Statistica Sinica: Supplement

A NONPARAMETRIC REGRESSION MODEL FOR PANEL COUNT DATA ANALYSIS

Huadong Zhao[1], Ying Zhang[2], Xingqiu Zhao[3]and Zhangsheng Yu[4*]

[1]*East China Normal University,* [2]*Indiana University,*

[3]*The Hong Kong Polytechnic University and* [2,4]*Shanghai Jiao Tong University*

Throughput the following theoretical arguments, $\mathbb{P}_n$ and P denote the usual empirical and true probability measures for the observed data. C is a universal constant that may vary from place to place.

Proof of Theorem 1 Let $\mathbb{M}_n(\theta) = \mathbb{P}_n m(\theta; X)$ and $\mathbb{M}(\theta) = Pm(\theta; X)$, where

$$m(\theta; X) = \sum_{j=1}^{K} \left[\mathbb{N}(T_j) \log \Lambda(T_j) + \mathbb{N}(T_j)\beta(Z) - \Lambda(T_j) \exp\{\beta(Z)\} \right].$$

For the consistency of θ, we need to show that

(i) $\sup_{\theta \in \mathcal{F}_1 \times \mathcal{F}_2} |\mathbb{M}_n(\theta) - \mathbb{M}(\theta)| \to 0$ in probability as $n \to \infty$;

(ii) $\sup_{\theta : d(\theta, \theta_0) \geq \epsilon} \mathbb{M}(\theta) \leq \mathbb{M}(\theta_0)$; and

(iii) $\mathbb{M}_n(\hat{\theta}_n) \geq \mathbb{M}_n(\theta_0) - o_p(1)$

according to Theorem 5.7 of van der Vaart (2000).

First, to show (i), we need to demonstrate that $\mathcal{M}_1 = \{m(\theta; X), \theta \in \Phi_{l_2,z} \times \psi_{l_1,t}\}$ is a Glivenko-Cantelli (G-C) Class. By Lemma 1 in Lu, Zhang, and Huang (2007), and Jackson type Theorem (De Boor (2001), page 149), there exists $\Lambda_n \in \psi_{l_1,t}$, and $\beta_n \in \Phi_{l_2,z}$ with order $l_1 \geq r+2, l_2 \geq r+2$, and knots of T and Z satisfying C2, such that $\|\Lambda_n - \Lambda_0\|_\infty = \sup_{t \in O[T]} |\Lambda_n(t) - \Lambda_0(t)| \leq C q_{n1}^{-r} = O(n^{-rv_1})$, $\|\beta_n - \beta_0\|_\infty = \sup_{z \in O[Z]} |\beta_n(z) - \beta_0(z)| \leq C q_{n2}^{-r} = O(n^{-rv_2})$. By the same argument as in Wellner and Zhang (2007), it follows that Λ_n is also uniformly bounded. Following the calculation of Shen and Wong (1994), for arbitrary $\epsilon > 0$, there exists a set of bracket $\{[\log \Lambda_i^L, \log \Lambda_i^U] : i = 1, 2, \ldots, [(1/\epsilon)^{Cq_{n1}}]\}$, such that for any $\Lambda \in \psi_{l_1,t}$, we have $\log \Lambda_i^L(t) \leq \log \Lambda(t) \leq \log \Lambda_i^U(t)$ for some $1 \leq i \leq [(1/\epsilon)^{Cq_{n1}}]$ and all $t \in [\sigma_1, \tau_1]$, and $\mathbb{P}_n |\log \Lambda_i^U(t) - \log \Lambda_i^L(t)| \leq C\epsilon$. Similarly there exists a set of brackets $\{[\beta_s^L, \beta_s^U] : s = 1, 2, \ldots, [(1/\epsilon)^{Cq_{n2}}]\}$, such that for any $\beta \in \Phi_{l_2,z}$, we have $\beta_s^L(z) \leq \beta(z) \leq \beta_s^U(z)$ for some $1 \leq s \leq [(1/\epsilon)^{c_2 q_{n2}}]$ and all $z \in [\sigma_2, \tau_2]$, $\mathbb{P}_n |\beta_s^U(Z) - \beta_s^L(Z)| \leq C\epsilon$. So we can construct a set of brackets $\left\{m_{i,s}^L, m_{i,s}^U : i = 1, 2, \ldots, [(1/\epsilon)^{Cq_{n1}}], s = 1, 2, \ldots, [(1/\epsilon)^{Cq_{n2}}]\right\}$. For any $m(\theta; X) \in \mathcal{M}_1$, there exist $i \in \{1, 2, \ldots, [(1/\epsilon)^{Cq_{n1}}]\}$ and $s \in \left\{1, 2, \ldots, [(1/\epsilon)^{Cq_{n2}}]\right\}$

such that $m(\theta;X) \in [m_{i,s}^L, m_{i,s}^U]$, where

$$m_{i,s}^L = \sum_{j=1}^{K} [\mathbb{N}(T_j)\log\Lambda_i^L(T_j) + \mathbb{N}(T_j)\beta_s^L(Z) - \Lambda_i^U(T_j)\exp\{\beta_s^U(Z)\}] \text{ and}$$

$$m_{i,s}^U = \sum_{j=1}^{K} [\mathbb{N}(T_j)\log\Lambda_i^U(T_j) + \mathbb{N}(T_j)\beta_s^U(Z) - \Lambda_i^L(T_j)\exp\{\beta_s^L(Z)\}].$$

By C1, C5 and Taylor's expansion, it follows that $\mathbb{P}_n|m_{i,s}^U - m_{i,s}^L| \leq C\epsilon$ for all $i \in \{1, 2, \ldots, [(1/\epsilon)^{Cq_{n1}}]\}$ and $s \in \{1, 2, \ldots, [(1/\epsilon)^{Cq_{n2}}]\}$. So the bracketing number for $\mathcal{M}_1$ with $L_1(\mathbb{P}_n)$ norm is bounded by $C(1/\epsilon)^{Cq_{n1}+Cq_{n2}}$. By the relationship of covering and bracketing numbers (page 84 of van der Vaart and Wellner (2000)), we know $N(\epsilon, \mathcal{M}_1, L_1(\mathbb{P}_n)) \leq N_{[\ ]}(2\epsilon, \mathcal{M}_1, L_1(\mathbb{P}_n))$, and it results in $\log N(\epsilon, \mathcal{M}_1, L_1(\mathbb{P}_n)) = O(Cq_{n1} + Cq_{n2}) = o_p(n)$. Hence $\mathcal{M}_1$ is a **G-C** class by Theorem 2.4.3 of van der Vaart and Wellner (2000).

