Technical Report for Marginal Distribution Model Checking Tests for Regression

Junyi Dong
Department of Mathematical Sciences, SUNY, Binghamton, NY 13850
and
Qiqing Yu
Department of Mathematical Sciences, SUNY, Binghamton, NY 13850
qyu@math.binghamton.edu

May 30, 2017

1 Proof of Example 2.1 and 2.2

Proof of Example 2.1. (Continued) Assume that the joint distribution of (Y,Z) is uniform on the region $A_1 \cup A_2$, where A_1 is the set bounded by the four straight lines $y=0$, $y=1$, $x-y=0$ and $x-y=-1$, and A_2 is the set bounded by $y=0$, $y=1$, $x=3$ and $x=4$. The family of distributions $\{F_{Y|Z}(\cdot|z):\ z \in (-1,1)\cup(3,4)\}$ does not satisfy the TIPH model and the simply linear regression model. We shall show that the functions S_Y, S_{Y^*} for fitting TIPH model and S_{Y^*} for fitting the linear regression model are all uniquely defined and are different, as well as β. In particular, $Y \sim unif(0,1)$,

$$S_{Y|Z}(y|x) = \begin{cases} 1-\frac{y}{1+x} & \text{if } y \in [0,1+x] \text{ and } x \in (-1,0] \\ 1-\frac{y-x}{1-x} & \text{if } y \in [x,1] \text{ and } x \in (0,1] \\ 1-y & \text{if } y \in [0,11 \text{ and } x \in [3,4], \end{cases} \quad \text{and } f_Z(x) = \begin{cases} \frac{1+x}{2} & \text{if } x \in [-1,0] \\ \frac{1-x}{2} & \text{if } x \in (0,1] \\ \frac{1}{2} & \text{if } x \in [3,4] \end{cases}.$$

To fit the TIPH model, for $y \in [0,1]$, $S_{Y^*|Z}(y|x) = (1-y)^{\exp(\beta x)}$ and $S_{Y^*}(y) = \int (1-y)^{\exp(\beta x)} f_Z(x)dx$. The numerical calculation yields $\beta \approx -0.046$, which uniquely maximizes $\mathscr{L}(\beta) = \beta E(Z_1) - E[\ln E(\mathbf{1}(Y_2 \geq Y_1)\exp(\beta Z_2)|Y_1)]$, where (Y_1,Z_1) and (Y_2,Z_2) are i.i.d. from $F_{Y,Z}$. Moreover, $S_Y \neq S_{Y^*}$, otherwise, it leads to a contradiction: $-1 = S'_Y(1-) = S'_{Y^*}(1-) = 0$, as $-1 = S'_Y(y) = S'_{Y^*}(y) = \int (1-y)^{\exp(\beta x)-1} e^{\beta x}(-1) f_Z(x)dx$, $\forall\ y \in (0,1)$.

To fit the simple linear regression model, $S_{Y^*|Z}(y|x) = \mathbf{1}(y-\beta x \leq 1) - (y-\beta x)\mathbf{1}(0 \leq y-\beta x \leq 1)$ and $S_{Y^*}(y) = \int [\mathbf{1}(y-\beta x \leq 1) - (y-\beta x)\mathbf{1}(0 \leq y-\beta x \leq 1)] f_Z(x)dx$, where $\beta = Cov(Z,Y)/Var(Z) = 2/153$. Moreover, $S_Y \neq S_{Y^*}$, as

$S_{Y^*}(1) = \int [\mathbf{1}(x \geq 0) - (1-\beta x)\mathbf{1}(1/\beta \geq x \geq 0)] f_Z(x)dx \geq \int \beta x \mathbf{1}(x \geq 0) f_Z(x)dx = (2/153)E(Z\mathbf{1}(Z > 0)) > 0 = S_Y(1)$.

Proof of Example 2.2 (continued). If Θ_0 is a simple linear regression, then $S_Y = S_{Y^*}$, $\beta = 1$, and for $x \in \{0,1\}$, $S_{Y|Z}(t|x) = \begin{cases} 1 & \text{if } t-x<0 \\ 1-(t-x) & \text{if } t-x \in [0,1], \end{cases}$ $S_Y(t) = E(S_o(t-Z)) = \begin{cases} 1 & \text{if } t<0 \\ 1-t/2 & \text{if } t \in [0,2), \end{cases}$

$S_o(t) = \begin{cases} 1 & \text{if } t<0 \\ 1-t & \text{if } t \in [0,1). \end{cases}$

If Θ_0 is a TIPH model, then $\beta = -\infty$, as $\mathscr{L}'(\beta) = 0.5e^{\beta}[\beta - \ln[1+e^{\beta}]] < 0$, where $\mathscr{L}(\boldsymbol{\theta}) = E\big[\boldsymbol{\theta}'\mathbf{G}(Y_1)\mathbf{Z}_1 - \ln E[e^{\boldsymbol{\theta}'\mathbf{G}(Y_1)\mathbf{Z}_2}\mathbf{1}(Y_2 \geq Y_1)|Y_1]\big]$. $S_{Y^*|Z}(t|x) = (S_o(t))^{e^{\beta x}} = \begin{cases} 1 & \text{if } t<0 \text{ and } x=0 \\ (1-t) & \text{if } t \in [0,1] \text{ and } x=0 \\ \mathbf{1}(t<1) & \text{if } x=1 \end{cases}$, and $S_{Y^*}(t) = E[(S_o(t))^{e^{\beta Z}}] = \begin{cases} 1 & \text{if } t<0 \\ 1-t/2 & \text{if } t \in [0,1]. \end{cases}$ Thus $S_Y(t) = (1-t/2) > 0 = S_{Y^*}(t)$ if $t \in [1,2)$. That is, $S_Y \neq S_{Y^*}$.

2 Methods to obtain $\hat{S}^*(t|\mathbf{x})$

1. $\Theta_0 \subset \Theta_L$. $\hat{S}^*(t|\mathbf{x}) = (\hat{S}_o(t))^{\exp(\hat{\boldsymbol{\beta}}'\mathbf{G}(t)\mathbf{x})}$, where $\hat{\boldsymbol{\beta}}$ is the SMLE (see (Wong and Yu, 2012)).

2. $\Theta_0 \subset \Theta_{ph} \cap \Theta_c$. $\hat{S}^*(t|\mathbf{x}) = e^{-\int_{s\leq t} \check{h}_o(s)\exp(\hat{\boldsymbol{\beta}}'\mathbf{G}(s)\mathbf{x} ds)}$, where $\hat{\boldsymbol{\beta}}$ is the MPLE. There are some simplified forms of $\hat{S}^*$ for the special cases as follows.

   (a) (Continuous TIPH model) $\hat{S}^*(t|\mathbf{x}) = (\hat{S}_o(t))^{\exp(\hat{\boldsymbol{\beta}}'\mathbf{x})}$.

   (b) (Continuous PWPH model) In the continuous PWPH model, say two cut-points a and b, $\mathbf{G}(t)$ is a 3×3 diagonal matrix with diagonal entries $\mathbf{1}(t<a)$, $\mathbf{1}(a \leq t < b)$, and

$\mathbf{1}(t \geq b)$. Then the estimator of $S^*(t|U,R,V)$, where $\mathbf{Z} = (U,R,V)'$, is

$$\hat{S}^*(t|U,R,V) = \begin{cases} (\hat{S}_o(t))^{\exp(\hat{\beta}_1 U)} & \text{if } t < a \\ (\hat{S}_o(a))^{\exp(\hat{\beta}_1 U)} \frac{(\hat{S}_o(t))^{\exp(\hat{\beta}_2 R)}}{(\hat{S}_o(a))^{\exp(\hat{\beta}_2 R)}} & \text{if } a \leq t < b \\ (\hat{S}_o(a))^{\exp(\hat{\beta}_1 U)} \frac{(\hat{S}_o(b))^{\exp(\hat{\beta}_2 R)}}{(\hat{S}_o(a))^{\exp(\hat{\beta}_2 R)}} \frac{(\hat{S}_o(t))^{\exp(\hat{\beta}_3 V)}}{(\hat{S}_o(b))^{\exp(\hat{\beta}_3 V)}} & \text{if } t \geq b. \end{cases}$$

(c) (Continuous TDPH model) In continuous TDPH model, $G(t) = (t-a)\mathbf{1}(t \geq a)$. Let $a = b_0 < b_1 < ... < b_k$ be the discontinuous points of $\hat{S}_o(t)$ for $t > a$. The estimator of $S^*(t|Z)$ is

$$\hat{S}^*(t|Z) = \begin{cases} \hat{S}_o(t) & \text{if } t < b_1 \\ \hat{S}_o(a) \prod_{i=1}^{j} (\frac{\hat{S}_o(b_i)}{\hat{S}_o(b_{i-1})})^{\exp((b_i - a)\hat{\beta} Z)} & \text{if } b_j \leq t < b_{j+1}. \end{cases}$$

3. $\Theta_0 \subset \Theta_{ph} \cap \Theta_d$. $\hat{S}^*(t|\mathbf{x}) = \prod_{s \leq t}(1 - \check{h}_o(s) \exp(\hat{\boldsymbol{\beta}}'\mathbf{G}(s)\mathbf{x}))$, where $\hat{h}_o(t) = \frac{\hat{S}_o(t-) - \hat{S}_o(t)}{\hat{S}_o(t-)}$, and $\hat{\boldsymbol{\beta}}$ is the MPLE.

