Supplementary file for – Single-index models for extreme value index regression–

Takuma Yoshida[1]
[1] *Graduate School of Science and Engineering, Kagoshima University*
1-21-35 Korimoto, Kagoshima 8908580, Kagoshima, Japan
E-mail: yoshida@sci.kagoshima-u.ac.jp

Supplementary file shows the proofs of lemmas inluded in Appendix B–D of main article. For convenience, the Lemmas are also mentioned here. In following, for simplicity, we write $P(Y > w_n(\boldsymbol{x}) \mid \boldsymbol{X} = \boldsymbol{x})$ as

$$P(Y > w_n(\boldsymbol{x}) \mid \boldsymbol{X}^T\boldsymbol{\theta} = \boldsymbol{x}^T\boldsymbol{\theta}, \boldsymbol{X} = \boldsymbol{x}).$$

Lemma 1. *Suppose that (C1)–(C6). Then, as* $K \to \infty$,

$$\sup_{x\in[a,b]} |\alpha_0(x) - s_0(x) - b_{s,n}(x)| = o(K^{-q}),$$

where $b_{s,n}(x)$ *is that given in Section 3.1.*

Proof of Lemma 1. From the property of B-spline, there exists the spline model s^* such that $\sup_{x\in[a,b]} |\alpha_0(x) - s^*(x)| \to 0$ as $K \to \infty$ (see, Barrow and Smith 1978). By the definition of s_0, and continuously and convexity of $L(\cdot)$, we have

$$|L(\alpha_0) - L(s_0)| \le |L(\alpha_0) - L(s^*)| \to 0$$

as $K \to \infty$. This implies that $\sup_{x\in\mathcal{X}} |\alpha_0(x) - s_0(x)| \to 0$. Therefore, for some $\eta \in (0, 1)$,

$$\begin{aligned}&|L(s_0) - L(\alpha_0)|\\ &\le |E\left[(\alpha_0(X_0) - s_0(X_0))\{\exp[\eta(\alpha_0(X_0) - s_0(X_0))]\exp[\alpha_0(X_0)]\log(Y/w_n(\boldsymbol{X})) - 1\}I(Y > w_n(\boldsymbol{X}))\right]|.\end{aligned}$$

From above, if we would like to minimize $|L(s_0) - L(\alpha_0)|$, $|s_0(x) - \alpha_0(x)|$ should be small as much as possible. Meanwhile, under given $\boldsymbol{X} = \boldsymbol{x}$ and $Y > w_n(\boldsymbol{x})$, $\exp[\alpha(x_0)]\log(Y/w_n(\boldsymbol{x}))$ is asymptotically distributed as standard exponential distribution. Therefore, in order to minimize $|L(s_0)-L(\alpha_0)|$, s_0 needs to be determined such as $\exp[\eta(s_0(\cdot)-\alpha_0(\cdot))]$ close to one. Consequently, we see that α_0 is the best L_∞ B-spline approximation to α. This implies that s_0 is equal or very close to s^*. Accordingly, from Zhou et al. (1998), we obtain

$$||s_0 - \alpha_0 - b_s||_\infty = o(K^{-q}). \tag{1}$$

This completes the proof. □

Lemma 2. *Suppose that (C1)–(C7). As* $n \to \infty$,

$$\dot{U}_n(\boldsymbol{b}_0) = \frac{1}{\sqrt{n}}\{E[P(Y > w_n(\boldsymbol{X})|\boldsymbol{X})\boldsymbol{B}(X_0)\boldsymbol{B}(X_0)^T]\}^{1/2}\boldsymbol{\varepsilon}+E\left[P(Y > w_n(\boldsymbol{X})|\boldsymbol{X})r_n(\boldsymbol{X})\boldsymbol{B}(X_0)\right]+\lambda\Delta_{m,K}\boldsymbol{b}_0,$$

where $\boldsymbol{\varepsilon}$ *is random variable with* K*-variate standard normal and* $\sup_{\boldsymbol{x}\in\mathcal{X}} |r_n(\boldsymbol{x})| \le O(K^{-q}) + O(\tau_n^{\beta_{min}})$.

Proof of Lemma 2. We note that

$$\dot{U}_n(\boldsymbol{b}_0) = \dot{U}_n(\boldsymbol{b}_0) - E[\dot{U}_n(\boldsymbol{b}_0)] + E[\dot{U}_n(\boldsymbol{b}_0)].$$

Here,

$$\begin{aligned} E[\dot{U}_n(\boldsymbol{b}_0)] &= E\left[\left\{\exp[\boldsymbol{B}(X_0)^T\boldsymbol{b}_0]\log\left(\frac{Y}{w_n(\boldsymbol{X})}\right)-1\right\} I(Y > w_n(\boldsymbol{X}))\boldsymbol{B}(X_0)\right] + \lambda\Delta_{m,K}\boldsymbol{b}_0 \\ &= E\left[P(Y > w_n(\boldsymbol{X})|\boldsymbol{X})r_n(\boldsymbol{X})\boldsymbol{B}(X_0)\right] + \lambda\Delta_{m,K}\boldsymbol{b}_0, \end{aligned}$$

where $x_0 = \boldsymbol{x}^T\boldsymbol{\theta}$ and

$$r_n(\boldsymbol{x}) = E\left[\exp[\boldsymbol{B}(x_0)^T\boldsymbol{b}_0]\log\left(\frac{Y}{w_n(\boldsymbol{x})}\right) - 1 \,\middle|\, \boldsymbol{X} = \boldsymbol{x}, Y > w_n(\boldsymbol{x})\right].$$

The above $E[\dot{U}_n(\boldsymbol{b}_0)]$ is the second term of right hand side of $\dot{U}$ in Lemma 2. From the definition of he Pareto-type tailed model with Hall class, we have

$$\begin{aligned} &E\left[\frac{1}{\gamma_0(x_0)}\log\left(\frac{Y}{w_n(\boldsymbol{x})}\right)\,\middle|\, \boldsymbol{X} = \boldsymbol{x}, Y > w_n(\boldsymbol{x})\right] \\ &= \int_0^\infty P\left(\frac{1}{\gamma_0(x_0)}\log\left(\frac{Y}{w_n(\boldsymbol{x})}\right) > z \,\middle|\, \boldsymbol{X} = \boldsymbol{x}, Y > w_n(\boldsymbol{x})\right) dz \\ &= 1 - \frac{\ell_1(\boldsymbol{x})}{\ell_0(\boldsymbol{x})}\frac{\beta(\boldsymbol{x})}{\beta(\boldsymbol{x})+1} w_n(\boldsymbol{x})^{-\beta(\boldsymbol{x})/\gamma_0(x_0)}(1+o(1)). \end{aligned} \tag{2}$$

Lemma 1 shows that $|\boldsymbol{B}(x_0)^T\boldsymbol{b}_0 - \alpha_0(x_0)| = O(K^{-q})$. Therefore, we find that

$$\sup_{\boldsymbol{x}\in\mathcal{X}} |r_n(\boldsymbol{x})| \leq O(K^{-q}) + O(\tau_n^{\beta_{min}}).$$

We next show from Lyapnov's central limit theorem and Cramér-Wold device that $\dot{U}_n(\boldsymbol{b}_0) - E[\dot{U}_n(\boldsymbol{b}_0)]$ is asymptically distributed as normal. It is sufficient to consider the leading term of $\dot{U}_n(\boldsymbol{b}_0) - E[\dot{U}_n(\boldsymbol{b}_0)]$ since it has mean zero. For this, we use the fact that

$$E_i = \frac{1}{\gamma_0(X_{0i})}\log\left(\frac{Y_i}{w_n(\boldsymbol{X}_i)}\right)$$

is asymptotically distributed as standard exponential distribution under given $Y_i > w_n(\boldsymbol{x}_i)$. For any vector $\boldsymbol{u} \in \mathbb{R}^K$, we consider

$$(\dot{U}_n(\boldsymbol{b}_0) - E[\dot{U}_n(\boldsymbol{b}_0)])^T\boldsymbol{u} = \sum_{i=1}^n (Z_i - E[Z_i]),$$

where

$$Z_i = \frac{1}{n}\boldsymbol{B}(X_{0i})^T\boldsymbol{u}(E_i - 1)I(Y_i > w_n(\boldsymbol{X}_i))(1+o(1)).$$

It is easy to show that $V[Z_i] = n^{-2}\boldsymbol{u}^T E[P(Y > w_n(\boldsymbol{X}))\boldsymbol{B}(X_0)\boldsymbol{B}(X_0)^T]\boldsymbol{u}(1+o(1))$. Condition (C5) and Lemma 6.3 of Zhou et al. (1998) implies that there exist constants $c_1, c_2 > 0$ such that

$$c_1\tau_n \leq \boldsymbol{u}^T E[P(Y > w_n(\boldsymbol{X}))\boldsymbol{B}(X_0)\boldsymbol{B}(X_0)^T]\boldsymbol{u} \leq c_2\tau_n.$$