Second, to show (ii), we only need to prove $\mathbb{M}(\theta_0) - \mathbb{M}(\theta) \geq Cd^2(\theta, \theta_0)$. Following the same lines as given in Wellner and Zhang (2007), we have

$$\mathbb{M}(\theta_0) - \mathbb{M}(\theta) \geq C\mathrm{E}\Big(\sum_{j=1}^{\mathrm{K}} [\Lambda_0(\mathrm{T}_j)\exp\{\beta_0(Z)\} - \Lambda(\mathrm{T}_j)\exp\{\beta(Z)\}]^2\Big).$$

With conditions (C1)-(C5) and C7, by the same arguments as in Wellner and Zhang (2007)(page 2126-2127), yields that

$$\mathbb{M}(\theta_0) - \mathbb{M}(\theta) \geq \mathrm{C}\{\|\beta - \beta_0\|^2_{\mathrm{L}_2(\mu_2)} + \| \Lambda - \Lambda_0 \|^2_{\mathrm{L}_2(\mu_1)}\} = \mathrm{C}d^2(\theta, \theta_0).$$

Third, we use the relationship of P-Donsker Class and asymptotic equicontinuity to prove (iii). Similar to the proof for (i), for $(\beta_0, \Lambda_0) \in \mathcal{F}_1 \times \mathcal{F}_2$, there exists $\beta_n \in \Phi_{l_2,z}$ and $\log \Lambda_n \in \psi_{l_1,t}$ with order $l_1 \geq r+2, l_2 \geq r+2$ such that $\|\beta_n - \beta_0\|_\infty \leq C q_{n1}^{-r} = O(n^{-rv_1})$, $\|\log \Lambda_n - \log \Lambda_0\|_\infty \leq C q_{n2}^{-r} = O(n^{-rv_2})$. Now let $\theta_n = (\Lambda_n, \beta_n)$, we have

$$\begin{aligned}\mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) &= \mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_n) + \mathbb{M}_n(\theta_n) - \mathbb{M}_n(\theta_0) \\ &\geq \mathbb{M}_n(\theta_n) - \mathbb{M}_n(\theta_0) \\ &= (\mathbb{P}_n - P)\{m(\theta_n; X) - m(\theta_0; X)\} + \mathbb{M}(\theta_n) - \mathbb{M}(\theta_0).\end{aligned}$$

We consider the class: $\mathcal{M}_2 = \{m(\theta; X) : \theta \in \Phi_{l_2,z} \times \psi_{l_1,t}, \|\Lambda - \Lambda_0\|_\infty \leq C q_{n1}^{-r}, \|\beta - \beta_0\|_\infty \leq C q_{n2}^{-r}\}$. It is obvious that $m(\theta; X) \leq m^B(\theta; X)$ with

$$m^B(\theta; X) = \sum_{j=1}^{K} \big(\mathbb{N}(T_K) \log[\Lambda(T_j) \exp\{\beta(Z)\}] - \Lambda(T_j) \exp\{\beta(Z)\}\big)$$

By the boundness of $\beta(Z)$ and $\Lambda(t)$ in $\mathcal{M}_2$, we can have $a \leq \Lambda(T) \exp\{\beta(Z)\} \leq b$ for some $a < 1$ and $b > 1$ and then $\{a \leq \Lambda(T) \exp \beta(Z) \leq b\} = \{a \leq \Lambda(T) \exp \beta(Z) < 1\} \bigcup \{1 \leq \Lambda(T) \exp \beta(Z) \leq b\}$.

For $\{a \leq \Lambda(t) \exp \beta(Z) \leq 1\}$, denote $B_{1,j} = \big\{ \log[\Lambda(T_j) \exp\{\beta(Z)\}] / \log a \big\}$ and $\mathcal{G}_{1,j} = \{I_{[\sigma_1, T_j] \times [\sigma_2, Z]}, \sigma_1 \leq T_j \leq \tau_1, \sigma_2 \leq Z \leq \tau_2\}$. We know that

$$0 \leq \log[\Lambda(T_j) \exp\{\beta(Z)\}] / \log a \leq 1,$$

therefore $B_{1,j} \subseteq \overline{sconv} \mathcal{G}_{1,j}$, the closure of the symmetric convex hull of $\mathcal{G}_{1,j}$.

Hence we have

$$N(\varepsilon, \mathcal{G}_{1,j}, L_2(Q_{C_1,C_2})) \leq C\,(1/\varepsilon)^8$$

for any probability measure Q_{C_1,C_2} of (C_1, C_2). Since $V(\mathcal{G}_{1,j}) = 5$ and the envelop function of $\mathcal{G}_{1,j}$ is 1. The above equation yields that

$$\log N(\varepsilon, \overline{sconv}\mathcal{G}_{1,j}, L_2(Q_{C_1,C_2})) \leq C\,(1/\varepsilon)^{10/7},$$

according to Theorem of 2.6.9 in van der Vaart and Wellner (2000). Hence it follows that $\log N(\varepsilon, B_{1,j}, L_2(Q_{C_1,C_2})) \leq C\,(1/\varepsilon)^{10/7}$.

Let $B'_{1,j} = \left\{\mathbb{N}(T_K)\log[\Lambda(T_j)\exp\{\beta(Z)\}]\right\}$. Suppose the centers of ε-balls of $B_{1,j}$ are $f_{i,j}^{B_1}$, for $i = 1, 2, \ldots, [C(1/\varepsilon)^{10/7}]$, then for any probability measure Q,

$$\begin{aligned}
&\left\|\sum_{j=1}^{K}\mathbb{N}(T_K)\log[\Lambda(T_j)\exp\{\beta(Z)\}] - \sum_{j=1}^{K}\mathbb{N}(T_K)\log a f_{i,j}^{B_1}\right\|_{L_2(Q)}^2\\
=&Q\left(\sum_{j=1}^{K}\mathbb{N}(T_K)\log[\Lambda(T_j)\exp\{\beta(Z)\}] - \sum_{j=1}^{K}\mathbb{N}(T_K)\log a f_{i,j}^{B_1}\right)^2\\
\leq&CQ\left(\sum_{j=1}^{K}\mathbb{N}^2(T_K)\right)Q\left(\sum_{j=1}^{K}\left\{\log[\Lambda(T_j)\exp\{\beta(Z)\}] - \log a f_{i,j}^{B_1}\right\}\right)^2 \quad \text{by C9}\\
\leq&E\{e^{C\mathbb{N}(T_K)}\}(\log a)^2 Q\left(\sum_{j=1}^{K}\left\{\frac{\log[\Lambda(T_j)\exp\{\beta(Z)\}]}{\log a} - f_{i,j}^{B_1}\right\}\right)^2\\
\leq&C\varepsilon^2 \quad \text{by C10.}
\end{aligned}$$

Let $\tilde{\varepsilon} = \sqrt{C}\varepsilon$, then $\mathbb{N}(T_{K,K}))\log a f_{i,j}^{B_1}, i = 1, 2, \ldots, [C(1/\varepsilon)^{10/7}]$, are the centers of $\tilde{\varepsilon}$ balls of $B'_{1,j}$. Hence we have $\log N(\tilde{\varepsilon}, B'_{1,j}, L_2(Q)) \leq C(1/\varepsilon)^{10/7}$, and this yields that

$$\int_0^1 \sup_Q \sqrt{\log N(\tilde{\varepsilon}, B'_{1,j}, L_2(Q))} d\varepsilon \leq \int_0^1 \sqrt{C}(1/\varepsilon)^{5/7} d\varepsilon \leq \infty.$$