4. $\Theta_0 \subset \Theta_{lr}$. $\hat{S}^*(t|\mathbf{x}) = \hat{S}_o(t - \hat{\boldsymbol{\beta}}'\mathbf{x})$, where $\hat{\boldsymbol{\beta}} = (\overline{\mathbf{Z}\mathbf{Z}'} - \overline{\mathbf{Z}}(\overline{\mathbf{Z}})')^{-1}(\overline{\mathbf{Z}Y} - \overline{\mathbf{Z}}(\overline{Y}))$ if there is no tie in Y_i's, otherwise, $\hat{\boldsymbol{\beta}}$ is given by the SMLE (see (Yu and Wong, 2005)), which satisfies $\lim_{n\to\infty} P(\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}) = 1$ if $F_{Y,\mathbf{Z}} \in \Theta_{lr} \cap \Theta_d$.

5. $\Theta_0 \subset \Theta_{apl}$. $\hat{S}_{Y|\mathbf{Z}}(t|\mathbf{z}) = \hat{S}_o(t - \hat{\boldsymbol{\beta}}'\mathbf{z} - \sum_{j=1}^{q} \hat{f}_j(z_j))$, where $(\beta_1, ..., \beta_q) = (0, ..., 0)$. The estimators $\hat{f}_j(\cdot)$, $1 \leq j \leq q$ can be obtained by $gam()$ function in the R package $mgcv$. Let $\mathbf{U} = (Z_{q+1}, ..., Z_p)$, then $(\hat{\beta}_{q+1}, ..., \hat{\beta}_p)' = (\overline{\mathbf{U}\mathbf{U}'} - \overline{\mathbf{U}}(\overline{\mathbf{U}})')^{-1}[(\overline{\mathbf{U}(Y - \sum_{i=1}^{q} \hat{f}_i(Z_i))}) - (\overline{\mathbf{U}}\overline{Y - \sum_{i=1}^{q} \hat{f}_i(Z_i)})]$.

6. $\Theta_0 \subset \Theta_{gplsi}$. $\hat{S}_{Y^*|\mathbf{Z}} = \hat{S}_o(t - \hat{\boldsymbol{\gamma}}'\mathbf{Z} - \rho(\hat{\boldsymbol{\alpha}}'\mathbf{Z}))$, where $(\gamma_1, ..., \gamma_q, \alpha_{q+1}, \alpha_p) = (0, ..., 0)$ and $\boldsymbol{\beta} = (\alpha_1, ..., \alpha_q, \gamma_{q+1}, ..., \gamma_p)'$. The estimator $\hat{\boldsymbol{\beta}}$ and $\hat{\rho}$ can be obtained by the procedure proposed by Carroll (1997).

3 Simulation

The next table summarizes the simulation cases. The PH model is defined by

$$h(t|\mathbf{z}) = h_o(t)\exp(\boldsymbol{\beta}^T\mathbf{G}(t)\mathbf{z}) \tag{1}$$

where $\mathbf{G}(t)$ is a $p \times p$ diagonal matrix with diagonal elements $g_j(t)$, $j = 1, ..., p$. If $\mathbf{G}(t)$ is the identity matrix, then the PH model is called time-independent PH (TIPH) model. If $g(t) = \mathbf{1}(a < t < b)$, where a and b are cut-points, then the PH model is called the piece-wise PH (PWPH) model. For other $g(t) \neq 1$, we called the PH model a time-dependent PH (TDPH) model.

	True Data	Θ_0	are tests valid ?	reject H_0?	error	cases
1	TIPH	TIPH	both MD and residual	no	type I	cases 1, 8 (RC)
2	PWPH	TIPH	both MD and residual	yes	type II	cases 5(RC), 10, 12
3	TDPH	TIPH	both MD and residual	yes	type II	cases 2 (RC), 7
4	PWPH	PWPH	both MD and residual	no	type I	case 9
5	TIPH	TIPH	only MD method	yes	type II	case 3
6	TDPH	TIPH	only MD method	yes	type II	cases 6, 11
7	TDPH	PWPH	only MD method	yes	type II	case 4
8	non-PH	TIPH	only MD method	yes	type II	example 2.1

CASE 1 Complete data. (Y_1, Z_1), ..., $Y_n, Z_n)$ are from $h_{Y|Z}(t|z) = h_0(t)\exp(z)$, where $h_0(t) = \mathbf{1}(t > 0)$, and $Z \sim N(0,1)$. Let H_0: $h(t|z) = h_0(t)\exp(\beta z)$. The residual method assumes $h(t|z) = h_0(t)\exp(\beta z + \theta z t)$ and tests H_0^r: $\theta = 0$. v.s. H_1^r: $\theta \neq 0$. Both H_0 and H_0^r are correct. We compare the probability of type I error $P(H_1|H_0)$ of these two methods. The MD plots suggest that the model fits even for sample size $n = 50$. All tests achieve the nominal level of the tests.

Case 1 $P(H_1\|H_0)$	T_1	T_2	T_3	T_4	T_5	T_6	residual (1)	residual (2)
n=50	0.036	0.035	0.024	0.029	0.027	0.034	0.056	0.004
n= 100	0.040	0.040	0.036	0.042	0.036	0.046	0.054	0.004
n = 200	0.039	0.041	0.025	0.029	0.029	0.032	0.058	0.004

CASE 2 RC data. $h_{Y|Z}(t|z) = h_0(t)\exp(\beta z + \theta z^2 t)$, where $P(Z=-1) = P(Z=-2) = 0.5$, $h_0(t) = \mathbf{1}(t>0)$, $\beta = -0.5$, $\theta = 1$, and $C \equiv 0.7$. H_0: $h(t|z) = h_0(t)\exp(\beta z)$. The residual method assumes $h(t|z) = h_0(t)\exp(\beta z + \theta z^2 t)$ and tests H_0^r: $\theta = 0$ v.s. H_1^r: $\theta \neq 0$. Both methods assume the correct parameter space and should reject H_0. We compare $P(H_0|H_1)$ of these two methods. The MD plots clearly reject H_0 even if $n = 50$. The MD tests can detect the incorrect model when $n \geq 200$, except T_1 and T_2. The residual test almost never rejects the wrong model for $n \leq 200$. The MD tests are more powerful than the residual test except T_1 and T_2..

Case 8 $P(H_0\|H_1)$	T_1	T_2	T_3	T_4	T_5	T_6	residual (2)
n = 50	0.996	0.998	0.826	0.918	0.830	0.920	0.988
n=100	1	1	0.682	0.786	0.680	0.778	0.936
n= 200	1	1	0.406	0.514	0.406	0.506	0.798

CASE 3 Complete data. $h_{Y|Z}(t|z) = h_0(t)\exp(\beta z^2)$. $h_0(t) = \mathbf{1}(t>0)$, $\beta = 1$ and $Z \sim unif(-3,3)$. H_0: $h(t|z) = h_0(t)\exp(\beta z)$. The residual method assumes $h(t|z) = h_0(t)\exp(\beta z + \theta z t)$ and tests H_0^r: $\theta = 0$ v.s. H_1^r: $\theta \neq 0$. Both methods should reject H_0. Here we compute the probability of type II error for MD test and that of pseudo type II error for residual method. The MD plots clearly reject H_0 if $n \geq 100$. The MD tests perform similarly. The residual test does not reject the wrong model w.p.0.95.

	type II error						pseudo type II error	
Case 10 $P(H_0\|H_1)$	T_1	T_2	T_3	T_4	T_5	T_6	residual (1)	residual (2)
n= 100	0.004	0.004	0.054	0.008	0.064	0.008	0.932	1
n= 150	0	0	0.002	0	0.002	0	0.958	1
n = 200	0	0	0	0	0	0	0.952	1

CASE 4 Complete data. $h_{Y|Z}(t|z) = h_0(t)\exp(2z\mathbf{1}(t \geq 0.2) + 2zt\mathbf{1}(t \geq 0.6))$, where $h_0(t) = \mathbf{1}(t > 0)$ and $Z \sim unif(0,1)$. $H_0 : h(t|z) = h_0(t)\exp(\beta z\mathbf{1}(t \geq 0.1))$. The residual method assumes $h(t|z) = h_0(t)\exp((\beta z + \theta zt)\mathbf{1}(t \geq 0.1))$. Both methods should reject H_0. Here we compute $P(H_0|H_1)$ for MD test and the probability of pseudo type II error for residual test. The MD plots clearly reject H_0 if $n \geq 50$. The MD tests perform similarly, except T_1 and T_2. The residual tests are not as powerful as the MD tests T_3, T_4, T_5 and T_6.