Therefore, $\sigma_n \equiv \sqrt{\sum_{i=1}^n V[Z_i]} = O(n^{-1/2}\tau_n^{1/2})$. Similarly, we have $E[|Z_i - E[Z_i]|^4] = n^{-4}E[P(Y > w_n(X))(\boldsymbol{B}(X)^T\boldsymbol{u})^4] \leq c_3\tau_n n^{-4}$. Therefore, we have

$$\frac{\sum_{i=1}^n E[|Z_i - E[Z_i]|^4]}{\sqrt{\sigma_n^4}} = O((n\tau_n)^{-1}) = o(1)$$

as $n \to \infty$. Thus, Lyapnov's condition holds and hence,

$$\sqrt{n}(\dot{U}_n(\boldsymbol{b}_0) - E[\dot{U}_n(\boldsymbol{b}_0)])^T\boldsymbol{u} \overset{a.s.}{\sim} N(0, \boldsymbol{u}^T E[P(Y > w_n(\boldsymbol{X}))\boldsymbol{B}(X_0)\boldsymbol{B}(X_0)^T]\boldsymbol{u}).$$

Here, for random variable A_n and B_n, $A_n \overset{a.s}{\sim} B_n$ means that the distributions of A_n and B_n are asymptotically equivalent as $n \to \infty$. Using Cramér-Wold device, we obtain

$$\sqrt{n}E[P(Y > w_n(\boldsymbol{X}))\boldsymbol{B}(X_0)\boldsymbol{B}(X_0)^T]^{-1/2}\{\dot{U}_n(\boldsymbol{b}_0) - E[\dot{U}_n(\boldsymbol{b}_0)]\} \overset{a.s.}{\sim} N_K(\boldsymbol{0}, I_K),$$

where I_K is the K-square identity matrix. This completes the proof. □

Lemma 3. *For any vector $\boldsymbol{v} \in \mathbb{R}^K$,*

$$||\ddot{U}(\boldsymbol{b}_0)^{-1}\boldsymbol{v}||_{max} \leq C||\boldsymbol{v}||_{max}\frac{K}{\tau_n}$$

for some constant $C > 0$. Furthermore, $||(\ddot{U}_n(\boldsymbol{b}_0)^{-1} - \ddot{U}(\boldsymbol{b}_0)^{-1})\boldsymbol{v}||_{max} = o(||\boldsymbol{v}||_{max}K\tau_n^{-1})$

Proof of Lemma 3. Let

$$G_n(\boldsymbol{b}) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{B}(X_{0i})\boldsymbol{B}(X_{0i})^T \exp[\boldsymbol{B}(X_{0i})^T\boldsymbol{b}] \log\left(\frac{Y_i}{w_n(\boldsymbol{X}_i)}\right) I\left(Y_i > w_n(\boldsymbol{X}_i)\right)$$

and $G(\boldsymbol{b}) = E[G_n(\boldsymbol{b})]$. Then, $\ddot{U}(\boldsymbol{b}) = G(\boldsymbol{b}) + \lambda\Delta_{m,K}$. We have

$$\begin{aligned}
&G(\boldsymbol{b})\\
&= E\left[\boldsymbol{B}(X_0)\boldsymbol{B}(X_0)^T \exp[\boldsymbol{B}(X_0)^T\boldsymbol{b}] \log\left(\frac{Y}{w_n(\boldsymbol{X})}\right) I\left(Y > w_n(\boldsymbol{X})\right)\right]\\
&= E\left[P(Y > w_n(\boldsymbol{X})|\boldsymbol{X})\boldsymbol{B}(X_0)\boldsymbol{B}(X_0)^T E\left[\exp[\boldsymbol{B}(X_0)^T\boldsymbol{b}] \log\left(\frac{Y}{w_n(\boldsymbol{X})}\right)\middle| Y > w_n(\boldsymbol{X}), \boldsymbol{X}\right]\right].
\end{aligned}$$

Similar to (2), we have

$$E\left[\exp[\boldsymbol{B}(X_0)^T\boldsymbol{b}_0] \log\left(\frac{Y}{w_n(\boldsymbol{X})}\right)\middle| Y > w_n(\boldsymbol{X}), \boldsymbol{X}\right] = 1 + o(1).$$

Thus, for any vector $\boldsymbol{u} \in \mathbb{R}^K$, we obtain

$$\boldsymbol{u}^T G(\boldsymbol{b}_0)\boldsymbol{u} = \boldsymbol{u}^T E\left[P(Y > w_n(\boldsymbol{X})|\boldsymbol{X})\boldsymbol{B}(X_0)\boldsymbol{B}(X_0)^T\right]\boldsymbol{u}(1 + o(1)).$$

Together with (C5) and Lemma 6.2 of Zhou et al. (1998), $\rho_{min}(G(\boldsymbol{b}_0)) \geq c_1\tau_n K^{-1}$ and $\rho_{max}(G(\boldsymbol{b}_0)) \leq c_2\tau_n K^{-1}$ with $c_1, c_2 > 0$. Here, we have

$$\ddot{U}(\boldsymbol{b}_0) = G(\boldsymbol{b}_0)^{1/2}(I + \lambda G(\boldsymbol{b}_0)^{-1/2}\Delta_{m,K}G(\boldsymbol{b}_0)^{-1/2})G(\boldsymbol{b}_0)^{1/2}$$

From Proposition 4.2 of Xiao (2019), we have

$$\rho_{max}(I + \lambda G(\boldsymbol{b}_0)^{-1/2}\Delta_{m,K}G(\boldsymbol{b}_0)^{-1/2}) \leq 1 + c\frac{\lambda}{\tau_n}K^{2m}$$

for some constant $c > 0$, and

$$\rho_{min}(I + \lambda G(\boldsymbol{b}_0)^{-1/2}\Delta_{m,K}G(\boldsymbol{b}_0)^{-1/2}) \geq 1.$$

Therefore, we obtain $\rho_{min}(\ddot{U}(\boldsymbol{b})) > c_1\tau_n K^{-1}$ and

$$\rho_{max}(\ddot{U}(\boldsymbol{b})) < c_2\frac{\tau_n(1 + c\lambda\tau_n^{-1}K^{2m})}{K}.$$

Under the condition (C6), we have

$$\frac{\rho_{max}(\ddot{U}(\boldsymbol{b}))}{\rho_{min}(\ddot{U}(\boldsymbol{b}))} \leq c_3(1 + c\lambda\tau_n^{-1}K^{2m}) \leq c_4$$

for some positive constants c_1, c_2, c_3 and c_4. This and Theorem 2.2 of Demko (1977) yield that the (i,j)-element of $\ddot{U}(\boldsymbol{b})^{-1}$ has

$$|(\ddot{U}(\boldsymbol{b})^{-1})_{ij}| \leq \rho_{max}(\ddot{U}(\boldsymbol{b})^{-1})Lr^{|i-j|}$$

for some $r \in (0,1)$ and $L > 0$. For any square matrix A, $\rho_{max}(A^{-1}) = \rho_{min}(A)^{-1}$. This implies that

$$|(\ddot{U}(\boldsymbol{b})^{-1})_{ij}| \leq L\frac{K}{\tau_n}r^{|i-j|} \tag{3}$$

for some constant $L > 0$. Since $1 + r + r^2 + \cdots + r^K \leq (1-r)^{-1}$, for any $\boldsymbol{v} \in \mathbb{R}^K$,

$$||\ddot{U}(\boldsymbol{b})^{-1}\boldsymbol{v}||_{max} \leq C\frac{2}{1-r}||\boldsymbol{v}||_{max}\frac{K}{\tau_n}$$

for some constant $C > 0$.

We next evaluate

$$||(\ddot{U}_n(\boldsymbol{b})^{-1} - \ddot{U}(\boldsymbol{b})^{-1})\boldsymbol{v}||_{max} = ||\ddot{U}(\boldsymbol{b})^{-1}(\ddot{U}_n(\boldsymbol{b}) - \ddot{U}(\boldsymbol{b}))\ddot{U}_n(\boldsymbol{b})^{-1}\boldsymbol{v}||_{max}.$$

First, we show

$$||\ddot{U}_n(\boldsymbol{b}_0) - \ddot{U}(\boldsymbol{b}_0)||_{max} = ||G_n(\boldsymbol{b}_0) - G(\boldsymbol{b}_0)||_{max} = o(\tau_n/K). \tag{4}$$

The (j,k)-element of $G_n(\boldsymbol{b}_0) - G(\boldsymbol{b}_0)$ is

$$G_{n,ij} = \sum_{i=1}^n \frac{1}{n}\{Z_iB_j(X_{0i})B_k(X_{0i}) - E[Z_iB_j(X_{0i})B_k(X_{0i})]\} \equiv \sum_{i=1}^n H_i$$

for $|j-k| \leq d$ and $G_{n,ij} = 0$ otherwise, where

$$Z_i = \boldsymbol{B}(X_{0i})\boldsymbol{B}(X_{0i})^T\exp[\boldsymbol{B}(X_{0i})^T\boldsymbol{b}_0]\log\left(\frac{Y_i}{w_n(\boldsymbol{X}_i)}\right)I(Y_i > w_n(\boldsymbol{X}_i)).$$

We note again that under $Y_i > w_n(\boldsymbol{X}_i)$, $E_i = \gamma_0(x_{0i})^{-1}\log(Y/w_n(\boldsymbol{X}_i))$ is approximately distributed as standard exponential distribution. Therefore, if $\max_i E_i < M$ for some $M > 0$, we see that $\max_i |H_i| \leq cn^{-1}M$ for some constant $c > 0$. Thus, for some $\delta > 0$, we have

$$\begin{aligned} P\left(\left|\sum_{i=1}^n H_i\right| > \delta\right) &\leq P\left(\left|\sum_{i=1}^n H_i\right| > \delta \,\middle|\, \max_i |E_i| < M\right) + P(\max_i |E_i| > M) \\ &\leq P\left(\left|\sum_{i=1}^n H_i\right| > \delta \,\middle|\, \max_i |E_i| < M\right) + ne^{-M}. \end{aligned}$$