The envelop function of $B'_{1,j}$ is $-\mathbb{N}(T_{K,K}) \log a$, which has finite moments by C3, C5 and C10. Therefore $B'_{1,j}$ is a P-Donsker by Theorem 2.5.2 in van der Vaart and Wellner (2000). Similarly, for $\{1 \leq \Lambda(t) \exp \beta(Z) \leq b\}$, we can show that $B'_{1,j} = \left\{\mathbb{N}(T_K) \log[\Lambda(T_j) \exp\{\beta(Z)\}]\right\}$ is P-Donsker class, which implies that the class made by $\mathbb{N}(T_K) \log[\Lambda(T_j) \exp\{\beta(Z)\}]$ is P-Donsker for $\Lambda(T)$ and $\beta(Z)$ satisfying $a \leq \Lambda(T) \exp\{\beta(Z)\} \leq b$. Following the same argument, we can show that the class made by $\Lambda(T_j) \exp\{\beta(Z)\}$ is also P-Donsker and hence the class made by $m^B(\Lambda, \beta; X)$ is P-Donsker. Therefore $\mathcal{M}_2$ is P-Donsker due to the fact that every element in $\mathcal{M}_2$ is bounded by $m^B(\Lambda, \beta; X)$.

Moreover it is easily shown by dominated convergence theorem that

$$P\{m(\theta; X) - m(\theta_0; X)\}^2 \to \ 0 \ as \ n \to \infty$$

for any $m(\Lambda, \beta) \in \mathcal{M}_2$. Hence by Corollary 2.3.12 in van der Vaart and Wellner (2000), it follows that $(\mathbb{P}_n - P)\{m(\theta_n; X) - m(\theta_0; X)\} = o_p(n^{-1/2})$. Using the dominated convergence theorem again, it can be concluded that $\mathbb{M}(\theta_n) - \mathbb{M}(\theta_0) > -o(1)$ as $n \to \infty$. Hence $\mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) \geq o_p(n^{-1/2}) - o(1) = -o_p(1)$. Therefore, $d(\hat{\theta}_n, \theta_0) \to_p 0$ as $n \to \infty$.

Proof of Theorem 2 In order to derive the rate of convergence, we need to verify the conditions of theorem 3.2.5 of van der Vaart and Wellner (2000).

First, we have already shown that $\mathbb{M}(\theta_0) - \mathbb{M}(\theta) \geq Cd^2(\theta, \theta_0)$.

Second, in the previous proof, we know $\mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) \geq I_{1,n} + I_{2,n}$, where $I_{1,n} = (\mathbb{P}_n - P)\{m(\theta_n; X) - m(\theta_0; X)\}$ and $I_{2,n} = P(m(\theta_n; X) - m(\theta_0; X))$. Let $\theta_\xi = \theta_0 + \xi(\theta_n - \theta_0)$ for $0 < \xi < 1$. Taylor expansion of $m(\theta_n; X)$ at θ_0 leads to,

$$
\begin{aligned}
I_{1,n} =& (\mathbb{P}_n - P)\{\dot{m}_1(\theta_\xi; X)(\Lambda_n - \Lambda_0) + \dot{m}_2(\theta_\xi; X)(\beta_n - \beta_0)\} \\
&= n^{-rv_1+\varepsilon}(\mathbb{P}_n - P)\frac{\dot{m}_1(\theta_\xi; X)(\Lambda_n - \Lambda_0)}{n^{-rv_1+\varepsilon}} + n^{-rv_2+\varepsilon}(\mathbb{P}_n - P)\frac{\dot{m}_2(\theta_\xi; X)(\beta_n - \beta_0)}{n^{-rv_2+\varepsilon}}
\end{aligned}
$$

for some $0 < \xi < 1$ and $0 < \varepsilon < min\{1/2 - rv_1, 1/2 - rv_2\}$, here $\dot{m}_1(\theta_\xi; X) = \sum_{j=1}^K \left[\frac{\mathbb{N}(T_j)}{\Lambda_\xi} - \exp\{\beta_\xi\}\right]$, $\dot{m}_2(\theta_\xi; X) = \sum_{j=1}^K [\mathbb{N}(T_j) - \Lambda_\xi \exp\{\beta_\xi\}]$. Because $\|\beta_n - \beta_0\|_\infty \leq Cq_{n1}^{-r} = O(n^{-rv_1})$, $\|\Lambda_n - \Lambda_0\|_\infty \leq Cq_{n2}^{-r} = O(n^{-rv_2})$ and $\dot{m}_1(\theta_\xi; X)(\Lambda_n - \Lambda_0)$, $\dot{m}_2(\theta_\xi; X)(\beta_n - \beta_0)$ are uniformly bounded. We can conclude that $P\{\frac{\dot{m}_1(\theta_\xi;X)(\Lambda_n - \Lambda_0)}{n^{-rv_1+\varepsilon}}\}^2 \to 0$ and $P\{\frac{\dot{m}_2(\theta_\xi;X)(\beta_n - \beta_0)}{n^{-rv_2+\varepsilon}}\}^2 \to 0$. We know $\mathcal{M}_2$ is Donsker in the proof of consistency, according to corollary 2.3.12 of van der Vaart and Wellner (2000) again, we can obtain that $(\mathbb{P}_n - P)\{\frac{\dot{m}_1(\theta_\xi;X)(\Lambda_n - \Lambda_0)}{n^{-rv_1+\varepsilon}}\} + (\mathbb{P}_n - P)\{\frac{\dot{m}_2(\theta_\xi;X)(\beta_n - \beta_0)}{n^{-rv_2+\varepsilon}}\} = o_p(n^{-1/2})$. Hence $I_{1,n} = o_p(n^{-rv_1+\varepsilon}n^{-1/2}) + o_p(n^{-rv_2+\varepsilon}n^{-1/2}) = o_p(n^{-2r\max(v_1,v_2)})$. By the inequality of $h(x) = x\log x - x + 1 \leq (x-1)^2$ in the neighbourhood of $x = 1$, it can be easily to conclude that

$$\mathbb{M}(\theta_0) - \mathbb{M}(\theta_n) \leq C(\|\Lambda_0 - \Lambda_n\|^2_{L_2(\mu_1)} + \|\beta_0 - \beta_n\|^2_{L_2(\mu_2)}) = O(n^{-2\min\{rv_1, rv_2\}}).$$

So we conclude that $\mathbb{M}(\theta_n) - \mathbb{M}(\theta_0) \geq -O(n^{-2\min\{rv_1, rv_2\}})$. Thus, we conclude that $\mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) \geq -O(n^{-2\min\{rv_1, rv_2\}})$.