	type II error						pseudo type II error
Case 12 $P(H_0\|H_1)$	T_1	T_2	T_3	T_4	T_5	T_6	residual (1)
n=50	1	1	0.59	0.49	0.52	0.57	0.694
n= 100	1	1	0.14	0.10	0.12	0.08	0.484
n = 200	1	1	0.07	0.06	0.06	0.04	0.164

CASE 5 RC data. $h_{Y|Z}(t|z) = h_0(t)\exp(z\mathbf{1}(t \geq 1))$, where $h_0(t) = \mathbf{1}(t > 0)$, $C \sim unif(0,2)$, and $Z \sim bin(2,0.5)$. H_0: $h(t|z) = h_0(t)\exp(\beta z)$. The residual method assumes $h(t|z) = h_0(t)\exp(\beta z + \theta z\mathbf{1}(t \geq 1))$ and tests H_0^r: $\theta = 0$ v.s. H_1^r: $\theta \neq 0$. Both methods assume the correct parameter space and should reject H_0. We compare $P(H_0|H_1)$ of these two methods. Here, we only apply existing code, $cox.zph$, for residual method. The MD plots are unclear even if $n = 200$. The MD tests can detect the wrong model if $n = 400$ except T_1 and T_2. The residual tests are more powerful.

Case 5 $P(H_0\|H_1)$	T_1	T_2	T_3	T_4	T_5	T_6	residual (2)
n=100	0.84	0.86	0.64	0.71	0.64	0.69	0.594
n= 200	0.82	0.84	0.60	0.58	0.58	0.59	0.249
n = 400	1	1	0.1	0.17	0.13	0.30	0.029

CASE 6 Complete data. $h_{Y|Z}(t|z) = h_0(t)\exp(\beta z + \theta z^2 t)$, where $h_0(t) = \mathbf{1}(t > 0)$, $\beta = -0.5$, $\theta = 1$, and $P(Z = -1) = P(Z = -2) = 0.5$. H_0: $h(t|z) = h_0(t)\exp(\beta z)$. The residual method assumes $h(t|z) = h_0(t)\exp(\beta z + \theta z t)$ and tests H_0^r: $\theta = 0$ v.s. H_1^r: $\theta \neq 0$. H_0 is false. We compare $P(H_0|H_1)$ of these two methods. The MD plots clearly reject H_0 if $n \geq 100$. The MD tests and residual test (1) perform similarly, except for T_1, T_2 and residual test (2).

Case 6 $P(H_0\|H_1)$	T_1	T_2	T_3	T_4	T_5	T_6	residual (1)	residual (2)
n=50	0.996	0.996	0.774	0.872	0.750	0.856	0.764	0.998
n= 100	1	1	0.628	0.730	0.606	0.698	0.546	0.984
n = 200	1	1	0.362	0.428	0.338	0.404	0.226	0.970

CASE 7 Complete data. $h_{Y|Z}(t|z) = h_0(t)\exp(\beta z + \theta z t)$, where $h_0(t) = \mathbf{1}(t > 0)$, $\beta = -1$, $\theta = 5$ and $Z \sim unif(0,4)$. H_0: $h(t|z) = h_0(t)\exp(\beta z)$. The residual method assumes $h(t|z) = h_0(t)\exp(\beta z + \theta z t)$ and tests H_0^r: $\theta = 0$ v.s. H_1^r: $\theta \neq 0$. Both methods assume the correct parameter space and should reject H_0. We compare $P(H_0|H_1)$ of these two methods. The MD plots clearly reject H_0 even if $n = 50$. The MD tests perform similarly, except T_1 and T_2. The residual tests are not as powerful as the MD tests except T_1 and T_2.

Case 7 $P(H_0\|H_1)$	T_1	T_2	T_3	T_4	T_5	T_6	residual (2)
n=50	1.000	1.000	0.428	0.330	0.528	0.374	0.526
n= 100	1.000	1.000	0.116	0.104	0.158	0.138	0.162

CASE 8 RC data. $h_{Y|Z}(t|z) = h_0(t)\exp(z)$, where $h_0(t) = \mathbf{1}(t > 0)$, $Z \sim pois(1)$ and the censoring variable $C \sim unif(0,2)$. Let H_0: $h(t|z) = h_0(t)\exp(\beta z)$. The residual method assumes $h(t|z) = h_0(t)\exp(\beta z + \theta z \log(t))$ and tests H_0^r: $\theta = 0$ v.s. H_1^r: $\theta \neq 0$. Both H_0 and H_0^r are correct. We compare $(H_1|H_0)$ of these two methods. Here, we only apply existing code, *cox.zph*, for the residual method. The MD plots suggest that the model fits even for sample size $n = 50$. All tests achieve the nominal level of the tests.

Case 2 $P(H_1\|H_0)$	T_1	T_2	T_3	T_4	T_5	T_6	residual (2)
n=100	0.04	0.04	0.04	0.05	0.04	0.03	0.071
n= 200	0.01	0.04	0.03	0.06	0.07	0.06	0.074
n = 400	0.02	0.04	0.03	0	0.07	0	0.069

CASE 9 Complete data. $h_{Y|Z}(t|z) = h_0(t)\exp(\beta z\mathbf{1}(t \geq 1))$. $h_0(t) = \mathbf{1}(t > 0)$, $\beta = 1$ and $Z \sim N(0,1)$. H_0: $h(t|z) = h_0(t)\exp(\beta z\mathbf{1}(t \geq 1))$. The residual method assumes $h(t|z) = h_0(t)\exp(\beta z\mathbf{1}(t \geq 1) + \theta z t^2 \mathbf{1}(t \geq 1))$, and tests H_0^r: $\theta = 0$ v.s. H_1^r: $\theta \neq 0$. Both methods assume correct underlying model and should not reject H_0. We present $P(H_1|H_0)$ for both methods. The MD plots suggest that the model fits even for sample size $n = 50$. All tests achieve the nominal level of the tests.

Case 9 $P(H_1\|H_0)$	T_1	T_2	T_3	T_4	T_5	T_6	residual (1)
n=50	0.110	0.108	0.048	0.048	0.112	0.106	0.118
n= 100	0.002	0.002	0.002	0	0.008	0.006	0.092
n = 200	0.006	0.006	0	0	0.002	0.002	0.064

CASE 10 Complete data. $h_{Y|Z}(t|z) = h_0(t)\exp(\beta z + \theta z\mathbf{1}(t \geq 2))$, where $h_0(t) = \mathbf{1}(t > 0)$, $\beta = -2$, $\theta = 2$ and $Z \sim N(0,1)$. H_0: $h(t|z) = h_0(t)\exp(\beta z)$. H_0^r: $\theta = 0$ v.s. H_1^r: $\theta \neq 0$. Both methods assume the correct parameter space and should reject H_0. We compare the probability of type II error

$P(H_0|H_1)$ of these two methods. The MD plots clearly reject H_0 if $n \geq 100$, and unclear if $n = 50$. The MD tests perform similarly. The residual tests are more powerful.

Case 3 $P(H_0\|H_1)$	T_1	T_2	T_3	T_4	T_5	T_6	residual (1)	residual (2)
n=50	0.344	0.370	0.797	0.624	0.801	0.622	0.098	0.208
n= 100	0.080	0.085	0.508	0.252	0.505	0.265	0.010	0.018
n = 200	0.013	0.012	0.150	0.049	0.159	0.070	0	0

CASE 11 Complete data. $h_{Y|Z}(t|z) = h_0(t)\exp(\beta z^2 + \theta zt)$. $h_0(t) = \mathbf{1}(t > 0)$, let $\beta = 5$, $\theta = 1$ and $Z \sim N(0,1)$. H_0: $h(t|z) = h_0(t)\exp(\beta z)$. The residual method assumes $h(t|z) = h_0(t)\exp(\beta z + \theta zt)$ and tests H_0^r: $\theta = 0$ v.s. H_1^r: $\theta \neq 0$. Since one should reject H_0, we compute $P(H_0|H_1)$ for MD test. The MD plots clearly reject H_0 if $n \geq 50$. The MD tests perform similarly. The residual test does not reject the wrong model with a probability ≥ 0.77.