It is easy to show that $V[H_i] \le n^{-2}\tau_n K^{-1}$. Therefore, Bernstein's inequality (see, van de Geer (2000)) examines that

$$P\left(\left|\sum_{i=1}^n H_i\right| > \delta\right) \le \exp\left[-c_1 \frac{nK\tau_n^{-1}\delta^2}{1+c_2\delta K\tau_n^{-1}M}\right] + ne^{-M}$$

for some constants $c_1, c_2 > 0$. Putting $\delta = \tau_n K^{-1}\varepsilon_n$ and $M = 1/\varepsilon_n$ with $\varepsilon_n \to 0$, $\varepsilon_n \log(n) \to 0$ and $n\tau_n K^{-1}\varepsilon_n \to \infty$, we obtain

$$\begin{aligned} P\left(\left|\sum_{i=1}^n H_i\right| > \tau_n K^{-1}\varepsilon_n\right) &\le \exp\left[-c\frac{n\tau_n\varepsilon_n}{K}\right] + ne^{-\varepsilon_n^{-1}} \\ &\to 0. \end{aligned}$$

This means that (4) holds.

Similar to the proof of (3), (j,k)-element of $\ddot{U}_n(\boldsymbol{b})^{-1}$ satisfies

$$|(\ddot{U}_n(\boldsymbol{b})^{-1})_{jk}| \le \frac{K}{\tau_n} L q^{|j-k|}$$

for some $q \in (0,1)$ and $L > 0$. Thus, the property of inverse of band matrix examines $||\ddot{U}(\boldsymbol{b})^{-1}\boldsymbol{v}||_{\max} = ||\boldsymbol{v}||_{max} O(K/\tau_n)$. Next, using (4) and the fact that $\ddot{U}_n(\boldsymbol{b}_0) - \ddot{U}(\boldsymbol{b}_0)$ is band matrix, $||\{\ddot{U}_n(\boldsymbol{b}_0) - \ddot{U}(\boldsymbol{b}_0)\}\ddot{U}(\boldsymbol{b})^{-1}\boldsymbol{v}||_{max} = ||\boldsymbol{v}||_{max} o(1)$. Finally, the inverse of band matrix property of $\ddot{U}_n(\boldsymbol{b})^{-1}$ achieves

$$||\ddot{U}_n(\boldsymbol{b})(\ddot{U}_n(\boldsymbol{b}_0) - \ddot{U}(\boldsymbol{b}_0))\ddot{U}(\boldsymbol{b})\boldsymbol{v}||_{max} = ||\boldsymbol{v}||_{max} o(K/\tau_n),$$

which completes the proof. □

Lemma 4. *Suppose that (C1)–(C7). For any $\delta_n > 0$ satisfying $\delta_n \to 0$ and*

$$\frac{1}{\delta_n}\max\left\{\sqrt{\frac{K}{n\tau_n}}, \sqrt{\frac{\lambda K}{\tau_n}}, K^{-q}, \tau_n^{\beta_{min}}\right\} \to 0$$

as $n \to \infty$,

$$P(||\hat{\boldsymbol{b}} - \boldsymbol{b}_0|| > \delta_n) \to 0.$$

Proof of Lemma 4. Let

$$L_0(\boldsymbol{b}) = E\left[\left\{\exp[\boldsymbol{B}(X_0)^T\boldsymbol{b}]\log\left(\frac{Y}{w_n(\boldsymbol{X})}\right) - \boldsymbol{B}(X_0)^T\boldsymbol{b}\right\} I(Y > w_n(\boldsymbol{X}))\right]$$

and

$$\begin{aligned} L(\boldsymbol{b}) &= U_n(\boldsymbol{b}, \boldsymbol{\theta}_0|\lambda) \\ &= \frac{1}{n}\sum_{i=1}^n \left\{\exp[\boldsymbol{B}(X_{0i})^T\boldsymbol{b}]\log\left(\frac{Y_i}{w_n(\boldsymbol{X}_i)}\right) - \boldsymbol{B}(X_{0i})^T\boldsymbol{b}\right\} I(Y_i > w_n(\boldsymbol{X}_i)) \\ &\quad + \frac{\lambda}{2}\int_a^b \left\{\frac{d^m}{dx^m}\boldsymbol{B}(x)^T\boldsymbol{b}\right\}^2 dx. \end{aligned}$$

We note that L and L_0 are strictly convex functions. Therefore, $\boldsymbol{b}_0 = \mathrm{argmin}_{\boldsymbol{b}} L_0(\boldsymbol{b})$ and $\hat{\boldsymbol{b}} = \mathrm{argmin}_{\boldsymbol{b}} L(\boldsymbol{b})$. From Lemma 2 of Hjort and Pollard (2011), for $\delta_n > 0$,

$$P(||\hat{\boldsymbol{b}} - \boldsymbol{b}_0|| > \delta_n) \leq P\left(\sup_{||\boldsymbol{b}-\boldsymbol{b}_0||\leq\delta_n} |L(\boldsymbol{b}) - L_0(\boldsymbol{b})| \geq 2^{-1} \inf_{||\boldsymbol{b}-\boldsymbol{b}_0||=\delta_n} |L_0(\boldsymbol{b}) - L_0(\boldsymbol{b}_0)|\right).$$

We now consider the vector $\boldsymbol{b} \in \mathbb{R}^K$ satisfying $||\boldsymbol{b} - \boldsymbol{b}_0|| = \delta_n$. When $||\boldsymbol{b} - \boldsymbol{b}_0|| = \delta_n$, we can write $\boldsymbol{b} = \boldsymbol{b}_0 + \delta_n \boldsymbol{u}, \boldsymbol{u} \in \mathbb{R}^K$, where $||\boldsymbol{u}|| = 1$. By the straightforward calculation, for some $\theta \in (0, 1)$,

$$\begin{aligned}
&L_0(\boldsymbol{b}) - L_0(\boldsymbol{b}_0)\\
&= \delta_n E\left[\boldsymbol{B}(X_0)^T\boldsymbol{u}\left\{\exp[\boldsymbol{B}(X_0)^T\boldsymbol{b}_0]\log\left(\frac{Y}{w_n(\boldsymbol{X})}\right) - 1\right\} I(Y > w_n(\boldsymbol{X}))\right]\\
&\quad + \frac{\delta_n^2}{2}\boldsymbol{u}^T E\left[\boldsymbol{B}(X_0)\boldsymbol{B}(X_0)^T\exp[\theta\delta_n\boldsymbol{B}(X_0)^T\boldsymbol{u}]\exp[\boldsymbol{B}(X_0)^T\boldsymbol{b}_0]\log\left(\frac{Y}{w_n(\boldsymbol{X})}\right) I(Y > w_n(\boldsymbol{X}))\right]\boldsymbol{u}
\end{aligned}$$

We note that $\boldsymbol{B}(x)^T\boldsymbol{u} \neq 0$ for any $\boldsymbol{u}$ with $||\boldsymbol{u}|| = 1$. Furthermore, for any positive function $v(x) > 0$, the minimum eigen-value of $E[v(X_0)\boldsymbol{B}(X_0)\boldsymbol{B}(X_0)^T]$ has lower bound cK^{-1} for some constant $c > 0$(see, Lemma 6.1 of Zhou et al. 1998). Together with (2) and $|\boldsymbol{B}(X_0)^T\boldsymbol{b}_0 - \alpha_0(X_0)| = O(K^{-q})$, we have

$$|L_0(\boldsymbol{b}) - L_0(\boldsymbol{b}_0)| \geq c^*\frac{\delta_n\tau_n}{K}\left(\frac{1}{K^q} + \delta_n\right) \geq 2c^*\frac{\delta_n^2\tau_n}{K}$$

for some constant $c^* \in (0, 1)$. This c^* is used later. Accordingly, we obtain

$$\begin{aligned}
P(||\hat{\boldsymbol{b}} - \boldsymbol{b}_0|| > \delta_n) &\leq P\left(\sup_{||\boldsymbol{b}-\boldsymbol{b}_0||\leq\delta_n} |L(\boldsymbol{b}) - L_0(\boldsymbol{b})| \geq 2^{-1} \inf_{||\boldsymbol{b}-\boldsymbol{b}_0||=\delta_n} |L_0(\boldsymbol{b}) - L_0(\boldsymbol{b}_0)|\right)\\
&\leq P\left(\sup_{||\boldsymbol{b}-\boldsymbol{b}_0||\leq\delta_n} |L(\boldsymbol{b}) - L_0(\boldsymbol{b})| \geq c^*\frac{\delta_n^2\tau_n}{K}\right).
\end{aligned}$$

For simplicity, we put $\zeta_n = \delta_n^2\tau_n K^{-1}$. Then, our porpose is to show

$$P\left(\sup_{||\boldsymbol{b}-\boldsymbol{b}_0||\leq\delta_n} |L(\boldsymbol{b}) - L_0(\boldsymbol{b})| \geq c^*\zeta_n\right) \to 0. \tag{5}$$