Third, for any $\delta > 0$, define the class

$$\mathcal{M}_\delta(\theta_0) = \{m(\theta; X) - m(\theta_0; X) : \theta \in \Phi_{l_2,z} \times \psi_{l_1,t}, d(\theta, \theta_0) \leq \delta\}.$$

Some algebra yields that $|\mathbb{M}(\theta) - \mathbb{M}(\theta_0)| \leq C\delta^2$ for any $m(\theta) - m(\theta_0) \in \mathcal{M}_\delta(\theta_0)$. Hence, by the Lemma 3.4.3 in van der Vaart and Wellner (2000), we obtain

$$E_P \|n^{1/2}(\mathbb{P}_n - P)\|_{\mathcal{M}_\delta} \leq C J_{[\,]}(\delta, \mathcal{M}_\delta, \|\,.\,\|_{P,B}) \left\{1 + \frac{J_{[\,]}(\delta, \mathcal{M}_\delta, \|\,.\,\|_{P,B})}{\delta^2 n^{1/2}}\right\}$$

where $J_{[\,]}(\delta, \mathcal{M}_\delta, \|\,.\,\|_{P,B}) = \int_0^\delta \sqrt{1 + \log N_{[\,]}(\varepsilon, \mathcal{M}_\delta, \|\,.\,\|_{P,B})} d\varepsilon \leq C q_n^{1/2} \delta$, $q_n = q_{n1} + q_{n2}$. The right side of the last equation yields $\phi_n(\delta) = C(q_n^{1/2}\delta + q_n/n^{1/2})$. Because $\phi(\delta)/\delta$ is a decrease function of δ, and $r_n^2 \phi(1/r_n) = r_n q_n^{1/2} + r_n^2 q_n / n^{1/2} \leq n^{1/2}$ yields that $r_n \leq n^{(1-\max\{v_1, v_2\})/2}$, and we have proved that $\mathbb{M}_n(\hat{\theta}_n) - \mathbb{M}_n(\theta_0) \geq -O(n^{-2\min\{rv_1, rv_2\}})$ in the second part. So by theorem 3.2.5 of van der Vaart and Wellner (2000),

When $r_n = n^{\min\{\min\{rv_1, rv_2\}, (1-\max\{v_1, v_2\})/2\}}$, we conclude that $r_n d(\hat{\theta}_n, \theta_0) = O_p(1)$. If $v_1 = v_2 = 1/(1+2r)$, then $n^{r/(1+2r)} d(\hat{\theta}_n, \theta_0) = O_p(1)$.

Proof of Theorem 3 Define the set

$$\mathcal{H} \equiv \mathrm{H}_\Lambda \times \mathrm{H}_\beta = \{h = (h_1, h_2) : h_1 \in BV[\sigma_1, \tau_1], h_2 \in C[\sigma_2, \tau_2]\},$$

where $BV[\sigma_1, \tau_1]$ is the Banach space consisting of all the functions with bounded total variation in $[\sigma_1, \tau_1]$, and $C[\sigma_2, \tau_2]$ is the Banach space consisting of all the continuous functions in $[\sigma_2, \tau_2]$. We define a sequence of maps S_n mapping a neighborhood of (Λ_0, β_0), denoted by $\mathcal{U}$, in the parameter space for $\theta = (\beta, \Lambda)$ into $l^\infty(\mathcal{H})$ as:

$$\begin{aligned} S_n(\theta)[h_1, h_2] &= n^{-1}\frac{dl_n(\Lambda + \varepsilon h_1, \beta + \varepsilon h_2)}{d\varepsilon}|_{\varepsilon=0} = A_{n_1}(\theta)[h_1] + A_{n_2}(\theta)[h_2] \\ &= \mathbb{P}_n\varphi(\theta; X)[h], \end{aligned}$$

where

$$\begin{aligned} & l_n(\Lambda, \beta) = \sum_{i=1}^n m(\theta; X_i) \\ & A_{n_1}(\theta)[h_1] \equiv n^{-1}\sum_{i=1}^n\sum_{j=1}^{K_i}\Big\{\frac{\mathbb{N}(\mathrm{T}_{i,j})}{\Lambda(\mathrm{T}_{i,j})} - \exp\{\beta(Z_i)\}\Big\}h_1(\mathrm{T}_{i,j}), \\ & A_{n_2}(\theta)[h_2] \equiv n^{-1}\sum_{i=1}^n\sum_{j=1}^{K_i}\Big\{\mathbb{N}(\mathrm{T}_{i,j}) - \Lambda(\mathrm{T}_{i,j})\exp\{\beta(Z_i)\}\Big\}h_2(Z_i), \end{aligned}$$

and

$$\varphi(\theta; X)[h] = \sum_{j=1}^K\Big\{\frac{\mathbb{N}(T_j)}{\Lambda(T_j)} - \exp\{\beta(Z)\}\Big\}h_1(T_j) + \sum_{j=1}^K\Big\{\mathbb{N}(T_j) - \Lambda(T_j)\exp\{\beta(Z)\}\Big\}h_2(Z).$$

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

$$S(\theta) = A_1(\theta)[h_1] + A_2(\theta)[h_2] = P\varphi(\theta; X)[h],$$

where

$$A_1(\theta)[h_1] = P\left[\sum_{j=1}^{K}\left\{\frac{\mathbb{N}(T_j)}{\Lambda(T_j)} - \exp\{\beta(Z)\}\right\}h_1(T_j)\right],$$
$$A_2(\theta)[h_2] = P\left[\sum_{j=1}^{K}\{\mathbb{N}(T_j) - \Lambda(T_j)\exp\{\beta(Z)\}\}\,h_2(Z)\right].$$

To derive the asymptotic normality of a class of smooth functionals of the estimator of $(\hat{\beta}_n, \hat{\Lambda}_n)$, we need to verify the following five conditions given by Theorem 3.3.1 of van der Vaart and Wellner (2000).

(A1) $(S_n - S)(\hat{\beta}_n, \hat{\Lambda}_n)[h] - (S_n - S)(\beta_0, \Lambda_0)[h] = o_p(n^{-1/2})$.

(A2) $S(\beta_0, \Lambda_0)[h] = 0$ and $S_n(\hat{\beta}_n, \hat{\Lambda}_n)[h] = o_p(n^{-1/2})$.

(A3) $\sqrt{n}(S_n - S)(\beta_0, \Lambda_0)[h]$ converges in distribution to a tight Gaussian process on $l^\infty(\mathcal{H})$.

(A4) $S(\beta, \Lambda)[h]$ is Frechet-differential at (β_0, Λ_0) with a continuously invertible derivative $\dot{S}(\beta_0, \Lambda_0)[h]$.

(A5) $S(\hat{\beta}_n, \hat{\Lambda}_n)[h] - S(\beta_0, \Lambda_0)[h] - \dot{S}(\beta_0, \Lambda_0)(\hat{\Lambda}_n - \Lambda_0, \hat{\beta}_n - \beta_0)[h] = o_p(n^{-1/2})$.