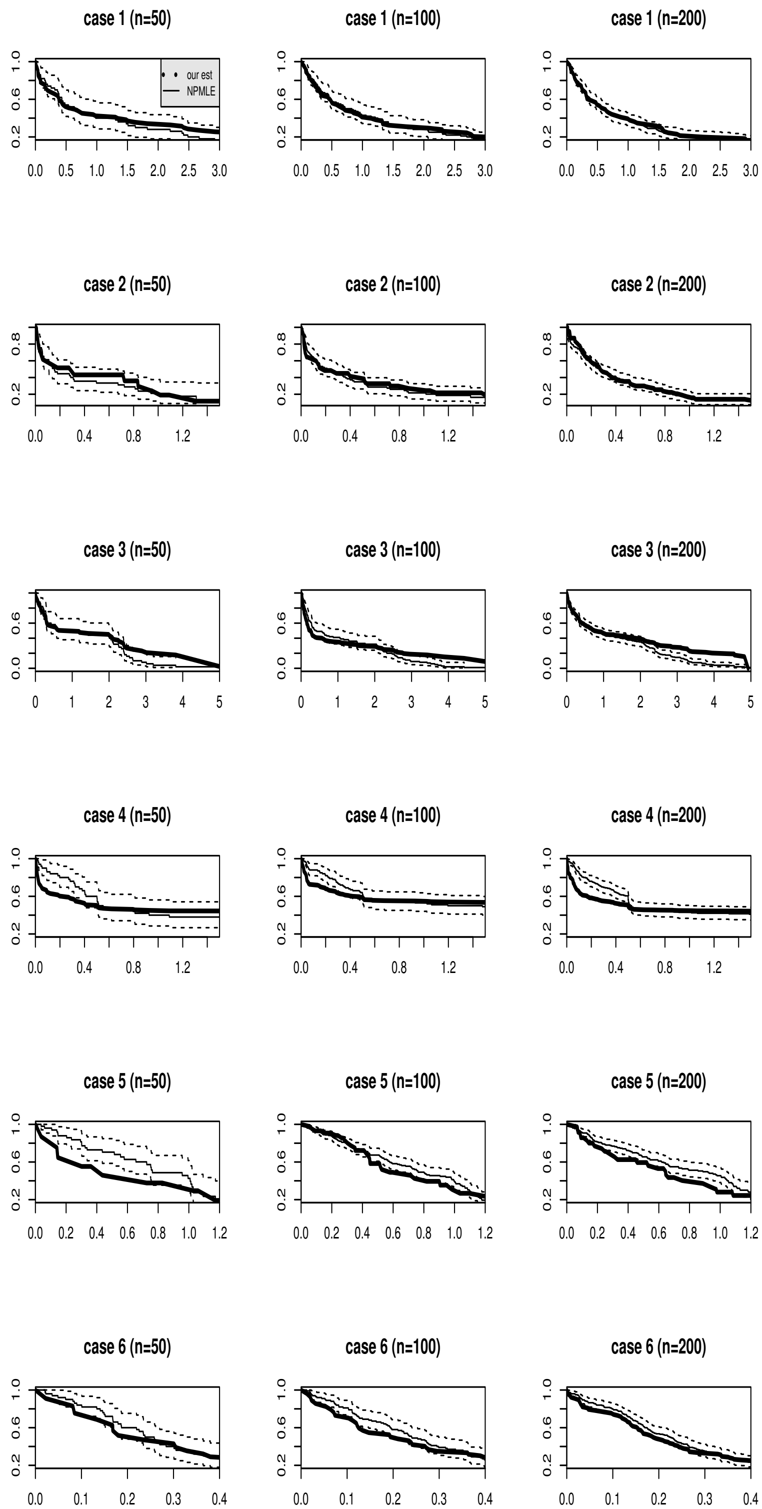
	type II error						pseudo type II error	
Case 11 $P(H_0\|H_1)$	T_1	T_2	T_3	T_4	T_5	T_6	residual (1)	residual (2)
n=50	0.608	0.604	0.850	0.846	0.842	0.838	0.906	1
n= 100	0	0	0.092	0.086	0.106	0.100	0.903	1
n = 200	0	0	0.002	0.002	0.002	0.002	0.774	1

CASE 12 Complete data. $h_{Y|Z}(t|z) = h_0(t)\exp(\beta z + \theta z\mathbf{1}(t \geq 0.5))$, where $h_0(t) = \mathbf{1}(t > 0)$, $\beta = 1$, $\theta = 5$, $Z \sim unif(-2,2)$. H_0: $h(t|z) = h_0(t)\exp(\beta z)$. H_0^r: $\theta = 0$ v.s. H_1^r: $\theta \neq 0$. Both methods assume the correct parameter space and should reject H_0. We compare $P(H_0|H_1)$ of these two methods. The MD plots are unclear even if $n = 200$. The MD tests can detect the wrong model for a large n, except T_1 and T_2. The residual tests are more powerful.

Case 4 $P(H_0\|H_1)$	T_1	T_2	T_3	T_4	T_5	T_6	residual (1)	residaul (2)
n=50	1	1	0.752	0.833	0.815	0.863	0	0
n= 100	1	1	0.169	0.319	0.391	0.467	0	0
n = 200	1	1	0.009	0.021	0.108	0.110	0	0

Example 2.1 (continued). Complete data are generated under the assumptions in Example 2.1. H_0: $h(t|z) = h_0(t)\exp(\beta z)$ v.s. H_1: $h(t|z) \neq h_0(t)\exp(\beta z)$. And residual method assumes $h(t|z) = h_0(t)\exp(\beta z + \theta z t)$ and tests H_0^r: $\theta = 0$ v.s. H_1^r: $\theta \neq 0$. The MD tests perform quite well even when $n = 100$, except for T_1 and T_2, but the residual tests make mistake most of the time and there is no tendency that $P(H_0|H_1)$ goes down as n becomes large.

	type II error						pseudo type II error	
PH model	T_1	T_2	T_3	T_4	T_5	T_6	residual (1)	residual (2)
n= 50	0.982	0.986	0.740	0.774	0.732	0.762	0.900	0.908
n=100	1.000	1.000	0.198	0.252	0.174	0.232	0.912	0.92
n=200	1.000	1.000	0	0	0	0	0.920	0.934


case 1 (n=50)
case 1 (n=100)
case 1 (n=200)
our est
NPMLE
case 2 (n=50)
case 2 (n=100)
case 2 (n=200)
case 3 (n=50)
case 3 (n=100)
case 3 (n=200)
case 4 (n=50)
case 4 (n=100)
case 4 (n=200)
case 5 (n=50)
case 5 (n=100)
case 5 (n=200)
case 6 (n=50)
case 6 (n=100)
case 6 (n=200)

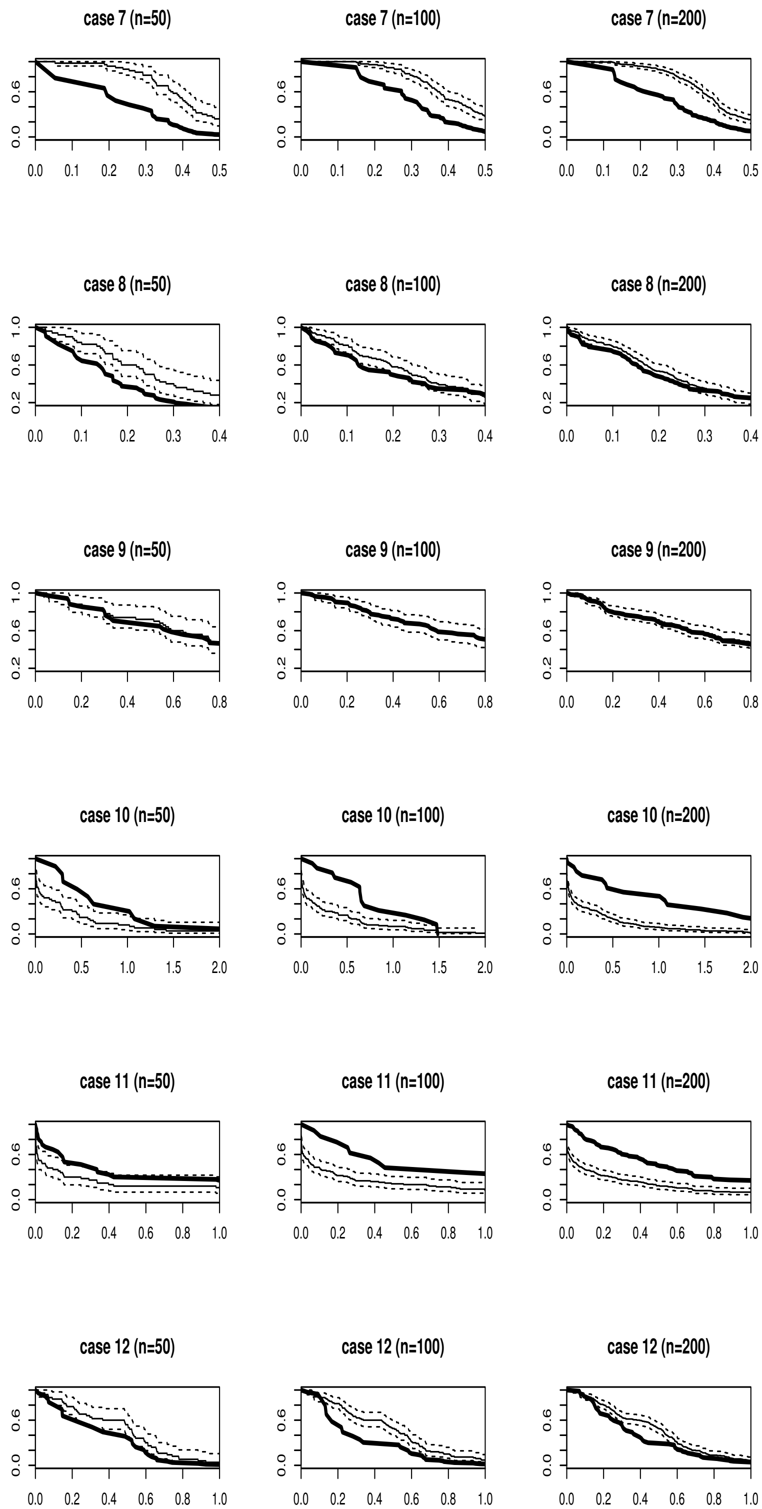


Figure 1. MD plots for Cases 1-12

4 Proof

In this section, we give the proofs for Lemma 2 and 3.