Again, we consider $\boldsymbol{b} = \boldsymbol{b}_0 + \delta_n\boldsymbol{u}, \boldsymbol{u} \in \mathbb{R}^K$, where $||\boldsymbol{u}|| \leq 1$. We then obtain

$$\begin{aligned}
&P\left(\sup_{||\boldsymbol{b}-\boldsymbol{b}_0||\leq\delta_n} |L(\boldsymbol{b}) - L_0(\boldsymbol{b})| \geq c^*\zeta_n\right)\\
&\leq P\left(|L(\boldsymbol{b}_0) - L_0(\boldsymbol{b}_0)| \geq 2^{-1}c^*\zeta_n\right)\\
&+ P\left(\sup_{||\boldsymbol{b}-\boldsymbol{b}_0||\leq\delta_n} |L(\boldsymbol{b}) - L(\boldsymbol{b}_0) - L_0(\boldsymbol{b}) + L_0(\boldsymbol{b}_0)| \geq 2^{-1}c^*\zeta_n\right)\\
&\equiv J_1 + J_2.
\end{aligned}$$

We evaluate J_1. Define

$$\ell(Y, \boldsymbol{X}|\boldsymbol{b}) = \left\{\exp[\boldsymbol{B}(X_0)^T\boldsymbol{b}]\log\left(\frac{Y}{w_n(\boldsymbol{X})}\right) - \boldsymbol{B}(X_0)^T\boldsymbol{b}\right\} I(Y > w_n(\boldsymbol{X})).$$

Since $\sup_{x\in[a,b]}|\alpha_0^{(m)}(x)-(d^m/dx^m)\boldsymbol{B}(x)^T\boldsymbol{b}_0|=O(K^{-(q-m)})$ (see, Barrow and Smith 1978), we obtain

$$L(\boldsymbol{b}_0)-L_0(\boldsymbol{b}_0)=\frac{1}{n}\sum_{i=1}^n \ell(Y_i,X_i|\boldsymbol{b}_0)-E[\ell(Y_i,X_i|\boldsymbol{b}_0)]+\frac{\lambda}{2}\int_a^b\{\alpha_0^{(m)}(x)\}^2dx(1+o(1)).$$

Under the condition of Lemma 4, we have

$$\frac{\lambda}{\zeta_n}\to 0.$$

Therefore, it is sufficient to show

$$P\left(\left|n^{-1}\sum_{i=1}^n \ell(Y_i,X_i|\boldsymbol{b}_0)-E[\ell(Y_i,X_i|\boldsymbol{b}_0)]\right|>2^{-1}c^*\zeta_n\right)\to 0.$$

We let $E_i=\gamma(X_{0i})^{-1}\log(Y_i/w_n(\boldsymbol{X}_i))$ under given $Y_i>w_n(\boldsymbol{X}_i)$. As we noted in the proof of Lemma 3, E_i is asymptotically distributed as standard exponential. We can then easily find that $V[\ell(Y_i,\boldsymbol{X}_i|\boldsymbol{b}_0)]\le c\tau_n$ for some constant $c>0$. Furthermore, if $\max_i|E_i|<M$ for a constant $M>0$, we have $\max_i|\ell(Y_i,X_i|\boldsymbol{b}_0)-E[\ell(Y,X|\boldsymbol{b}_0)]|<c_0M$ with additional constant $c_0>0$. Therefore, using Bernstein's inequality with $M=\tau_n\zeta_n^{-1}$, we obtain that for some constant $C^*>0$,

$$\begin{aligned}
&P\left(\left|n^{-1}\sum_{i=1}^n \ell(Y_i,X_i|\boldsymbol{b}_0)-E[\ell(Y,X|\boldsymbol{b}_0)]\right|>2^{-1}c^*\zeta_n\right)\\
&\le P\left(\left|n^{-1}\sum_{i=1}^n \ell(Y_i,X_i|\boldsymbol{b}_0)-E[\ell(Y_i,X_i|\boldsymbol{b}_0)]\right|>2^{-1}c^*\zeta_n\right|, \max_i|E_i|<M\right)+\sum_{i=1}^n P(E_i>M)\\
&\le c\exp\left[-C^*\frac{n\zeta_n^2}{\tau_n}\right]+cn\exp\left[-\tau_n\zeta_n^{-1}\right]\\
&\le c\exp\left[-C^*\frac{n\tau_n\delta_n^2}{K}\right]+cn\exp\left[-K\delta_n^{-2}\right]\\
&\to 0.
\end{aligned}$$

Next, we focus on J_2. Define

$$\ell_1(\boldsymbol{X},Y|\boldsymbol{u})=\boldsymbol{B}(X_0)^T\boldsymbol{u}\left\{\exp[\theta\delta_n\boldsymbol{B}(X_0)^T\boldsymbol{u}]\exp[\boldsymbol{B}(X_0)^T\boldsymbol{b}_0]\log\left(\frac{Y}{w_n(\boldsymbol{X})}\right)-1\right\}I(Y>w_n(\boldsymbol{X})).$$

Then, for some $\theta\in(0,1)$, we obtain

$$\begin{aligned}
&L(\boldsymbol{b})-L(\boldsymbol{b}_0)\\
&=\delta_n\frac{1}{n}\sum_{i=1}^n\boldsymbol{B}(X_{0i})^T\boldsymbol{u}\left\{\exp[\theta\delta_n\boldsymbol{B}(X_{0i})^T\boldsymbol{u}]\exp[\boldsymbol{B}(X_{0i})^T\boldsymbol{b}_0]\log\left(\frac{Y_i}{w_n(\boldsymbol{X}_i)}\right)-1\right\}I(Y_i>w_n(\boldsymbol{X}_i))\\
&\quad+\lambda\left\{\delta_n\boldsymbol{u}^T\Delta_{m,K}\boldsymbol{b}_0+\delta_n^2\boldsymbol{u}^T\Delta_{m,K}\boldsymbol{u}\right\}\\
&=\delta_n\frac{1}{n}\sum_{i=1}^n\ell_1(\boldsymbol{X}_i,Y_i|\boldsymbol{u})+\frac{\lambda}{2}\left[\int_a^b\left\{\frac{d^m}{dx^m}\boldsymbol{B}^{[d]}(x)^T\{\boldsymbol{b}_0+\delta_n\boldsymbol{u}\}\right\}^2dx-\int_a^b\left\{\frac{d^m}{dx^m}\boldsymbol{B}^{[d]}(x)^T\boldsymbol{b}_0\right\}^2dx\right].
\end{aligned}$$

Under $\delta_n \to 0$, we obtain

$$\lambda \left[\int_a^b \left\{ \frac{d^m}{dx^m} \boldsymbol{B}^{[d]}(x)^T \{\boldsymbol{b}_0 + \delta_n \boldsymbol{u}\} \right\}^2 dx - \int_a^b \left\{ \frac{d^m}{dx^m} \boldsymbol{B}^{[d]}(x)^T \boldsymbol{b}_0 \right\}^2 dx \right] = O(\lambda \delta_n).$$

This yields that

$$\begin{aligned}
&\sup_{||\boldsymbol{u}||<1} |L(\boldsymbol{b}) - L(\boldsymbol{b}_0) - L_0(\boldsymbol{b}) + L_0(\boldsymbol{b}_0)| \\
&\le \delta_n \sup_{||\boldsymbol{u}||<1} \left| \frac{1}{n} \sum_{i=1}^n \{\ell_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}) - E[\ell_1(\boldsymbol{X}, Y|\boldsymbol{u})]\} \right| + O(\lambda \delta_n).
\end{aligned}$$

Under (C7), we have $\lambda \delta_n / \zeta_n \to 0$. Thus, the remaining proof is to show

$$P\left(\sup_{||\boldsymbol{u}||<1} \left| \frac{1}{n} \sum_{i=1}^n \{\ell_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}) - E[\ell_1(\boldsymbol{X}, Y|\boldsymbol{u})]\} \right| \ge 2^{-1} c^* \zeta_n \delta_n^{-1} \right) \to 0.$$