For (A1), define

$$\mathcal{G}_n^\delta[h] = \Bigg\{\varphi(\theta; X)[h] : \sup_{\sigma_1 \le t \le \tau_1} |\Lambda(t) - \Lambda_0(t)| < \delta, \sup_{\sigma_2 \le z \le \tau_2} |\beta(z) - \beta_0(z)| < \delta, \Lambda \in \psi_{l_1,t}, \beta \in \Phi_{l_2,z}, (h_1, h_2) \in \mathcal{H}\Bigg\}.$$

Similar to the same argument as that in the proof of consistency, we can show that $\mathcal{G}_n^\delta[h]$ is P-Donsker.

By Corollary 2.3.12 of van der Vaart and Wellner (2000), we can obtain

$$(\mathbb{P}_n - P)(\varphi(\hat{\theta}_n; X)[h] - \varphi(\theta_0; X)[h]) = o_p(n^{-1/2}),$$

uniformly in h. Thus, (A1) holds.

For (A2), the assumption of the proportional mean model immediately leads to $S(\theta_0)[h] = 0$ for $h \in \mathcal{H}$. Next we show that $S_n(\hat{\theta}_n)[h] = o_p(n^{-1/2})$.

Note that $\hat{\theta}_n$ maximizes $l_n(\Lambda, \beta)$ over $\Lambda \in \psi_{l_1,t}$ and $\beta \in \Phi_{l_2,z}$. It implies that

$$0 \equiv \frac{\partial l_n(\hat{\Lambda}_n + \varepsilon h_{n1}, \hat{\beta}_n + \varepsilon h_{n2})}{\partial \varepsilon},$$

for any $h_{n1} \in \psi_{l_1,t}$ and $h_{n2} \in \Phi_{l_2,z}$, which yields $S_n(\hat{\theta}_n)[h_{n1}, h_{n2}] = 0$.

For any $h = (h_1, h_2) \in \mathcal{H}$, there exist $h_n = (h_{n1}, h_{n2})$ for $h_{n_1} \in \psi_{l_1,t}$ and $h_{n_2} \in \Phi_{l_2,z}$ such that $\|h_{n_1} - h_1\|_\infty = O(n^{-rv_1}), \|h_{n_2} - h_2\|_\infty = O(n^{-rv_2})$. Then it suffices to show that

$$S_n(\hat{\theta}_n)[h - h_n] = S_n(\hat{\theta}_n)[h_1 - h_{n1}, h_2 - h_{n2}] = o_p(n^{-1/2})$$

Note that

$$
\begin{aligned}
S_n(\hat{\theta}_n)[h-h_n] &= \mathbb{P}_n\varphi(\hat{\theta}_n;X)[h-h_n] \\
&= (\mathbb{P}_n-P)\varphi(\hat{\theta}_n;X)[h-h_n]+P\varphi(\hat{\theta}_n;X)[h-h_n] \\
&= (\mathbb{P}_n-P)\varphi(\hat{\theta}_n;X)[h-h_n]+P\left(\varphi(\hat{\theta}_n;X)-\varphi(\theta_0;X)\right)[h-h_n] \\
&= I_{1n}+I_{2n}.
\end{aligned}
$$

Because $\mathcal{G}_n^\delta[h]$ is P-Donsker demonstrated for (A1) and $P\left(\varphi(\hat{\theta}_n;X,Z)[h-h_n]\right)^2 \rightarrow_p 0$ due to the approximation of h_n to h, it follows $I_{1n}=o_p(n^{-1/2})$ by Corollary 2.3.12 of van der Vaart and Wellner (2000). The rate of convergence of $\hat{\theta}_n$ and the approximation of h_n to h immediately leads to $I_{2n}=o_p(n^{-1/2})$. Hence (A2) is justified.

Condition (A3) holds because $\mathcal{H}$ is P-Donsker and the functionals A_{n1}, A_{n2} are bounded Lipschitz functions with respect to $\mathcal{H}$ (the same argument as in van der Vaart and Wellner (2000),Example 3.3.7 on page 312).

For (A4), by the smoothness of $S(\beta,\lambda)$, the Frechet differentiability holds and the derivative of S at (Λ_0,β_0) , denoted by $\dot{S}(\beta_0,\Lambda_0)$, is a map from the space $\{(\Lambda-\Lambda_0,\beta-\beta_0):(\Lambda,\beta)\in\mathcal{U}\}$ to $l^\infty(\mathcal{H})$ and

$$
\begin{aligned}
&\dot{S}(\beta_0,\Lambda_0)(\Lambda-\Lambda_0,\beta-\beta_0)[h] \\
=&\frac{d}{d\varepsilon}\{A_1(\theta_0+\varepsilon(\theta-\theta_0))[h_1]\}\mid_{\varepsilon=0}+\frac{d}{d\varepsilon}\{A_2(\theta_0+\varepsilon(\theta-\theta_0))[h_2]\}\mid_{\varepsilon=0} \\
=&\int(\beta(z)-\beta(z_0))dQ_1(h_1,h_2)(z)+\int(\Lambda(t)-\Lambda_0(t))dQ_2(h_1,h_2)(t),
\end{aligned}
$$

where

$$Q_1(h_1,h_2)(z) = P\Big\{\exp(\beta_0(Z))I(Z\le z)\sum_{j=1}^{K}\Big(h_1(T_j)+\Lambda_0(T_j)h_2(Z)\Big)\Big\}$$

$$Q_2(h_1,h_2)(t) = P\Big\{\exp(\beta_0(Z))\sum_{j=1}^{K}\frac{I((T_j)\le t)}{\Lambda_0(T_j)}\Big(h_1(T_j)+\Lambda_0(T_j)h_2(Z)\Big)\Big\}.$$

To demonstrate $\dot{S}(\beta_0,\Lambda_0)[h]$ is invertible, we need to show that $Q=(Q_1,Q_2)$ is one to one and it is equivalent to show that for $h\in\mathcal{H}$, if $Q(h_1,h_2)=0$, then $h_1=0, h_2=0$. Suppose that $Q(h_1,h_2)=0$. Then $\dot{S}(\beta_0,\Lambda_0)(\Lambda-\Lambda_0,\beta-\beta_0)[h_1,h_2]=0$ for any (β,Λ) in the neighborhood $\mathcal{U}$. In particular, we take $\Lambda=\Lambda_0+\varepsilon h_1$ and $\beta=\beta_0+\varepsilon h_2$, for a small constant ε. A simple algebra leads to

$$\dot{S}(\beta_0,\Lambda_0)(\Lambda-\Lambda_0,\beta-\beta_0)[h_1,h_2] = -\varepsilon P\Big[\exp(\beta_0(Z))\sum_{j=1}^{K}\Lambda_0(T_j)\Big\{\frac{h_1(T_j)}{\Lambda_0(T_j)}+h_2(Z)\Big\}^2\Big],$$

which yields

$$\frac{h_1(T_j)}{\Lambda_0(T_j)}+h_2(Z)=0,\quad j=1,\dots,K,\quad a.e.$$

and so $h_1\equiv 0, h_2\equiv 0$ by C6.

