\hat{\beta}_n)\|_\infty\\ &\quad \le \|\dot{g}_0(t)P(X|\epsilon_0 \ge t) - \dot{\hat{g}}_n(t)\bar{X}(t;\hat{\beta}_n)\|_\infty\\ &\qquad + \|\dot{\hat{g}}_n(t_{\hat{\beta}_n})\bar{X}(t_{\hat{\beta}_n};\hat{\beta}_n) - \dot{\hat{g}}_n(t_{\hat{\beta}_n})\bar{X}(t;\hat{\beta}_n)\|_\infty\\ &\qquad + \|\dot{\hat{g}}_n(t_{\hat{\beta}_n})\bar{X}(t;\hat{\beta}_n) - \dot{\hat{g}}_n(t)\bar{X}(t;\hat{\beta}_n)\|_\infty,\end{aligned} \tag{3.1}$$

where the first term on the right hand side of inequality (3.1) is

$$\begin{aligned}&\|\dot{g}_0(t)P(X|\epsilon_0 \ge t) - \dot{\hat{g}}_n(t)\bar{X}(t;\hat{\beta}_n)\|_\infty\\ &\quad \le \|\dot{g}_0(t) - \dot{\hat{g}}_n(t)\|_\infty \|P(X|\epsilon_0 \ge t)\|_\infty\\ &\qquad + \|P(X|\epsilon_0 \ge t) - \bar{X}(t;\hat{\beta}_n)\|_\infty \|\dot{\hat{g}}_n(t)\|_\infty\\ &\quad = O_p(n^{-2v}) + O_p(n^{-1/2}) = O_p(n^{-2v})\end{aligned}$$

by Lemma 7.4 and Corollary 6.21 in [2] for the first term and straightforward argument using empirical process theory for Donsker classes for the second term, together with the boundedness of $\|P(X|\epsilon_0 \geq t)\|_\infty$ and $\|\dot{\hat{g}}_n(t)\|_\infty$, and it is straightforward to see that the remaining two terms on the right hand side of inequality (3.1) is $O_p(n^{-1/2})$. Thus we have $I_{1n} = o_p(1)$.

That $I_{2n} = o_p(1)$ can be argued directly by the dominated convergence theorem. We now have proved the theorem.

References.

[1] Ding, Y. (2010). *Some New Insights about the Accelerated Failure Time Model.* Ph.D. Thesis, Biostatistics, University of Michigan.

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

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

[4] Wellner, J. A. and Zhang, Y. (2007). Two Likelihood-based Semiparametric Estimation Methods for Panel Count Data with Covariates. *The Annals of Statistics* **35** 2106–2142.

Department of Biostatistics
University of Michigan
1420 Washington Heights
Ann Arbor, MI 48109-2029
E-mail: yingding@umich.edu
bnan@umich.edu