Let $\mathcal{U} = \{\boldsymbol{u} \in \mathbb{R}^K : ||\boldsymbol{u}|| < 1\}$ be the vector space and $\mathcal{U}_1, \ldots, \mathcal{U}_N$ be a covering of $\mathcal{U}$ with the diameter $R_n = C\zeta_n/(4\delta_n n^\nu)$ for some constant $C > 0$ and $\nu > 0$. That is, $\mathcal{U} \subseteq \cup_{i=1}^N \mathcal{U}_i$. Then, Lemma 2.5 of van de Geer (2000) yields that it is sufficient to set $N \le C(n^\nu \delta_n/\zeta_n)^K$. Let $\boldsymbol{u}_k \in \mathcal{U}_k, k = 1, \ldots, N$ and let M_1 be positive constant. Then, for some constant $c_1 > 0$, we have

$$\begin{aligned}
&P\left(\sup_{||\boldsymbol{u}||<1} \left| \frac{1}{n} \sum_{i=1}^n \{\ell_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}) - E[\ell_1(\boldsymbol{X}, Y|\boldsymbol{u})]\} \right| \ge 2^{-1} c^* \zeta_n \delta_n^{-1} \right) \\
&\le \sum_{k=1}^N P\left(\left| \frac{1}{n} \sum_{i=1}^n \{\ell_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}_k) - E[\ell_1(\boldsymbol{X}, Y|\boldsymbol{u}_k)]\} \right| + \sup_{||\boldsymbol{u}||\in\mathcal{U}_k} \left| \frac{1}{n} \sum_{i=1}^n \ell_2(\boldsymbol{X}_i, Y_i|\boldsymbol{u}, \boldsymbol{u}_k) \right| \ge 2^{-1} c^* \zeta_n \delta_n^{-1} \right) \\
&\le \sum_{k=1}^N P\left(\left| \frac{1}{n} \sum_{i=1}^n \{\ell_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}_k) - E[\ell_1(\boldsymbol{X}, Y|\boldsymbol{u}_k)]\} \right| \right. \\
&\left. + \sup_{||\boldsymbol{u}||\in\mathcal{U}_k} \left| \frac{1}{n} \sum_{i=1}^n \ell_2(\boldsymbol{X}_i, Y_i|\boldsymbol{u}, \boldsymbol{u}_k) \right| \ge 2^{-1} c^* \zeta_n \delta_n^{-1} \,\middle|\, \max_i E_i \le M_1 \right) + c_1 N n \exp[-M_1]
\end{aligned}$$

where

$$\ell_2(\boldsymbol{X}_i, Y_i|\boldsymbol{u}, \boldsymbol{u}_k) = \ell_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}) - E[\ell_1(\boldsymbol{X}, Y|\boldsymbol{u})] - \{\ell_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}_k) - E[\ell_1(\boldsymbol{X}, Y|\boldsymbol{u}_k)]\}, \quad \boldsymbol{u} \in \mathcal{U}_k.$$

We evaluate

$$\sup_{||\boldsymbol{u}||\in\mathcal{U}_k} \left| \frac{1}{n} \sum_{i=1}^n \ell_2(\boldsymbol{X}_i, Y_i|\boldsymbol{u}, \boldsymbol{u}_k) \right|.$$

By the Taylor expansion, there exists $\eta \in (0, 1)$ such that

$$\begin{aligned}
&\ell_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}) - \ell_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}_k) \\
&= \boldsymbol{B}(X_{0i})^T(\boldsymbol{u} - \boldsymbol{u}_k) \left\{ \exp[\theta\delta_n \boldsymbol{B}(X_{0i})^T \boldsymbol{u}_k] \exp[\boldsymbol{B}(X_{0i})^T \boldsymbol{b}_0] \log\left(\frac{Y_i}{w_n(\boldsymbol{X}_i)} \right) - 1 \right\} I(Y_i > w_n(\boldsymbol{X}_i)) \\
&+ \theta\delta_n (\boldsymbol{u} - \boldsymbol{u}_k)^T \boldsymbol{B}(X_{0i}) \boldsymbol{B}(X_{0i})^T \boldsymbol{u}_k \\
&\times \exp[\theta\delta_n \boldsymbol{B}(X_{0i})^T \boldsymbol{u}_k] \exp[\eta\theta\delta_n \boldsymbol{B}(X_{0i})^T (\boldsymbol{u} - \boldsymbol{u}_k)] \exp[\boldsymbol{B}(X_{0i})^T \boldsymbol{b}_0] \log\left(\frac{Y_i}{w_n(\boldsymbol{X}_i)} \right) I(Y_i > w_n(\boldsymbol{X}_i)).
\end{aligned}$$

We see that under $\delta_n \to 0$,

$$\exp[\theta\delta_n \boldsymbol{B}(X_{0i})^T\boldsymbol{u}]\exp[\eta\theta\delta_n \boldsymbol{B}(X_{0i})^T(\boldsymbol{u}-\boldsymbol{u}_k)]\exp[\boldsymbol{B}(X_{0i})^T\boldsymbol{b}_0]\log\left(\frac{Y_i}{w_n(\boldsymbol{X}_i)}\right)$$

is approximately distributed as E_i. Therefore, under $\max_i E_i \le M_1$, for some constant $C > 0$, we obtain

$$\left|\frac{1}{n}\sum_{i=1}^n \ell_2(\boldsymbol{X}_i, Y_i|\boldsymbol{u},\boldsymbol{u}_k)\right| \le 2CM_1\frac{\delta_n\tau_n}{K}||\boldsymbol{u}-\boldsymbol{u}_k|| \le C\frac{\delta_n\tau_n R_n}{K}M_1 \le C\frac{\zeta_n}{\delta_n}\frac{M_1\delta_n\tau_n}{Kn^\nu}.$$

Putting $M_1 = c_2\log(Nn\log(n))$ for some constant $c_2 > 0$, we obtain $M_1\delta_n\tau_n/(Kn^\nu) \to 0$. This yields that under $\max_i E_i \le M_1$,

$$\sup_{||\boldsymbol{u}||\in\mathcal{U}_k}\left|\frac{1}{n}\sum_{i=1}^n \ell_2(\boldsymbol{X}_i, Y_i|\boldsymbol{u},\boldsymbol{u}_k)\right| < 2^{-1}c\frac{\zeta_n}{\delta_n}.$$

Furthermore, we have

$$Nn\exp[-M_1] \to 0.$$

Lastly, we evaluate

$$\sum_{k=1}^N P\left(\left|\frac{1}{n}\sum_{i=1}^n\{\ell_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}_k) - E[\ell_1(\boldsymbol{X}, Y|\boldsymbol{u}_k)]\}\right| \ge 2^{-1}c^*\zeta_n\delta_n^{-1}\middle|\max_i E_i \le M_1\right) \to 0.$$

Under $\max_i E_i \le M_1$, it is easy to find that $\max_i|\ell_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}_k)| \le c_0\delta_n M_1$ for some constant $c > 0$. Furthermore, we obtain $V[\ell_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}_k)] \le c_0\delta_n^2\tau_n K^{-1}$. This and Bernstein's inequality yield that

$$\begin{aligned}
&P\left(\left|\frac{1}{n}\sum_{i=1}^n\{\ell_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}_k) - E[\ell_1(\boldsymbol{X}, Y|\boldsymbol{u}_k)]\}\right| \le c\frac{\zeta_n}{\delta_n}\middle|\max_i |E_i| < M_1\right)\\
&\le C\exp\left[-C\frac{\zeta_n^2/\delta_n^2}{\delta_n^2\tau_n/(nK) + M_1\delta_n\zeta_n/(n\delta_n)}\right]\\
&\le C\exp\left[-C\frac{n}{\log(n)}\right]
\end{aligned}$$

for some constant $C > 0$. Accordingly, we obtain

$$\sum_{k=1}^N P\left(\left|\frac{1}{n}\sum_{i=1}^n\{\ell_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}_k) - E[\ell_1(\boldsymbol{X}, Y|\boldsymbol{u}_k)]\}\right| \le c\frac{\zeta_n}{\delta_n}\middle|\max_i |E_i| < M_1\right) \to 0.$$

Consequently, $P(||\hat{\boldsymbol{b}} - \boldsymbol{b}_0|| > \delta_n) \to 0$ holds. □

Lemma 5. *Suppose that (C1)–(C9). Let δ_n be the positive sequence satisfying $\delta_n \to 0$, $(n\tau_n)^{1/2}\delta_n \to \infty$, $\delta_n\log(n) \to 0$ and $\eta_n/\delta_n^2 \to 0$. Then, as $n \to \infty$,*

$$P(||\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0|| > \delta_n) \to 0.$$

Proof of Lemma 5. Let

$$L_n(\boldsymbol{\phi}) = \frac{1}{n}\sum_{i=1}^{n}\left\{\exp[\alpha_0(\boldsymbol{X}_i^T\boldsymbol{\theta}(\boldsymbol{\phi})|\boldsymbol{\theta}(\boldsymbol{\phi}))]\log(Y_i/w_n(\boldsymbol{X}_i)) - \alpha_0(\boldsymbol{X}_i^T\boldsymbol{\theta}(\boldsymbol{\phi})|\boldsymbol{\theta}(\boldsymbol{\phi}))\right\} I(Y_i > w_n(\boldsymbol{X}_i))$$

and let

$$\hat{L}_n(\boldsymbol{\phi}) = \frac{1}{n}\sum_{i=1}^{n}\left\{\exp[\hat{\alpha}(\boldsymbol{X}_i^T\boldsymbol{\theta}(\boldsymbol{\phi})|\boldsymbol{\theta}(\boldsymbol{\phi}))]\log(Y_i/w_n(\boldsymbol{X}_i)) - \hat{\alpha}(\boldsymbol{X}_i^T\boldsymbol{\theta}(\boldsymbol{\phi})|\boldsymbol{\theta}(\boldsymbol{\phi}))\right\} I(Y_i > w_n(\boldsymbol{X}_i)).$$