Next we show that (A5) holds. By Taylor expansion

$$\begin{aligned}
S(\hat\beta_n,\hat\Lambda_n)[h] &- S(\beta_0,\Lambda_0)[h]\\
&= \dot{S}(\beta_0,\Lambda_0)(\hat\Lambda_n-\Lambda_0,\hat\beta_n-\beta_0)[h]+O_p\left(\|\hat\Lambda_n-\Lambda_0\|^2_{L_2(\mu_1)}+\|\hat\beta_n-\beta_0\|^2_{L_2(\mu_2)}\right)\\
&= \dot{S}(\beta_0,\Lambda_0)(\hat\Lambda_n-\Lambda_0,\hat\beta_n-\beta_0)[h]+o_p(n^{-1/2})
\end{aligned}$$

by the rate of convergence of $\hat{\theta}_n$ given in Theorem 2.

Finally, it follows that

$$\sqrt{n}\int(\hat{\Lambda}_n(t)-\Lambda_0(t))dQ_2(h_1,h_2)(t)+\sqrt{n}\int(\hat{\beta}_n(z)-\beta(z_0))dQ_1(h_1,h_2)(z)$$
$$=\sqrt{n}(S_n-S)(\beta_0,\Lambda_0)[h]+o_p(1) \qquad \text{by (A1) and (A2).}$$

For any $h=(h_1,h_2)\in\mathcal{H}$, since Q is invertible, there exists an $h^*=(h_1^*,h_2^*)\in\mathcal{H}$ such that

$$Q_2(h_1^*,h_2^*)=h_1, \quad Q_1(h_1^*,h_2^*)=h_2.$$

Therefore, we have

$$\sqrt{n}\int(\hat{\Lambda}_n(t)-\Lambda_0(t))dh_1(t)+\sqrt{n}\int(\hat{\beta}_n(z)-\beta_0(z))dh_2(z)$$
$$=\sqrt{n}(S_n-S)(\beta_0,\Lambda_0)[h^*]+o_p(1)\to_d N(0,\sigma^2),$$

where $\sigma^2=E\{\varphi^2(\theta_0;X)[h^*]\}$. The proof is complete.

In fact, we can establish the asymptotic normality for the functionals of $\hat{\Lambda}_n(t)$ and $\hat{\beta}_n(z)$ separately by choosing a proper h^*. For example, if we take

$$h_1^*(T_j)=\frac{-\Lambda_0(T_j)E\{h_2^*(Z)\exp(\beta_0(Z))|K,T_j\}}{E\{\exp(\beta_0(Z))|K,T_j\}}, \quad \text{for all } \ j=1,2,\cdots,K$$

then

$$
\begin{aligned}
&Q_2(h_1^*, h_2^*)(t) \\
&\quad = E\left[\sum_{j=1}^{K} \exp(\beta_0(Z)) \frac{I(T_j \le t)}{\Lambda_0(T_j)} \{\Lambda_0(T_j) h_2^*(Z) + h_1^*(T_j)\}\right] \\
&\quad = E\left[\sum_{j=1}^{K} I(T_j \le t) \left\{ E\{h_2^*(Z)\exp(\beta_0(Z))|K, T_j\} + \frac{h_1^*(T_j)}{\Lambda_0(T_j)} E\{\exp(\beta_0(Z))|K, T_j\} \right\}\right] \\
&\quad = 0.
\end{aligned}
$$

Furthermore, for this chosen h^*, we have

$$
\begin{aligned}
&Q_1(h_1^*, h_2^*)(z) \\
&\quad = E\left[\exp(\beta_0(Z)) I(Z \le z) \sum_{j=1}^{K} \Lambda_0(T_j) \left\{ h_2^*(Z) - \frac{E\{h_2^*(Z)\exp(\beta_0(Z))|K, T_j\}}{E\{\exp(\beta_0(Z))|K, T_j\}} \right\}\right]
\end{aligned}
$$

and

$$
\sigma_\beta^2 = E\left[\sum_{j=1}^{K} \left\{ (\mathbb{N}(T_j) - \Lambda_0(T_j)\exp(\beta_0(Z))) \left(h_2^*(Z) - \frac{E\{h_2^*(Z)\exp(\beta_0(Z))|K, T_j\}}{E\{\exp(\beta_0(Z))|K, T_j\}} \right) \right\}\right]^2.
$$

Then Theorem 3 results in

$$
\sqrt{n} \int (\hat{\beta}_n(z) - \beta_0(z)) dQ_1(h_1^*, h_2^*)(z) \to_d N(0, \sigma_\beta^2).
$$

Validity of bootstrap nonparametric inference

Finally, we provide a justification for validating the test statistic described in Section 3. Following the discussion above, we can choose a specific $h^* = (h_1^*, h_2^*)$ such that

$$
Q_1(h_1^*, h_2^*)(t) = 0 \quad \text{and} \quad Q_2(h_1^*, h_2^*)(z) = H(z)
$$

and

$$\sqrt{n}\int\left(\hat{\beta}_n(z)-\beta(z)\right)dH(z)\rightarrow_d N(0,\sigma_\beta^2).$$

In the following, let $\mathbb{P}_n$ and P denote the empirical and true probability measures of Z, respectively, then we can rewrite the above asymptotic normality as

$$\sqrt{n}P(\hat{\beta}_n-\beta)\rightarrow_d N(0,\sigma_\beta^2).$$

Note that

$$\begin{aligned}
&\sqrt{n}\left(\int\hat{\beta}_n(z)d\mathbb{H}_n(z)-\int\beta(z)dH(z)\right)=\sqrt{n}(\mathbb{P}_n\hat{\beta}_n(Z)-P\beta(Z))\\
&=\sqrt{n}\left[(\mathbb{P}_n-P)\hat{\beta}_n(Z)+P(\hat{\beta}_n(Z)-\beta(Z))\right]\\
&=\sqrt{n}(\mathbb{P}_n-P)\beta(Z)+\sqrt{n}(\mathbb{P}_n-P)(\hat{\beta}_n(Z)-\beta(Z))+\sqrt{n}P(\hat{\beta}_n(Z)-\beta(Z))
\end{aligned}$$

By the ordinary central limit theorem, it follows that

$$\sqrt{n}(\mathbb{P}_n-P)\beta(Z)\rightarrow_d N\left(0,P(\beta(Z)-P\beta(Z))^2\right)$$

Using the same empirical process theorem arguments as above, we can show that of $\mathcal{G}^1=\{(\beta_n-\beta);\beta_n\in\Phi_{l_2,z}\}$ is P-Donsker.By the consistency $\hat{\beta}_n$, $P(\hat{\beta}_n-\beta)^2\rightarrow_p 0$ and the asymptotic equicontinuity theorem (Corollary 2.3.12 of van der Vaart and Wellner (2000), it follows that $\sqrt{n}(\mathbb{P}_n-P)(\hat{\beta}_n(Z)-\beta(Z))=o_p(1)$ and hence

$$\sqrt{n}\left(\int\hat{\beta}_n(z)d\mathbb{H}_n(z)-\int\beta(z)dH(z)\right)\rightarrow_d N(0,\Omega)$$

for some Ω in a complicated form. Therefore proposed test statistic

$$T_n = \int \hat{\beta}_n(z) dd\mathbb{H}_n(z) = \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_n(Z_i)$$

is asymptotically normal with mean zero and variance Ω/n in a complicated form under H_0: $\beta(z) = 0$ for all z. The variance can be estimated through the bootstrap method with the validity justified by the asymptotic normality just proved.