Remark 1. *Let Ω_1 be the event that $\hat{F}_{Y,\mathbf{Z}}(t,\mathbf{z}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(Y_i \le t, \boldsymbol{Z}_i \le \boldsymbol{z}) \to F_{Y,\mathbf{Z}}(t,\boldsymbol{z})$ and $\hat{F}_Y(t) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(Y_i \le t) \to F_Y(t)$ and let $\Omega_{\boldsymbol{z}}$ be the event that $\hat{G}(\mathbf{s}) = \frac{1}{n}\sum_{i=1}^{n}\mathbf{1}(\boldsymbol{Z}_i \le \boldsymbol{s}) \to F_{\mathbf{Z}}(\boldsymbol{s})$, then by the SLLN, $P(\Omega_{\boldsymbol{z}}) = 1$ and $P(\Omega_1) = 1$. Let $\Omega_h = \{\omega \in \Omega, \check{h}_o(t)(\omega) \to h_o(t)\}$ and $\Omega_s = \{\omega \in \Omega : \sup|\hat{S}_o(t)(\omega) - S_o(t)| \to 0\}$ and $\Omega_2 = \Omega_h \cap \Omega_s \cap \Omega_1 \cap \Omega_{\boldsymbol{z}}$, then by Lemma 2 and 3, $P(\Omega_2) = 1$.*

Lemma 4. *Let $(X,\mathscr{F},P)$ be a probability space. Let $\mu_n(t,\omega)$, $t \in \mathbb{R}$ and $\omega \in X$, be a sequence of measure. Let f_n and g_n be measurable functions, $\Omega_a = \{\omega \in X : \mu_n(\cdot,\omega) \to \mu(\cdot,\omega)$ set-wisely$\}$, $\Omega_b = \{\omega \in X : f_n(t,\omega) \to f(t,\omega)$ point-wisely in $t\}$, and $\Omega_c = \{\omega \in X : g_n(t,\omega) \to g(t,\omega)$ point-wisely in $t\}$. If $P(\Omega_a) = P(\Omega_b) = P(\Omega_c) = 1$, $|f_n| \le g_n$, and $\int g_n d\mu_n \to \int g d\mu < \infty$ almost surely, then $\int f_n d\mu_n \to \int f d\mu$ almost surely.*

Proof. Let $\Omega = \Omega_a \cap \Omega_b \cap \Omega_c$, then $P(\Omega) = 1$. For each $\omega \in \Omega$, $\mu_n(\cdot,\omega) \to \mu(\cdot,\omega)$ set-wisely, $f_n(t,\omega) \to f(t,\omega)$ point-wisely in t, and $f_n(t,\omega) \to f(t,\omega)$ point-wisely in t. Since $|f_n| \le g_n$ and $\int g_n d\mu_n \to \int g d\mu < \infty$ almost surely, by the General Convergence Theorem Royden (1988), $\lim \int f_n(t,\omega) d\mu_n(t,\omega) = \int f(t,\omega) d\mu(t,\omega)$. Since $P(\Omega) = 1$, $\int f_n d\mu_n \to \int f d\mu$ almost surely.

Lemma 5. *Assume that S_o is continuous and $(Y_i, \boldsymbol{Z}_i)$, i=1,...,n, are i.i.d copies from $(Y, \boldsymbol{Z})$ where $\boldsymbol{Z} \in \mathbb{R}^p$. If $\hat{\eta} \to \eta$ almost surely, then $\frac{\sum_{i=1}^{n}\mathbf{1}(Y_i>\hat{\eta},||\mathbf{Z}_i||<c_n)}{\sum_{j=1}^{n}\mathbf{1}(||\mathbf{Z}_j||<c_n)} \to P(Y > \eta|\boldsymbol{Z} = \boldsymbol{0})$ almost surely, where c_n satisfies the conditions in Lemma 1.*

Proof. Notice that $h_n := \frac{\sum_{i=1}^{n}\mathbf{1}(Y_i>\hat{\eta},||\mathbf{Z}_i||<c_n)}{\sum_{j=1}^{n}\mathbf{1}(||\mathbf{Z}_j||<c_n)} = \frac{\sum_{i=1}^{n}[\mathbf{1}(Y_i>\hat{\eta}\ge\eta)+\mathbf{1}(Y_i>\eta>\hat{\eta})+\mathbf{1}(\eta>Y_i>\hat{\eta})]\mathbf{1}(||\mathbf{Z}_i||<c_n)}{\sum_{j=1}^{n}\mathbf{1}(||\mathbf{Z}_j||<c_n)}$. Then

$$\frac{\sum_{i=1}^{n}\mathbf{1}(Y_i>\eta,||\mathbf{Z}_i||<c_n)}{\sum_{j=1}^{n}\mathbf{1}(||\mathbf{Z}_j||<c_n)} - \frac{\sum_{i=1}^{n}\mathbf{1}(\hat{\eta}\ge Y_i>\eta,||\mathbf{Z}_i||<c_n)}{\sum_{j=1}^{n}\mathbf{1}(||\mathbf{Z}_j||<c_n)} \le h_n \le \frac{\sum_{i=1}^{n}\mathbf{1}(Y_i>\eta,||\mathbf{Z}_i||<c_n)}{\sum_{j=1}^{n}\mathbf{1}(||\mathbf{Z}_j||<c_n)} + \frac{\sum_{i=1}^{n}\mathbf{1}(\eta\ge Y_i>\hat{\eta},||\mathbf{Z}_i||<c_n)}{\sum_{j=1}^{n}\mathbf{1}(||\mathbf{Z}_j||<c_n)}.$$

The first terms of the upper bound and the lower bound converge almost surely to $P(Y > \eta|\mathbf{Z} = \mathbf{0})$ by Stute (1986). Let $\epsilon > 0$, when n is large enough, since $\hat{\eta} \to \eta$ almost surely, then $|\eta - \hat{\eta}| < \epsilon$ almost surely. The second term in the upper bound $\frac{\sum_{i=1}^n \mathbf{1}(\eta \ge Y_i > \hat{\eta}, ||\mathbf{Z}_i|| < c_n)}{\sum_{j=1}^n \mathbf{1}(||\mathbf{Z}_j|| < c_n)} \le \frac{\sum_{i=1}^n \mathbf{1}(\eta \ge Y_i > \eta - \epsilon, ||\mathbf{Z}_i|| < c_n)}{\sum_{j=1}^n \mathbf{1}(||\mathbf{Z}_j|| < c_n)} \to$ $P(\eta - \epsilon < Y \le \eta | \mathbf{Z} = \mathbf{0})$ $(= S_o(\eta - \epsilon) - S_o(\eta))$ almost surely by Stute (1986). Since it is true for any ϵ and S_o is continuous, the second term in the upper bound converges almost surely to 0. Similarly, it can be shown that the second term in the lower bound converges almost surely to 0. By the Squeeze Theorem, $\frac{\sum_{i=1}^n \mathbf{1}(Y_i > \hat{\eta}, ||\mathbf{Z}_i|| < c_n)}{\sum_{j=1}^n \mathbf{1}(||\mathbf{Z}_j|| < c_n)} \to P(Y > \eta | \mathbf{Z} = \mathbf{0})$ almost surely.

Lemma 6. *If $f_n(x) \to f(x)$ on [a,b], $f_n''(x)$ exists and are uniformly bounded on [a,b] and $h_n \to 0$, then $\frac{f_n(x+h_n) - f(x)}{h_n} \to f'(x)$ uniformly on [a,b].*

Proof. If $f_n(x) \to f(x)$ on [a,b], $f_n''(x)$ exists and are uniformly bounded on [a,b], then by Corollary D of Theorem 3 in Frink (1935), $\lim_{n\to\infty} f_n'(x) \to f'(x)$ on [a,b] uniformly. Also, by Lemma in Frink (1935), $\frac{f_n(x+h_n) - f_n(x)}{h_n} \to f'(x)$ uniformly on [a,b].