Then, $\hat{\boldsymbol{\phi}} = \operatorname{argmin} \hat{L}_n(\boldsymbol{\phi}|\hat{\alpha}(\cdot|\boldsymbol{\theta}(\boldsymbol{\phi})))$. From the theorems in Section 3.1, we see that for any $\boldsymbol{\phi} \in \mathcal{S}_c$, $|\hat{\alpha}(\boldsymbol{x}^T\boldsymbol{\theta}(\boldsymbol{\phi})|\boldsymbol{\theta}(\boldsymbol{\phi})) - \alpha_0(\boldsymbol{x}^T\boldsymbol{\theta}(\boldsymbol{\phi})|\boldsymbol{\theta}(\boldsymbol{\phi}))| = O(\eta_n)$. Therefore, from the definition of L_n, we have $|\hat{L}_n(\boldsymbol{\phi}) - L_n(\boldsymbol{\phi})| \leq C_L\eta_n\tau_n$ for some constant $C_L > 0$. Next, we define

$$L(\boldsymbol{\phi}) = E\left[\left\{\exp[\alpha_0(\boldsymbol{X}^T\boldsymbol{\theta}(\boldsymbol{\phi})|\boldsymbol{\theta}(\boldsymbol{\phi}))]\log(Y/w_n(\boldsymbol{X})) - \alpha_0(\boldsymbol{X}^T\boldsymbol{\theta}(\boldsymbol{\phi})|\boldsymbol{\theta}(\boldsymbol{\phi}))\right\} I(Y > w_n(\boldsymbol{X}))\right].$$

The true single index parameter vector $\boldsymbol{\phi}_0$ is the minimizer of $L_0(\boldsymbol{\phi}) = -E[\log f(Y|\alpha_0(\cdot|\boldsymbol{\theta}(\boldsymbol{\phi})), \boldsymbol{\theta}(\boldsymbol{\phi}))]$. By the definition of L and $\boldsymbol{\theta}_0 = \boldsymbol{\theta}(\boldsymbol{\phi}_0)$, we have

$$|-E[\log f_{w_n}(Y|\alpha_0(\cdot|\boldsymbol{\theta}_0), \boldsymbol{\theta}_0)] - L(\boldsymbol{\phi}_0)| \leq C_\ell \tau_n^{\beta_{min}+1}$$

for some constant $C_\ell > 0$. We note that $\alpha_0(\cdot) = \alpha_0(\cdot|\boldsymbol{\theta}_0)$. Thanks to Hijort and Pollard (2011), for any sufficiently small $\delta_n > 0$ and some condtant $C > 0$,

$$\begin{aligned}
&P(||\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0|| > \delta_n)\\
&\leq P\left(\sup_{||\boldsymbol{\phi}-\boldsymbol{\phi}_0||<\delta_n} |\hat{L}_n(\boldsymbol{\phi}) - L_0(\boldsymbol{\phi})| > 2^{-1}\inf_{||\boldsymbol{\phi}-\boldsymbol{\phi}_0||=\delta_n}|L_0(\boldsymbol{\phi}) - L_0(\boldsymbol{\phi}_0)|\right)\\
&\leq P\left(\sup_{||\boldsymbol{\phi}-\boldsymbol{\phi}_0||<\delta_n} |L_n(\boldsymbol{\phi}) - L(\boldsymbol{\phi})| + C\{\tau_n\eta_n + \tau_n^{\beta_{min}}\} > 2^{-1}\inf_{||\boldsymbol{\phi}-\boldsymbol{\phi}_0||=\delta_n}|L(\boldsymbol{\phi}) - L(\boldsymbol{\phi}_0)|\right)
\end{aligned}$$

Here, we write

$$A_1(\boldsymbol{x}|\boldsymbol{\phi}) = \frac{d\alpha_0(\boldsymbol{x}^T\boldsymbol{\theta}(\boldsymbol{\phi})|\boldsymbol{\theta}(\boldsymbol{\phi}))}{d\boldsymbol{\phi}}$$

and

$$A_2(\boldsymbol{x}|\boldsymbol{\phi}) = \frac{d^2\alpha_0(\boldsymbol{x}^T\boldsymbol{\theta}(\boldsymbol{\phi})|\boldsymbol{\theta}(\boldsymbol{\phi}))}{d\boldsymbol{\phi}d\boldsymbol{\phi}^T}.$$

If $||\boldsymbol{\phi} - \boldsymbol{\phi}_0|| = \delta_n$, there exists $\boldsymbol{u} \in \mathbb{R}^{p-1}$ such that $||\boldsymbol{u}|| = 1$ and $\boldsymbol{\phi} = \boldsymbol{\phi}_0 + \delta_n\boldsymbol{u}$. By the Taylor expansion, there exists $\boldsymbol{\phi}^*$ satisfying $||\boldsymbol{\phi}_0 - \boldsymbol{\phi}^*|| < ||\boldsymbol{\phi}_0 - \boldsymbol{\phi}|| = \delta_n$ such that $\boldsymbol{\theta}^* = \boldsymbol{\theta}(\boldsymbol{\phi}^*)$ and

$$\begin{aligned}
&L(\boldsymbol{\phi}) - L(\boldsymbol{\phi}_0)\\
&= \delta_n\boldsymbol{u}^T E\left[A_1(\boldsymbol{X}|\boldsymbol{\phi}_0)\left\{\exp[\alpha_0(\boldsymbol{X}^T\boldsymbol{\theta}_0)]\log(Y/w_n(\boldsymbol{X})) - 1\right\} I(Y > w_n(\boldsymbol{X}))\right]\\
&+\delta_n^2\boldsymbol{u}^T E\left[A_2(\boldsymbol{X}|\boldsymbol{\phi}^*)\left\{\exp[\alpha_0(\boldsymbol{X}^T\boldsymbol{\theta}^*|\boldsymbol{\theta}^*)]\log(Y/w_n(\boldsymbol{X})) - 1\right\} I(Y > w_n(\boldsymbol{X}))\right]\boldsymbol{u}\\
&+\delta_n^2\boldsymbol{u}^T E\left[A_1(\boldsymbol{X}, \boldsymbol{\phi}_0)A_1(\boldsymbol{X}, \boldsymbol{\phi}^*)^T \exp[\alpha_0(\boldsymbol{X}^T\boldsymbol{\theta}^*|\boldsymbol{\theta}^*)]\log(Y/w_n(\boldsymbol{X}))I(Y > w_n(\boldsymbol{X}))\right]\boldsymbol{u}\{1 + o_P(1)\}.
\end{aligned}$$

Since

$$\begin{aligned}
&|E\left[\exp[\alpha_0(\boldsymbol{X}^T\boldsymbol{\theta}_0)]\log(Y/w_n(\boldsymbol{X})) - 1|Y > w_n(\boldsymbol{x}), \boldsymbol{X}^T\boldsymbol{\theta}_0 = \boldsymbol{x}^T\boldsymbol{\theta}_0, \boldsymbol{X} = \boldsymbol{x}\right]|\\
&= \frac{|\ell_1(\boldsymbol{x})|}{\ell_0(\boldsymbol{x})}\frac{\beta(\boldsymbol{x})}{\beta(\boldsymbol{x}) + 1}w_n(\boldsymbol{x})^{-\beta(\boldsymbol{x})/\gamma_0(\boldsymbol{x}^T\boldsymbol{\theta}_0)}(1 + o(1))\\
&= O(\tau_n^{\beta(\boldsymbol{x})})\\
&\geq C\tau_n^{\beta_{max}}.
\end{aligned}$$

for some constant $C > 0$, where $\beta_{max} = \sup_{\boldsymbol{x}\in\mathcal{X}} \beta(\boldsymbol{x})$. Here, we are allowed to use $\beta_{max} = \infty$. Thus, under (C9), we have

$$\inf_{||\boldsymbol{\phi}-\boldsymbol{\phi}_0||=\delta_n} |L(\boldsymbol{\phi}) - L(\boldsymbol{\phi}_0|)| \geq C\delta\tau_n \left\{\tau_n^{\beta_{max}} + \delta\tau_n^{\beta_{max}} + \delta\right\} \geq C\delta^2\tau_n.$$

By the assumption, we see that $\min\{\tau_n\eta_n, \tau_n^{\beta_{min}+1}\}/\tau_n\delta_n^2 \to 0$. Thus, our purpose is to show

$$P\left(\sup_{||\boldsymbol{\phi}-\boldsymbol{\phi}_0||\leq\delta_n} |L_n(\boldsymbol{\phi}|\alpha_0(\cdot|\boldsymbol{\theta}(\boldsymbol{\phi}))) - L(\boldsymbol{\phi}|\alpha_0)| > C\delta_n^2\tau_n\right) \to 0.$$

We define $\boldsymbol{\phi} = \boldsymbol{\phi}_0 + \delta_n\boldsymbol{u}$, where $||\boldsymbol{u}|| \leq 1$. Then, we have

$$\begin{aligned}
&P\left(\sup_{||\boldsymbol{\phi}-\boldsymbol{\phi}_0||\leq\delta_n} |L_n(\boldsymbol{\phi}) - L(\boldsymbol{\phi})| > C\delta_n^2\tau_n\right)\\
&\leq P\left(|L_n(\boldsymbol{\phi}_0) - L(\boldsymbol{\phi}_0)| > 2^{-1}C\delta_n^2\tau_n\right)\\
&+P\left(\sup_{||\boldsymbol{\phi}-\boldsymbol{\phi}_0||\leq\delta_n} |L_n(\boldsymbol{\phi}) - L_n(\boldsymbol{\phi}_0) - L(\boldsymbol{\phi}) + L(\boldsymbol{\phi}_0)| > 2^{-1}C\delta_n^2\tau_n\right)\\
&\equiv J_1 + J_2.
\end{aligned}$$