Figure 1 for simulation of spline-based semiparametric model.

Some reserved simulation results

Here we just kept the following simulation results under sample size 100 and 400.

S2. Linear regression functions $\beta(Z) = 0.5 * Z$

S3. Nonlinear regression functions $\beta(Z) = 0.5{*}Beta(Z, 2, 2)$, where $Beta(\cdot)$ is the $Beta$ density function.

S4. Nonlinear regression functions that oscillate at 0: $\beta(Z) = 1.5 \sin(2\pi Z) I(Z \leq 0.5)$ where $I(\cdot)$ is the indicator function

References

De Boor, C. (2001). *A Practical Guide to Splines*. Springer-Verlag, New York.

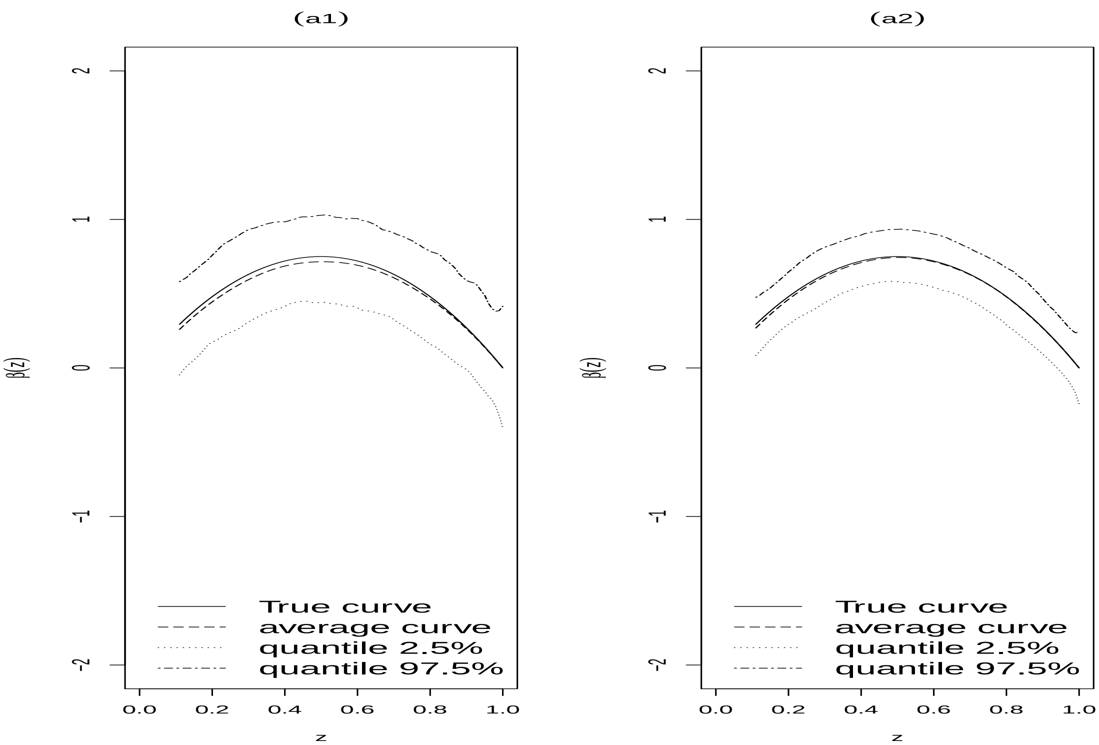

Figure 1: The solid curve is the true regression function $\beta(z)$, the dotted, dashed and dash dotted curves are the pointwise 2.5-quantile, mean and 97.5 quantile of $\hat{\beta}_n(z)$s; (a1)-(a3) are the results of $\beta(Z) = 0.5 * Beta(Z, 2, 2)$ (where $Beta(\cdot)$ is the $Beta$ density function.) under sample size 100 and 400;

Du, P., Jiang, Y. and Wang, Y. (2011). Smoothing spline ANOVA frail ty model for recurrent event data.*Biometrics* **67**, 1330-1339.

Hu, XJ., Lagakos, SW., and Lockhart. RA. (1998). Marginal Analysis of Panel Counts through

Table 1: SP, Sample size; M-C, Monte Carlo; ASE, average of standard errors; SD, standard deviation.

Parameter	SP	True value	Bias	M-C SD	ASE	Probability of rejecting H_0
$\beta(Z) = 0.5 * Z$	100	0·25	0·012	0·218	0·178	0·374
	400		0·001	0·127	0·117	0·581
$\beta(Z) = 0.5 * Beta(Z, 2, 2)$	100	0·5	-0·011	0·216	0·173	0·772
	400		-0·01	0·124	0·113	0·969
$1.5 \sin(2\pi Z) I(Z \leq 0.5)$	100	0·477	-0·080	0·225	0·189	0·559
	400		-0·028	0·116	0·109	0·983

Estimating Functions. *Biometrika* **96**, 445-456.

Hua L, Zhang Y, Tu W. (2014) A spline-based semiparametric sieve likelihood method for over-dispersed panel count data. *Canadian Journal of Statistics-revue Canadienne De Statistique* **42**, 217-245.

Huang, C. Y., Qin, J., and Wang, M. C. (2010). Semiparametric analysis for recurrent event data with time-dependent covariates and informative censoring. *Biometrics* **66**, 39-49.

Lange, K.(1994) An adaptive barrier method for convex programming. *Methods and Applications of Analysis***1**, 392-402.

Lee, EW. and Kim, MY. (1998). The analysis of correlated panel data using a continuous-time Markov model. *Biometrics.* **54**, 1638-1644.

Lu, M., Zhang, Y. and Huang, J. (2007). Estimation of the mean function with panel count data using monotone polynomial splines. *Biometrika* **94**, 705-718.

Lu, M., Zhang, Y. and Huang, J. (2009). Semiparametric estimation methods for panel count data using monotone B-splines. *J. Amer. Statist. Assoc.* **104**, 1060-1070.

Nielsen, J. D. and Dean, C. B. (2008). Clustered mixed nonhomogeneous Poisson process spline models for the analysis of recurrent event panel data. *Biometrics* **64**, 751-761.