A.3. Proof of Lemma 2. Under the assumptions that all expectations exist and that $\Theta_0 \subset \Theta_{lr}$, by the SLLN, $\overline{\mathbf{Z}\mathbf{Z}^T} - \overline{\mathbf{Z}}\,\overline{\mathbf{Z}}^T \to E[\mathbf{Z}\mathbf{Z}^T] - E[\mathbf{Z}]E[\mathbf{Z}^T] = \Sigma_{\mathbf{Z}}$ almost surely and $\overline{\mathbf{Z}Y} - \overline{\mathbf{Z}}\,\overline{Y} \to E[\mathbf{Z}Y] - E[\mathbf{Z}]E[Y]$ almost surely, then $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}$ almost surely. Thus statement (a) holds.

Now assume $F_{Y,\mathbf{Z}} \in \Theta_{lr}$ and $\epsilon \perp \mathbf{Z}$, $\hat{S}^*(t; \boldsymbol{\beta}|\mathbf{x}) = \hat{S}_o(t - \boldsymbol{\beta}^T\mathbf{x}) \to S_o(t - \boldsymbol{\beta}^T\mathbf{x})$ almost surely and $S_{Y^*|\mathbf{Z}}(t; \boldsymbol{\beta}|\mathbf{x}) = S_o(t - \boldsymbol{\beta}^T\mathbf{x})$. Thus statement (b) holds. Moreover, $\hat{S}^*(t; \hat{\boldsymbol{\beta}}|\mathbf{x}) = \hat{S}_o(t - \hat{\boldsymbol{\beta}}^T\mathbf{x}) = \frac{\sum_{i=1}^n \mathbf{1}(Y_i > t - \hat{\boldsymbol{\beta}}^T\mathbf{x}, ||\mathbf{Z}_i|| < c_n)/n}{\sum_{i=1}^n \mathbf{1}(||\mathbf{Z}_i|| < c_n)/n}$. Let $\hat{\eta} = t - \hat{\boldsymbol{\beta}}^T\mathbf{x}$ and $\eta = t - \boldsymbol{\beta}^T\mathbf{x}$. Since $\hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}$ almost surely by Lemma 2(a), $\hat{\eta} \to \eta$ almost surely. Since S_o is continuous, by Lemma 5, $\hat{S}^*(t; \hat{\boldsymbol{\beta}}|\mathbf{x}) \to P(Y > t - \boldsymbol{\beta}^T\mathbf{x} | \mathbf{Z} = \mathbf{0}) = S_o(t - \boldsymbol{\beta}^T\mathbf{x}) = S_{Y^*|\mathbf{Z}}(t; \boldsymbol{\beta}|\mathbf{x})$ almost surely. Thus statement (c) holds.

A.4. Proof of Lemma 3.

Proof of (a). Let $L_n(\boldsymbol{\alpha}) = \prod_{i=1}^n \frac{\exp(\boldsymbol{\alpha}^T \mathbf{G}(Y_i)\mathbf{Z}_i)}{\sum_{k: Y_k \ge Y_i} \exp(\boldsymbol{\alpha}^T \mathbf{G}(Y_i)\mathbf{Z}_k)}$ and $\mathscr{L}_n(\boldsymbol{\alpha}) = \frac{1}{n}\ln L_n(\boldsymbol{\alpha}) + \ln n$, then

$\mathscr{L}_n(\boldsymbol{\alpha}) = \frac{1}{n}\sum_{i=1}^n \boldsymbol{\alpha}^T \mathbf{G}(Y_i)\boldsymbol{Z}_i - \frac{1}{n}\sum_{i=1}^n \ln[\frac{1}{n}\sum_{k=1}^n e^{\boldsymbol{\alpha}^T \mathbf{G}(Y_i)\mathbf{Z}_k} \mathbf{1}(Y_k \ge Y_i)]$. We shall show that its limit is

$\mathscr{L}(\boldsymbol{\alpha}) = E\big[\boldsymbol{\alpha}^T\mathbf{G}(Y_1)\mathbf{Z}_1 - \ln p(\boldsymbol{\alpha}, Y_1)\big]$, where $p(\boldsymbol{\alpha}, Y_1) = E[e^{\boldsymbol{\alpha}^T\mathbf{G}(Y_1)\mathbf{Z}_2}\mathbf{1}(Y_2 \geq Y_1)|Y_1]$,

Let $p_n(\boldsymbol{\alpha}, Y_1) = \frac{1}{n}\sum_{k=1}^n e^{\boldsymbol{\alpha}^T\mathbf{G}(Y_1)\boldsymbol{Z}_k}\mathbf{1}(Y_k \geq Y_1)$, $f_n(\boldsymbol{\alpha}, Y_1) = \ln p_n(\boldsymbol{\alpha}, Y_1)$ and $f(\boldsymbol{\alpha}, Y_1) = \ln p(\boldsymbol{\alpha}, Y_1)$.

By assumption, $||\mathbf{G}(Y_1)\boldsymbol{Z}_1|| \leq M < \infty$. If $\boldsymbol{\alpha}$ is finite, then $|\boldsymbol{\alpha}^T\mathbf{G}(Y_1)\boldsymbol{Z}_1|$ is bounded by some real number K and $e^{-K} \leq \exp(\boldsymbol{\alpha}^T\mathbf{G}(Y_1)\boldsymbol{Z}_1) \leq e^K$. Also $\frac{1}{n}\sum_{i=1}^n \boldsymbol{\alpha}^T\mathbf{G}(Y_i)\mathbf{Z}_i \to E[\boldsymbol{\alpha}^T\mathbf{G}(Y_1)\mathbf{Z}_1]$ almost surely by the SLLN.

$$e^K \geq p_n(\boldsymbol{\alpha}, Y_1) \geq e^{-K}\frac{1}{n}\sum_{k=1}^n \mathbf{1}(Y_k \geq Y_1) \Rightarrow K \geq f_n(\boldsymbol{\alpha}, Y_1) \geq g_n(\boldsymbol{\alpha}, Y_1),$$

where $g_n(\boldsymbol{\alpha}, Y_1) = -K + \ln[\frac{1}{n}\sum_{k=1}^n \mathbf{1}(Y_k \geq Y_1)]$. Notice that

$$\begin{aligned}\int g_n(\boldsymbol{\alpha}, t)d\hat{F}_{Y_1}(t) &= \frac{1}{n}\sum_{j=1}^n \big[\ln[\frac{1}{n}\sum_{k=1}^n \mathbf{1}(Y_k \geq Y_j)] - K\big] \\ &= \frac{n-1}{n}\frac{1}{n-1}\sum_{i=1}^{n-1}\ln(\frac{i}{n(n-1)} + 1 - \frac{i}{n-1}) - K.\end{aligned}$$

Let $h_1(x) = \ln(\frac{x}{n} + 1 - x)$, $0 < x < 1$, then $h_1'(x) < 0$. Using right-endpoints estimation,

$R_{n-1} = \frac{1}{n-1}\sum_{i=1}^{n-1}\ln(\frac{i}{n(n-1)} + 1 - \frac{i}{n-1})$ underestimates $\int_0^1 h_1(x)dx$.

Hence $\int \ln[\frac{\hat{F}_Y(t)}{n} + (1 - \hat{F}_Y(t))]d\hat{F}_Y(t) \leq \int_0^1 \ln(\frac{x}{n} + 1 - x)dx < \infty$ if $n \geq 2$.

Let $h_2(x) = \ln(\frac{x}{n} + 1 - x)$, $\frac{1}{n-1} < x < 1 + \frac{1}{n-1}$, then $h_2'(x) < 0$. Using left-endpoints estimation,

$L_{n-1} = \frac{1}{n-1}\sum_{i=0}^{n-2}\ln(\frac{i+1}{n(n-1)} + 1 - \frac{i+1}{n-1}) = \frac{1}{n-1}\sum_{i=1}^{n-1}\ln(\frac{i}{n(n-1)} + 1 - \frac{i}{n-1})$ overestimates $\int_{\frac{1}{n-1}}^{1+\frac{1}{n-1}} h_2(x)dx$.

Then $\int \ln[\frac{\hat{F}_Y(t)}{n} + (1 - \hat{F}_Y(t))]d\hat{F}_Y(t) \geq \int_{\frac{1}{n-1}}^{1+\frac{1}{n-1}} \ln(\frac{x}{n} + 1 - x)dx > -\infty$ if $n \geq 3$. Also, since

$\int_0^1 \ln(\frac{x}{n} + 1 - x)dx \to \int_0^1 \ln(1 - x)dx = -1$ and $\int_{\frac{1}{n-1}}^{1+\frac{1}{n-1}} \ln(\frac{x}{n} + 1 - x)dx \to \int_0^1 \ln(1 - x)dx = -1$,

$\int \ln[\frac{\hat{F}_Y(t)}{n} + (1 - \hat{F}_Y(t))]d\hat{F}_Y(t) \to \int_0^1 \ln(1 - x)dx = -1$.