Let

$$\begin{aligned}
h(\boldsymbol{X}_i, Y_i|\boldsymbol{u}) \quad = \quad & \left\{\exp[\alpha_0(\boldsymbol{X}_i^T\boldsymbol{\theta}(\boldsymbol{\phi}_0 + \delta_n\boldsymbol{u})|\boldsymbol{\theta}(\boldsymbol{\phi}_0 + \delta_n\boldsymbol{u}))]\log(Y_i/w_n(\boldsymbol{X}_i))\right.\\
&\left.-\alpha_0(\boldsymbol{X}_i^T\boldsymbol{\theta}(\boldsymbol{\phi}_0 + \delta_n\boldsymbol{u})|\boldsymbol{\theta}(\boldsymbol{\phi}_0 + \delta_n\boldsymbol{u}))\right\} I(Y_i > w_n(\boldsymbol{X}_i)).
\end{aligned}$$

Then,

$$L_n(\boldsymbol{\phi}_0) - L(\boldsymbol{\phi}_0) = \frac{1}{n}\sum_{i=1}^n \{h(\boldsymbol{X}_i, Y_i|\boldsymbol{0}) - E[h(\boldsymbol{X}, Y|\boldsymbol{0})]\}.$$

Here, under $Y_i > w_n(\boldsymbol{X}_i)$, $E_i = \exp[\alpha_0(\boldsymbol{X}_i^T\boldsymbol{\theta}_0)]\log(Y_i/w_n(\boldsymbol{X}_i))$ is asymptotically distributed as exponential distribution. Let M be the positive constant. Then, we have

$$J_1 \leq P\left(|L_n(\boldsymbol{\phi}_0)) - L_0(\boldsymbol{\phi}_0)| > 2^{-1}C\delta^2\tau_n \Big| \max_i E_i < M\right) + n\exp[-M].$$

It is easy to show that $V[h(\boldsymbol{X}_i, Y_i|\boldsymbol{0})] \leq C_1\tau_n$ for some constant $C_1 > 0$. Under $\max_i E_i < M$, we have $\max_i |h_1(\boldsymbol{X}_i, Y_i|\boldsymbol{0})| \leq C_2 M$ for some constant $C_2 > 0$. We set $M = 1/\delta_n$. Then, the Bernstein's inequality yields that

$$\begin{aligned}
&P\left(|L_n(\boldsymbol{\phi}_0) - L_0(\boldsymbol{\phi}_0)| > C\delta^2\tau_n \Big| \max_i E_i < M\right)\\
&\leq C_3 \exp\left[-C_3 \frac{\delta_n^2\tau_n}{\frac{\tau_n}{n} + \frac{M\delta_n^2\tau_n}{n}}\right]\\
&\leq C_3 \exp\left[-C_3 n\delta_n\right].
\end{aligned}$$

for some constant $C_3 > 0$. Since $\delta = \delta_n$ satisfies $\log(n) < \delta_n^{-1} < n$, we have $J_1 \to 0$ as $n \to \infty$.

We next evaluate J_2. From Taylor expansion, we obtain that there exists $\xi \in (0,1)$ such that

$$
\begin{aligned}
&h(\boldsymbol{X}_i, Y_i|\boldsymbol{u}) - h(\boldsymbol{X}_i, Y_i|0) \\
&= \delta_n \boldsymbol{u}^T A_1(\boldsymbol{X}_i|\boldsymbol{\phi}_0 + \xi\delta_n\boldsymbol{u}) \\
&\quad \times\{\exp[\alpha_0(\boldsymbol{X}_i^T\boldsymbol{\theta}(\boldsymbol{\phi}_0 + \xi\delta_n\boldsymbol{u})|\boldsymbol{\theta}(\boldsymbol{\phi}_0 + \xi\delta_n\boldsymbol{u}))]\log(Y_i/w_n(\boldsymbol{X}_i)) - 1\}I(Y_i > w_n(\boldsymbol{X}_i)) \\
&\equiv h_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}).
\end{aligned}
$$

Then, in J_2, we can write

$$
L_n(\boldsymbol{\phi}_0 + \delta\boldsymbol{u}) - L_n(\boldsymbol{\phi}_0) - L(\boldsymbol{\phi}_0 + \delta\boldsymbol{u}) + L(\boldsymbol{\phi}_0) = \frac{1}{n}\sum_{i=1}^n h_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}) - E[h_1(\boldsymbol{X}, Y|\boldsymbol{u})].
$$

Let $\mathcal{U} = \{\boldsymbol{u} \in \mathbb{R}^{p-1}|||\boldsymbol{u}|| \le 1\}$ be the vector space and $\mathcal{U}_1, \dots, \mathcal{U}_N$ be a covering of $\mathcal{U}$ with the diameter $R_n = C\delta_n^2/n^\nu$ for some constant $C > 0$ and $\nu > 0$. Then, Lemma 2.5 of van de Geer (2000) yields that $N \le C(n^\nu/\delta^2\tau_n)^{p-1}$. Let $\boldsymbol{u}_k \in \mathcal{U}_k, k = 1, \dots, N$ and let M_1 be positive constant. Define

$$
h_2(\boldsymbol{X}_i, Y_i|\boldsymbol{u}, \boldsymbol{u}_k) = h_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}) - E[h_1(\boldsymbol{X}, Y|\boldsymbol{u})] - \{h_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}_k) - E[h_1(\boldsymbol{X}, Y|\boldsymbol{u}_k)]\}.
$$

Then, we have

$$
\begin{aligned}
&P\left(\sup_{||\boldsymbol{u}||\le 1}\left|\frac{1}{n}\sum_{i=1}^n h_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}) - E[h_1(\boldsymbol{X}, Y|\boldsymbol{u})]\right| \ge 2^{-1}C\delta^2\tau_n\right) \\
&\le \sum_{k=1}^N P\left(\left|\frac{1}{n}\sum_{i=1}^n h_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}_k) - E[h_1(\boldsymbol{X}, Y|\boldsymbol{u}_k)]\right| + \sup_{\boldsymbol{u}\in\mathcal{U}_k}\left|\frac{1}{n}\sum_{i=1}^n h_2(\boldsymbol{X}_i, Y_i|\boldsymbol{u}, \boldsymbol{u}_k)\right| \ge 2^{-1}C\delta^2\tau_n\right) \\
&\le \sum_{k=1}^N P\left(\left|\frac{1}{n}\sum_{i=1}^n \{h_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}_k) - E[h_1(\boldsymbol{X}, Y|\boldsymbol{u}_k)]\}\right|\right. \\
&\quad \left. + \sup_{\boldsymbol{u}\in\mathcal{U}_k}\left|\frac{1}{n}\sum_{i=1}^n h_2(\boldsymbol{X}_i, Y_i|\boldsymbol{u}, \boldsymbol{u}_k)\right| \ge 2^{-1}C\delta^2\tau_n \middle| \max_i E_i \le M_1\right) + Nn\exp[-M_1].
\end{aligned}
$$

We set $M_1 = 1/\delta_n$. Since δ_n satisfies $\delta_n\log(n) \to 0$, $Nn\exp[-M_1]$ converges to zero. Similar to the evaluation of ℓ_2 in the proof of Lemma 4, under $\max_i E_i \le M_1$, we obtain that for some constant $c > 0$,

$$
\sup_{\boldsymbol{u}\in\mathcal{U}_k}\left|\frac{1}{n}\sum_{i=1}^n h_2(X_i, Y_i|\boldsymbol{u}, \boldsymbol{u}_k)\right| < cM_1\delta_n\tau_nR_n < 2^{-1}C\delta_n^2\tau_n.
$$

Lastly, we will evaluate

$$
P\left(\left|\frac{1}{n}\sum_{i=1}^n h_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}_k) - E[h_1(\boldsymbol{X}, Y|\boldsymbol{u}_k)]\right| \ge 2^{-1}C\delta_n^2\tau_n\right) \to 0.
$$

It is easy to show that $\max_i |h_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}_k)| \le C_1\delta_nM_1$ and $V[h_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}_k)] \le C_2\delta_n^2\tau_n$ for some constant $C_1, C_2 > 0$. Therefore, Bernsteins inequality realizes that

$$
\begin{aligned}
&P\left(\left|\frac{1}{n}\sum_{i=1}^n h_1(\boldsymbol{X}_i, Y_i|\boldsymbol{u}_k) - E[h_1(\boldsymbol{X}, Y|\boldsymbol{u}_k)]\right| \ge 2^{-1}C\delta^2\tau_n\right) \\
&\le C_3\exp\left[-\frac{\delta^4\tau_n^2}{C_1\frac{\delta^2\tau_n}{n} + C_2\frac{M_1\delta^2\tau_n}{n}}\right] \\
&\le C_3\exp\left[-n\tau_n\delta^2\right]
\end{aligned}
$$

for some constant $C_3 > 0$. Since $\delta_n (n\tau_n)^{1/2} \to \infty$, we obtain $J_2 \to 0$ as $n \to \infty$. This implies that $P(||\hat{\boldsymbol{\phi}} - \boldsymbol{\phi}_0|| > \delta_n) \to 0$ as $n \to \infty$.