Nielsen, J. D. and Dean, C. B. (2008). Adaptive functional mixed NHPP models for the analysis of recurrent event panel data.*Computational Statistics and Data Analysis* **52**, 3670-3685.

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

Shen, X. and Wong, WH. (1994). Convergence rate of sieve estimates. *Ann. Statist.* **22**, 580-615.

Sun, J. and Kalbfleisch, JD. (1995). Estimation of the mean function of point processes based on panel count data. *Statist. Sinica.* **5**, 279-290.

Sun, J. and Zhao, X. (2013). *Statistical Analysis of Panel Count Data*. Springer-Verlag, New York.

Sun, J. and Wei, LJ. (2000). Regression analysis of panel count data with covariate-dependent observation and censoring times. *J. Roy.Statist. Soc. B.* **62**, 293-302.

Sun, L., Zhu, L., and Sun, J.(2009). Regression analysis of multivariate recurrent event data with time-varying covariate effects. *J Multivariate Anal.* **100**, 2214-2223.

Thall, PF. and Lachin, JM. (1988) Analysis of recurrent events: Nonparametric methods for random-interval count data. *J. Amer. Statist. Assoc.* **83**, 339-347.

van der Vaart, AW. (2000). *Asymptotic Statistics.* Cambridge university press.

van der Vaart, AW. and Wellner, JA. (2000). *Weak Convergence and Empirical Processes with Applications to Statistics.* Springer-Verlag, New York.

Wellner, JA. and Zhang, Y. (2000). Two estimators of the mean of a counting process with panel count data. *Ann. Statist.* **28**, 779-814.

Wellner, JA. and Zhang, Y. (2007). Two likelihood-based semiparametric estimation methods for panel count data with covariates. *Ann. Statist.* **35**, 2106-2142.

Wu, Y. and Zhang, Y. (2012). Partially monotone tensor spline estimation of the joint distribution function with bivariate current status data. *Ann. Statist.* **40**, 1609-1636.

Yao, W., Barb-Tuana, F.M., Llapur, C.J., Jones, M.H., Tiller, C., Kimmel, R., Kisling, J., Nguyen, E.T., Nguyen, J., Yu, Z. and Kaplan, M.H. (2010). Evaluation of airway reactivity and immune characteristics as risk factors for wheezing early in life. *Journal of Allergy and Clinical Immunology* **126**, 483-488.

Yu, Z., Liu, L., Bravata, DM., Williams, LS., Tepper, RS. (2013). A semiparametric recurrent events model with time-varying coefficients. *Statistics in Medicine.* **32**, 1016-1026.

Zhang, Y. (2002). A semiparametric pseudolikelihood estimation method for panel count data. *Biometrika* **89**, 39-48.

Zhu, L., Zhao, H., Sun, J., Leisenring, W., and Robison, LL. (2015). Regression analysis of mixed recurrent-event and panel-count data with additive rate models. *Biometrics* **71**, 71-79.

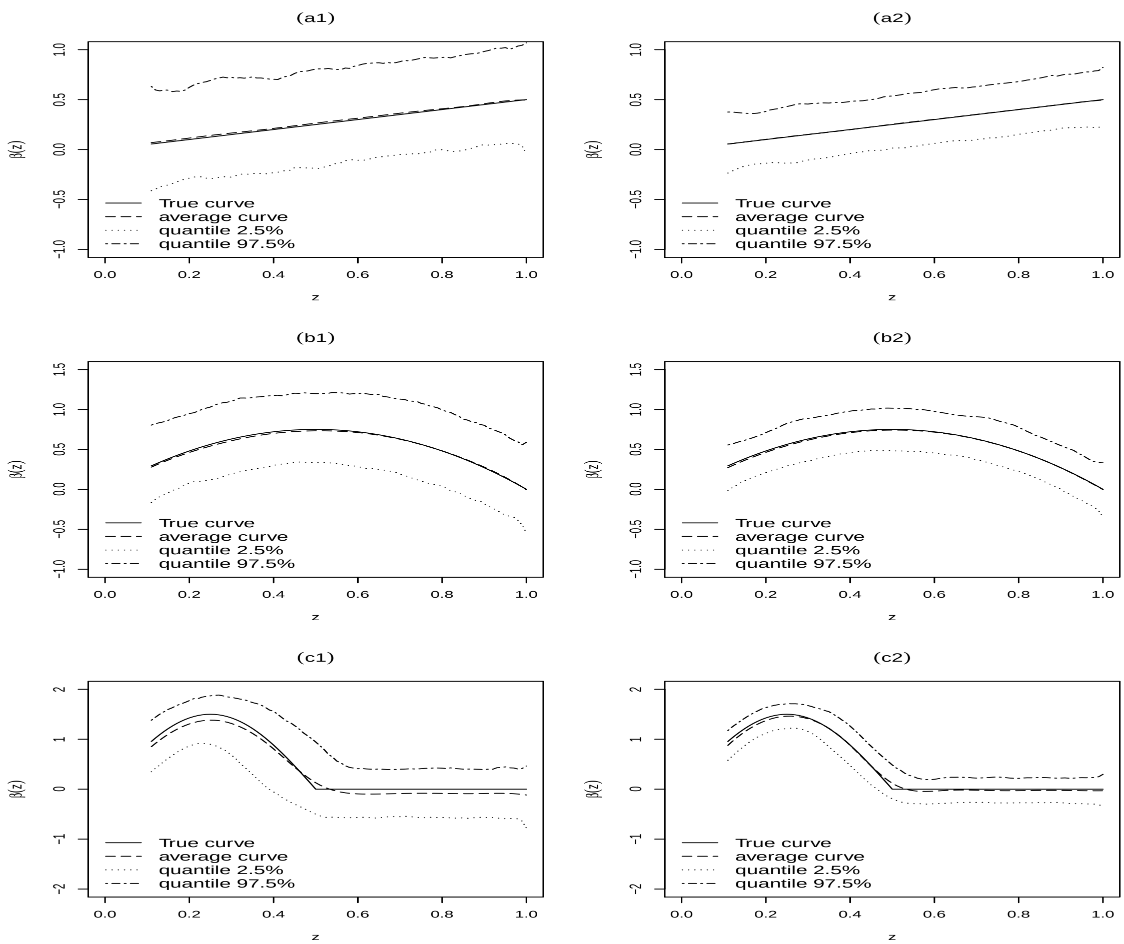

Figure 2: Estimation results for the regression function: The solid curve is the true regression function $\beta(z)$, the dotted, dashed and dash dotted curves are the pointwise 2.5-quantile, mean and 97.5 quantile of $\hat{\beta}_n(z)$s; (a1)-(a2) are the results of $\beta(Z) = 0.5 * Z$ under sample sizes 100 and 400; (b1)-(b2) are the results of $\beta(Z) = 0.5 * Beta(Z, 2, 2)$ under sample sizes 100 and 400. (c1)-(c2) are the results of $\beta(Z) = 1.5 \sin(2\pi Z) I(Z \leq 0.5)$ under sample sizes 100 and 400.