Then $\lim \int g_n(\boldsymbol{\alpha}, t)d\hat{F}_{Y_1}(t)$ is finite. Since f_n is bounded by the integrable function g_n and $f_n \to f$ almost surely, by Lemma 4,

$$\int f_n d\hat{F}_Y(t) = \frac{1}{n}\sum_{i=1}^n f_n(Y_i) \to \int f dF_Y(t) = E\big[\ln E[e^{\boldsymbol{\alpha}^T\mathbf{G}(Y_1)\mathbf{Z}_2}\mathbf{1}(Y_2 \geq Y_1)|Y_1]\big]$$

almost surely and $\mathscr{L}_n(\boldsymbol{\alpha}) \to \mathscr{L}(\boldsymbol{\alpha})$ almost surely, for each $\boldsymbol{\alpha} \in \mathbb{R}^p$.

By assumption, $B = \{\boldsymbol{\beta} : \boldsymbol{\beta} = \arg\sup_{\boldsymbol{\alpha}\in\mathbb{R}^p} \mathscr{L}(\boldsymbol{\alpha})\}$ is a singleton set, $\boldsymbol{\beta}_0 = \arg\sup_{\boldsymbol{\alpha}\in\mathbb{R}^p} \mathscr{L}(\boldsymbol{\alpha})$ is uniquely determined. Let $\hat{\boldsymbol{\beta}}_n = \arg\sup_{\boldsymbol{\alpha}\in\mathbb{R}^p} \mathscr{L}_n(\boldsymbol{\alpha})$. Then $\mathscr{L}(\boldsymbol{\beta}_0) = \sup_{\boldsymbol{\alpha}\in\mathbb{R}^p} \mathscr{L}(\boldsymbol{\alpha}) \geq \mathscr{L}(\boldsymbol{\alpha})$ and $\mathscr{L}_n(\hat{\boldsymbol{\beta}}_n) = \sup_{\boldsymbol{\alpha}\in\mathbb{R}^p} \mathscr{L}_n(\boldsymbol{\alpha}) \geq \mathscr{L}_n(\boldsymbol{\alpha})$ for any $\boldsymbol{\alpha} \in \mathbb{R}^p$. Since $\mathscr{L}_n(\hat{\boldsymbol{\beta}}_n) \geq \mathscr{L}_n(\boldsymbol{\beta}_0)$,

$$\liminf_{n\to\infty}\mathscr{L}_n(\hat{\boldsymbol{\beta}}_n) \geq \liminf_{n\to\infty}\mathscr{L}_n(\boldsymbol{\beta}_0) = \mathscr{L}(\boldsymbol{\beta}_0) \text{ almost surely} \tag{2}$$

Let $\boldsymbol{\beta}^*$ be a limiting point of $\hat{\boldsymbol{\beta}}_n$ in the sense that there exists a subsequence of $\hat{\boldsymbol{\beta}}_n(\omega)$, say $\hat{\boldsymbol{\beta}}_{n_l}(\omega)$, such that $\hat{\boldsymbol{\beta}}_{n_l}(\omega) \to \boldsymbol{\beta}^*$ $(= \boldsymbol{\beta}^*(\omega))$. By the assumption in the Lemma 3, $P(\overline{\lim_{n\to\infty}} ||\hat{\boldsymbol{\beta}}_n|| < \infty) = 1$. Let Ω_1 be as defined in Remark 1 and let $\Omega_* = \Omega_1 \cap \{\overline{\lim_{n\to\infty}} ||\hat{\boldsymbol{\beta}}_n|| < \infty\}$. For each $\omega \in \Omega_*$, $\boldsymbol{\beta}^*(\omega)$ is finite. Then

$$\mathscr{L}_n(\hat{\boldsymbol{\beta}}_{n_l}(\omega)) = \frac{1}{n}\sum_{i=1}^{n} \hat{\boldsymbol{\beta}}_{n_l}(\omega)^T \mathbf{G}(Y_i)\boldsymbol{Z}_i - \frac{1}{n}\sum_{i=1}^{n} \ln[\frac{1}{n}\sum_{k=1}^{n} e^{\hat{\boldsymbol{\beta}}_{n_l}(\omega)^T \mathbf{G}(Y_i)\boldsymbol{Z}_k} \mathbf{1}(Y_k \geq Y_i)]$$

Since $\hat{\boldsymbol{\beta}}_{n_l}(\omega) \to \boldsymbol{\beta}^*$ and $\boldsymbol{\beta}^*$ is finite, $\hat{\boldsymbol{\beta}}_{nl}(\omega)$ is bounded. By the similar argument as above, $\mathscr{L}_n(\hat{\boldsymbol{\beta}}_{n_l}(\omega)) \to \mathscr{L}(\boldsymbol{\beta}^*)$ almost surely. Since $\lim \mathscr{L}_n(\hat{\boldsymbol{\beta}}_{n_l}) \geq \liminf_{n\to\infty}\mathscr{L}_n(\hat{\boldsymbol{\beta}}_n) \geq \mathscr{L}(\boldsymbol{\beta}_0)$ almost surely, we have $\mathscr{L}(\boldsymbol{\beta}^*) \geq \mathscr{L}(\boldsymbol{\beta}_0)$ almost surely. Then $\mathscr{L}(\boldsymbol{\beta}^*) = \mathscr{L}(\boldsymbol{\beta}_0)$, that is, $\boldsymbol{\beta}^* = \boldsymbol{\beta}_0$, as B is a singleton set. Since every convergent subsequence of $\hat{\boldsymbol{\beta}}_n$ converges to $\boldsymbol{\beta}_0$, $\hat{\boldsymbol{\beta}}_n(\omega) \to \boldsymbol{\beta}_0$ $\forall\ \omega \in \Omega_*$. That is, $\hat{\boldsymbol{\beta}}_n \to \boldsymbol{\beta}_0$ almost surely.

Proof of (b). Let $\Omega_s = \{\omega \in \Omega : \sup|\hat{S}_o(t)(\omega) - S_o(t)| \to 0\}$ and

$\Omega_{s'} = \{\omega \in \Omega : sup|\check{S}_o(t)(\omega) - S_o(t)| \to 0\}$. Since $\hat{S}_o(t)$ and $\check{S}_o(t)$ has the same asymptotic properties and $P(\Omega_s) = 1$ (Yu and Li, 1994), $P(\Omega_{s'}) = 1$.

Let $\omega \in \Omega_{s'}$. Since h_o is a piece-wise constant baseline hazard function, for each $t \notin \{a_j : j \geq 0\}$, there exist $p, q \geq 1$ such that $a_p < t < a_q$ and $h_o(t)$ is constant on (a_p, a_q). If $h_o(t) = 0$ on (a_p, a_q), then $\check{h}_o(t)(\omega) = 0$ if n is large enough, as there is no observation in (a_p, a_q), $t \in (a_p, a_q)$, and $\eta_n \to 0$. If $\check{h}_o(t)(\omega) > 0$, assume that the sample size n is large enough such that there exist at least two observations $Y_{(j-1)}$ and $Y_{(j)}$, $j \in \{2,...,m\}$ in (a_p, a_q) such that $a_p < Y_{(j-1)} < t \le Y_{(j)} < a_q$. When $j < m$, let $\eta_n \to 0$, then there are three cases:

(i) $a_p < Y_{(j)} - \eta_n \le Y_{(j-1)} < t \le Y_{(j)} < a_q$,

(ii) $a_p < Y_{(j-1)} < Y_{(j)} - \eta_n < t \le Y_{(j)} < a_q$,

(iii) $a_p < Y_{(j-1)} < t < Y_{(j)} - \eta_n < Y_{(j)} < a_q$.

In case (i), notice that $-\ln\check{S}_o(t)(\omega) = \int_{s\le t} \check{h}_o(s)(\omega)ds$, $-\ln\check{S}_o(t)(\omega) + \ln\check{S}_o(Y_{(j-1)})(\omega) = \int_{Y_{(j-1)}<s\le t} \check{h}_o(s)(\omega)ds = h_j(t - Y_{(j-1)})$. Let $h_n = t - Y_{(j-1)} \ge 0$, since $h_n \le \eta_n$ and $\eta_n \to 0$, $h_n \to 0$. Then $\check{h}_o(t)(\omega) = h_j = \frac{-\ln\check{S}_o(t)(\omega)+\ln\check{S}_o(t-h_n)(\omega)}{h_n}$. Let $f_n(t) = -\ln\check{S}_o(t)(\omega) = -\ln\check{S}_o(Y_{j-1})(\omega) + h_j(t - Y_{(j-1)})$ and $f(t) = -\ln S_o(t) = H_o(t)$, then $f_n(t) \to f(t)$ and $f_n''(t) = 0$. By Lemma 6, $\check{h}_o(t)(\omega) \to (-\ln S_o(t))' = h_o(t)$.