$\square$

Lemma 6. *Suppose that (C1)–(C9). Then, as $n \to \infty$,*

$$\boldsymbol{B}(\boldsymbol{x}^T\boldsymbol{\theta}_0)^T(\hat{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{b}_0) = \boldsymbol{B}(\boldsymbol{x}^T\boldsymbol{\theta}_0)^T \ddot{U}_n(\boldsymbol{b}_0)^{-1}\dot{U}_n(\boldsymbol{b}_0)(1 + o_P(1)).$$

Proof of Lemma 6. Let

$$\begin{aligned}
&Q_n(\boldsymbol{u}) \\
&= \frac{1}{n}\sum_{i=1}^n \left\{\exp[\boldsymbol{B}^{[d]}(\boldsymbol{X}_i^T\hat{\boldsymbol{\theta}})^T\{\boldsymbol{b}_0 + \boldsymbol{u}/\eta_n\}]\log\left(\frac{Y_i}{w_n(\boldsymbol{X}_i)}\right) - \boldsymbol{B}^{[d]}(\boldsymbol{X}_i^T\hat{\boldsymbol{\theta}})^T\{\boldsymbol{b}_0 + \boldsymbol{u}/\eta_n\}\right\} I(Y_i > w_n(\boldsymbol{X}_i)) \\
&\quad + \frac{\lambda}{2}\{\boldsymbol{b}_0 + \boldsymbol{u}/\eta_n\}^T\Delta_{m,K}\{\boldsymbol{b}_0 + \boldsymbol{u}/\eta_n\}.
\end{aligned}$$

Then, the minimizer of Q is obtained as

$$\hat{\boldsymbol{u}} = \eta_n\{\hat{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{b}_0\}.$$

By the Taylor expansion, we have

$$Q_n(\boldsymbol{u}) = Q_n(\boldsymbol{0}) + \frac{1}{\eta_n}\boldsymbol{W}_n(\hat{\boldsymbol{\theta}})^T\boldsymbol{u} + \frac{1}{\eta_n^2}\boldsymbol{u}^T A_n(\hat{\boldsymbol{\theta}})\boldsymbol{u} + \frac{\lambda}{\eta_n}\boldsymbol{u}^T\Delta_{m,K}\{\boldsymbol{b}_0 + \boldsymbol{u}/\eta_n\}$$

where

$$\boldsymbol{W}_n(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{B}^{[d]}(\boldsymbol{X}_i^T\hat{\boldsymbol{\theta}})\left\{\exp[\boldsymbol{B}^{[d]}(\boldsymbol{X}_i^T\hat{\boldsymbol{\theta}})^T\boldsymbol{b}_0]\log\left(\frac{Y_i}{w_n(\boldsymbol{X}_i)}\right) - 1\right\} I(Y_i > w_n(\boldsymbol{X}_i)),$$

$||\boldsymbol{u}^*|| < ||\boldsymbol{u}||$ and

$$A_n(\hat{\boldsymbol{\theta}}) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{B}^{[d]}(\boldsymbol{X}_i^T\hat{\boldsymbol{\theta}})\boldsymbol{B}^{[d]}(\boldsymbol{X}_i^T\hat{\boldsymbol{\theta}})^T\exp[\boldsymbol{B}^{[d]}(\boldsymbol{X}_i^T\hat{\boldsymbol{\theta}})^T\{\boldsymbol{b}_0 + \boldsymbol{u}^*/\eta_n\}]\log\left(\frac{Y_i}{w_n(\boldsymbol{X}_i)}\right) I(Y_i > w_n(\boldsymbol{X}_i))$$

From Theorem 4, we obtain

$$\boldsymbol{W}_n(\hat{\boldsymbol{\theta}}) = \boldsymbol{W}_n(\boldsymbol{\theta}_0) + \boldsymbol{B}_n^T(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

where

$$\begin{aligned}
&\boldsymbol{B}_n \\
&= \frac{1}{n}\sum_{i=1}^n \frac{\partial \boldsymbol{B}^{[d]}(\boldsymbol{X}_i^T\boldsymbol{\theta}^*)^T\boldsymbol{u}}{\partial\boldsymbol{\theta}}\left\{\exp[\boldsymbol{B}^{[d]}(\boldsymbol{X}_i^T\boldsymbol{\theta}^*)^T\boldsymbol{b}_0]\log\left(\frac{Y_i}{w_n(\boldsymbol{X}_i)}\right) - 1\right\} I(Y_i > w_n(\boldsymbol{X}_i)) \\
&\quad + \frac{1}{n}\sum_{i=1}^n \boldsymbol{B}^{[d]}(\boldsymbol{X}_i^T\boldsymbol{\theta}^*)^T\boldsymbol{u}\frac{\partial \boldsymbol{B}^{[d]}(\boldsymbol{X}_i^T\boldsymbol{\theta}^*)^T\boldsymbol{b}_0}{\partial\boldsymbol{\theta}}\exp[\boldsymbol{B}^{[d]}(\boldsymbol{X}_i^T\boldsymbol{\theta}^*)^T\boldsymbol{b}_0]\log\left(\frac{Y_i}{w_n(\boldsymbol{X}_i)}\right) I(Y_i > w_n(\boldsymbol{X}_i))
\end{aligned}$$

and $||\boldsymbol{\theta}^* - \boldsymbol{\theta}_0|| < ||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0||$. From the proof of Theorem 1, we see that $||\boldsymbol{B}_n|| = O_P(\tau_n^{\beta_{min}+1})$ and $||\boldsymbol{W}_n(\hat{\boldsymbol{\theta}}) - \boldsymbol{W}_n(\boldsymbol{\theta}_0)|| = O_P(\tau_n^{\beta_{min}+1}||\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0||) = o_P(1)$ uniformly for $\boldsymbol{u} \in \mathcal{U} \subset \mathbb{R}^K$, where $\mathcal{U}$ is any compact subset of $\mathbb{R}^K$. Similarly, simple each element of $A_n(\hat{\boldsymbol{\theta}}) - A_n(\boldsymbol{\theta}_0)$ has an order

$o_P(1)$ uniformly for $\boldsymbol{u} \in \mathcal{U}$. Therefore, from the convexity lemma of Pollard (1991), we have that $Q_n(\boldsymbol{u}) - Q_n(\boldsymbol{0})$ is asymptotically equivalent to

$$\frac{1}{\eta_n}\boldsymbol{W}_n(\boldsymbol{\theta}_0)^T\boldsymbol{u} + \frac{1}{\eta_n^2}\boldsymbol{u}^T A_n(\boldsymbol{\theta}_0)\boldsymbol{u} + \frac{\lambda}{\eta_n}\boldsymbol{u}^T\Delta_{m,K}\{\boldsymbol{b}_0 + \boldsymbol{u}/\eta_n\}$$

uniformly for $\boldsymbol{u} \in \mathcal{U} \subset \mathbb{R}^K$. We note that $\dot{U}_n(\boldsymbol{b}_0, \boldsymbol{\theta}_0|\lambda) = \boldsymbol{W}_n(\boldsymbol{\theta}_0)$ and $\ddot{U}_n(\boldsymbol{b}_0, \boldsymbol{\theta}_0|\lambda) = A_n(\boldsymbol{\theta}_0) + \lambda\Delta_{m,K}$. This implies that

$$\begin{aligned} \eta_n\boldsymbol{B}(\boldsymbol{x}^T\boldsymbol{\theta}_0)^T(\hat{\boldsymbol{b}}(\hat{\boldsymbol{\theta}}) - \boldsymbol{b}_0) &= \eta_n\boldsymbol{B}(\boldsymbol{x}^T\boldsymbol{\theta}_0)^T(A_n(\boldsymbol{\theta}_0) + \lambda\Delta_{m,K})^{-1}\boldsymbol{W}_n(\boldsymbol{\theta}_0)(1 + o_P(1)) \\ &= \eta_n\boldsymbol{B}(\boldsymbol{x}^T\boldsymbol{\theta}_0)^T\ddot{U}_n(\boldsymbol{b}_0, \boldsymbol{\theta}_0|\lambda)^{-1}\dot{U}_n(\boldsymbol{b}_0, \boldsymbol{\theta}_0|\lambda)(1 + o_P(1)). \end{aligned}$$

Using Lemmas 2 and 3, we obtain

$$\eta_n\boldsymbol{B}(\boldsymbol{x}^T\boldsymbol{\theta}_0)^T\ddot{U}_n(\boldsymbol{b}_0, \boldsymbol{\theta}_0|\lambda)^{-1}\dot{U}_n(\boldsymbol{b}_0, \boldsymbol{\theta}_0|\lambda) = O_P(1).$$

□

References

[1] Barrow, D.L. and Smith, P.W. (1978). Asymptotic properties of best $L_2[0,1]$ approximation by spline with variable knots. *Quarterly of Applied Mathematics* **36** 293–304.

[2] Demko, S. (1977). Inverses of band matrices and local convergence of spline projections. *SIAM. Journal on Numerical Analysis.* **14** 616–619.

[3] Hjort ,N.L. and Pollard, D. (2011). Asymptotics for minimisers of convex processes. *https://arxiv.org/abs/1107.3806.*

[4] Pollard, D. (1991). Asymptotics for least absolute deviation regression estimators. *Econometric Theory.* **7** 186–199.

[5] van de Geer, S. (2000). *Empirical Processes in M-Estimation.* Cambridge University Press.

[6] Xiao, L. (2019). Asymptotic theory of penalized splines. *Electronic Journal of Statistics.* **13** 747–794.

[7] Zhou, S., Shen, X., and Wolfe, D.A. (1998). Local asymptotics for regression splines and confidence regions. *Annals of Statistics* **26** 1760–1782.