In case (ii), $-\ln\check{S}_o(t)(\omega) + \ln\check{S}_o(Y_{(j-1)})(\omega) = \int_{Y_{(j)}-\eta_n}^{t} \check{h}_o(s)(\omega)ds = h_j(t - (Y_{(j)} - \eta_n))$. Let $h_n = t - (Y_{(j)} - \eta_n) \le \eta_n$, then $h_n \to 0$. Notice that $\check{S}_o(Y_{(j-1)})(\omega) = \check{S}_o(Y_{(j)} - \eta_n)(\omega)$, then $\check{h}_o(t)(\omega) = h_j = \frac{-\ln\check{S}_o(t)(\omega)+\ln\check{S}_o(t-h_n)(\omega)}{h_n}$. Let $f_n(t) = -\ln\check{S}_o(t)(\omega) = -\ln\check{S}_o(Y_{(j-1)})(\omega) + h_j(t - (Y_j - \eta_n))$, $f(t) = -\ln S_o(t) = H_o(t)$, then $f_n(t) \to f(t)$, $f_n''(t) = 0$ and by Lemma 6, $\check{h}_o(t)(\omega) \to (-\ln S_o(t))' = h_o(t)$.

In case (iii), we have $\check{S}_o(t)(\omega) = \hat{S}_o(t)$. Since $\hat{h}_o(t)(\omega) \to h_o(t)$ (Hansen, 2004), $\check{h}_o(t)(\omega) = \hat{h}_o(t)(\omega) \to h_o(t)$.

Finally, if $j = m$, that is, $a_p < Y_{(m-1)} < t \le Y_{(m)} < a_q$, one can define $\check{h}_o(t)(\omega) = 0$, then $\check{S}_o(t)(\omega) = \hat{S}_o(t)(\omega)$. By similar argument as in case (iii), $\check{h}_o(t)(\omega) \to h_o(t)$.

Since $\omega \in \Omega_{s'}$ where $P(\Omega_{s'}) = 1$, $\check{h}_o(t) \to h_o(t)$ almost surely for each t which is not a cut-point.

Proof of (c). Let Ω_s be as defined in (b) and let $\Omega_h = \{\omega \in \Omega, \check{h}_o(t)(\omega) \to h_o(t)\}$, then by (b), $P(\Omega_h) = P(\Omega_s) = 1$. Let $\omega \in \Omega_h \cap \Omega_s$. Since $\boldsymbol{\beta}$ is finite and $\mathbf{G}(t)\mathbf{Z}$ is bounded, there exists $0 < \gamma_1 \le \gamma_2 < \infty$ such that $\gamma_1 \le \exp(\boldsymbol{\beta}^T\mathbf{G}(s)\mathbf{x}) \le \gamma_2$. Then, $\gamma_1\check{h}_o(s)(\omega) \le \check{h}_o(s)(\omega)\exp(\boldsymbol{\beta}^T\mathbf{G}(s)\mathbf{x}) \le$

$\gamma_2\check{h}_o(s)(\omega)$. Let $\tau = \inf\{t \mid S_o(t) = 0\}$. If $\tau = \infty$, then for any t, $\exp(-\int_0^t \check{h}_o(s)(\omega)ds) = \check{S}_o(t)(\omega) \to S_o(t) = \exp(-\int_0^t h_o(s)ds) > 0$ and $\int_0^t \check{h}_o(s)(\omega)ds \to \int_0^t h_o(s)ds < \infty$. Since $f_n(s) = \check{h}_o(s)(\omega)\exp(\boldsymbol{\beta}^T\mathbf{G}(s)\mathbf{x}) \to h_o(s)\exp(\boldsymbol{\beta}^T\mathbf{G}(s)\mathbf{x})$, $\int_0^t \check{h}_o(s)(\omega)\exp(\boldsymbol{\beta}^T\mathbf{G}(s)\mathbf{x})ds \to \int_0^t h_o(s)\exp(\boldsymbol{\beta}^T\mathbf{G}(s)\mathbf{x})ds$ by the dominated convergence theorem. Hence

$$\hat{S}^*(t;\boldsymbol{\beta}|\mathbf{x}) = \exp(-\int_0^t \check{h}_o(s)(\omega)\exp(\boldsymbol{\beta}^T\mathbf{G}(s)\mathbf{x})ds) \to \exp(-\int_0^t h_o(s)\exp(\boldsymbol{\beta}^T\mathbf{G}(s)\mathbf{x})ds) = S_{Y^*|\mathbf{Z}}(t;\boldsymbol{\beta}|\mathbf{x}).$$

If $\tau < \infty$, it is sufficient to show that $\hat{S}^*(t;\boldsymbol{\beta}|\mathbf{x}) \to 0 = S_{Y^*|\mathbf{Z}}(t;,\boldsymbol{\beta}|\mathbf{x})$ for $t \geq b$. Notice that $S_o(t) = 0$ and $S_{Y^*|\mathbf{Z}}(t;\boldsymbol{\beta}|\mathbf{x}) = 0$ when $t \geq \tau$. And

$$\exp(-\int_0^t \check{h}_o(s)(\omega)\gamma_1 ds) \geq \exp(-\int_0^t \check{h}_o(s)(\omega)\exp(\boldsymbol{\beta}^T\mathbf{G}(s)\mathbf{x})ds) = \hat{S}^*(t;\boldsymbol{\beta}|\mathbf{x}) \geq \exp(-\int_0^t \check{h}_o(s)(\omega)\gamma_2 ds).$$

Since $\check{S}_o(t)(\omega) \to S_o(t)$, $\exp(-\int_0^t \check{h}_o(s)(\omega)\gamma_1 ds) = [\check{S}_o(t)(\omega)]^{\gamma_1} \to S_o(t)^{\gamma_1} = 0$,

$\exp(-\int_0^t \check{h}_o(s)(\omega)\gamma_2 ds) = [\check{S}_o(t)(\omega)]^{\gamma_2} \to S_o(t)^{\gamma_2} = 0$ and then $\hat{S}^*(t;\boldsymbol{\beta}|\mathbf{x}) \to 0 = S_{Y^*;\boldsymbol{\beta}|\mathbf{Z}}(t|\mathbf{x})$.

Since $P(\Omega_h \cap \Omega_s) = 1$, $\hat{S}^*(t|\boldsymbol{\beta}|\mathbf{x}) \to S_{Y^*|\mathbf{Z}}(t;\boldsymbol{\beta}|\mathbf{x})$ almost surely for each t.

In addition, let $\Omega_b = \{\omega \in \Omega : \hat{\boldsymbol{\beta}} \to \boldsymbol{\beta}\}$, then by Lemma 3(a), $P(\Omega_b) = 1$. Let $\omega \in \Omega_h \cap \Omega_s \cap \Omega_b$. Since $\boldsymbol{\beta}$ is finite when sample size is large enough, $\hat{\boldsymbol{\beta}}$ is bounded. Also, since $\mathbf{G}(s)\mathbf{Z}$ is bounded, there exists $0 < \gamma_1 \leq \gamma_2 < \infty$ such that $\gamma_1 \leq \exp(\hat{\boldsymbol{\beta}}^T\mathbf{G}(s)\mathbf{x}) \leq \gamma_2$. Then, $\gamma_1\check{h}_o(s)(\omega) \leq \check{h}_o(s)(\omega)\exp(\hat{\boldsymbol{\beta}}^T\mathbf{G}(s)\mathbf{x}) \leq \gamma_2\check{h}_o(s)(\omega)$. By the similar argument as above, $\hat{S}^*(t;\hat{\boldsymbol{\beta}}|\mathbf{x}) \to 0 = S_{Y^*|\mathbf{Z}}(t;\boldsymbol{\beta}|\mathbf{x})$. Since $P(\Omega_h \cap \Omega_s) = 1$, $\hat{S}^*(t;\hat{\boldsymbol{\beta}}|\mathbf{x}) \to S_{Y^*|\mathbf{Z}}(t;\boldsymbol{\beta}|\mathbf{x})$ almost surely for each t.

References

R.J. Carroll, J. Fan, I. Gijbels, and M.P. Wand. Generalized partially linear single-index models. *Journal of the American Statistical Association- Theory and Methods*, 92(436):477–489, 1997.

O. Frink. Differentiation of sequences. *Bull. Amer. Math. Soc*, 41(8):553–560, 1935.

B. E. Hansen. Nonparametric conditional density estimation, 2004.

H.L. Royden. *Real Analysis*. Pearson, 1988. ISBN 0024041513.

W. Stute. On almost sure convergence of conditional empirical distribution functions. *The Annals of Probability*, 1986.

G. Y. C. Wong and Q. Yu. Estimation under the lehmann regression model with interval-censored data. *Communications in Statistics - Simulation and Computation*, 41(8):1489–1500, 2012.

Q. Yu and L. Li. On the strong consistency of the product limit estimator. *Sankhya, A*, (56):416–430, 1994.

Q. Yu and G.Y.C. Wong. A modified semi-parametric mle in linear regression analysis with complete data or right-censored data. *Technometrics*, 47(1):34–42, 2005.