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 \(\{F_y|z : z \in (-1, 1) \cup (3, 4)\}\) 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 \(\beta\). In particular,

\[
Y \sim \text{unif}(0, 1),
\]

\[
S_{y|z}(y|x) = \begin{cases} 
1 - \frac{y}{1+x} & \text{if } y \in [0, 1 + x] \text{ and } x \in (-1, 0) \\
1 - \frac{y-x}{1-x} & \text{if } y \in [x, 1) \text{ and } x \in (0, 1) \\
1 - y & \text{if } y \in [0, 11 \text{ and } x \in [3, 4],}
\end{cases}
\]

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 \(\mathcal{L}(\beta) = \beta E(Z_1) - E[\ln E(1(Y_2 \geq Y_1) \exp(\beta Z_2)|Y_1)]\), 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, \(S_{y^*|z}(y|x) = 1(y - \beta x \leq 1) - (y - \beta x) 1(0 \leq y - \beta x \leq 1)\) and \(S_{y^*}(y) = \int [1(y - \beta x \leq 1) - (y - \beta x) 1(0 \leq y - \beta x \leq 1)] f_z(x) \, dx\), where \(\beta = Cov(Z, Y) / Var(Z) = 2/153\). Moreover, \(S_y \neq S_{y^*}\), as \(S_{y^*}(1) = \int [1(x \geq 0) - (1 - \beta x) 1(1/\beta \geq x \geq 0)] f_z(x) \, dx \geq \int \beta x 1(x \geq 0) f_z(x) \, dx = (2/153) E(Z 1(Z \geq 0)) > 0 = S_y(1)\).

Proof of Example 2.2 (continued). If \(\Theta_0\) 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 $\Theta_0$ is a TIPH model, then $\beta = -\infty$, as $L'(\beta) = 0.5 e^\beta - \ln[1 + e^\beta] < 0$, where $L(\theta) = E[\theta'G(Y)Z - \ln E[e^{\theta'G(Y)Z}1(Y_2 \geq Y_1)|Y_1]]$. $S_{Y|Z}(t|x) = (S_o(t))^e^{\beta x}$

$E[(S_o(t))^e^{\beta z}] = \begin{cases} 1 & \text{if } t < 0 \\ 1 - t/2 & \text{if } t \in [0,1). \end{cases}$ 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 $\hat{S}^* (t|x)$

1. $\Theta_0 \subset \Theta_L$, $\hat{S}^* (t|x) = (\hat{S}_o(t))^{exp(\hat{\beta}G(t)x)}$, where $\hat{\beta}$ is the SMLE (see [Wong and Yu, 2012]).

2. $\Theta_0 \subset \Theta_{ph} \cap \Theta_c$. $\hat{S}^* (t|x) = e^{-\int_{s=t}^{\infty} \hat{h}_o(s) exp(\hat{\beta}G(s)x(s)) ds}$, where $\hat{\beta}$ is the MPLE. There are some simplified forms of $\hat{S}^*$ for the special cases as follows.

(a) (Continuous TIPH model) $\hat{S}^* (t|x) = (\hat{S}_o(t))^{exp(\hat{\beta}x)}$.

(b) (Continuous PWPH model) In the continuous PWPH model, say two cut-points $a$ and $b$, $G(t)$ is a $3 \times 3$ diagonal matrix with diagonal entries $1(t < a), 1(a \leq t < b),$ and
\[ \textbf{1}(t \geq b). \text{ Then the estimator of } S^*(t|U, R, V), \text{ where } Z = (U, R, V)', \text{ is} \]

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

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

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

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

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

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

6. \( \Theta_0 \subset \Theta_{gplsi} \). \( \hat{S}_{Y*|Z} = \hat{S}_o(t - \hat{\gamma}' Z - \rho(\hat{\alpha}' Z)), \) where \( (\gamma_1, ..., \gamma_q, \alpha_{q+1}, \alpha_p) = (0, ..., 0) \) and \( \beta = (\alpha_1, ..., \alpha_q, \gamma_{q+1}, ..., \gamma_p)' \). The estimator \( \hat{\beta} \) 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

\[ h(t|z) = h_0(t) \exp(\beta^T G(t)z) \]  

(1)

where \( G(t) \) is a \( p \times p \) diagonal matrix with diagonal elements \( g_j(t), j = 1, ..., p \). If \( G(t) \) is the identity matrix, then the PH model is called time-independent PH (TIPH) model. If \( g(t) = 1 (a < t < b) \), 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.

<table>
<thead>
<tr>
<th>True Data</th>
<th>( \Theta_0 )</th>
<th>are tests valid?</th>
<th>reject ( H_0 )?</th>
<th>error</th>
<th>cases</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>TIPH</td>
<td>TIPH</td>
<td>both MD and residual</td>
<td>no</td>
<td>type I</td>
</tr>
<tr>
<td>2</td>
<td>PWPH</td>
<td>TIPH</td>
<td>both MD and residual</td>
<td>yes</td>
<td>type II</td>
</tr>
<tr>
<td>3</td>
<td>TDPH</td>
<td>TIPH</td>
<td>both MD and residual</td>
<td>yes</td>
<td>type II</td>
</tr>
<tr>
<td>4</td>
<td>PWPH</td>
<td>PWPH</td>
<td>both MD and residual</td>
<td>no</td>
<td>type I</td>
</tr>
<tr>
<td>5</td>
<td>TIPH</td>
<td>TIPH</td>
<td>only MD method</td>
<td>yes</td>
<td>type II</td>
</tr>
<tr>
<td>6</td>
<td>TDPH</td>
<td>TIPH</td>
<td>only MD method</td>
<td>yes</td>
<td>type II</td>
</tr>
<tr>
<td>7</td>
<td>TDPH</td>
<td>PWPH</td>
<td>only MD method</td>
<td>yes</td>
<td>type II</td>
</tr>
<tr>
<td>8</td>
<td>non-PH</td>
<td>TIPH</td>
<td>only MD method</td>
<td>yes</td>
<td>type II</td>
</tr>
</tbody>
</table>

**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 \( h_0(t) = 1 (t > 0) \), and \( Z \sim N(0,1) \). Let \( H_0: h(t|z) = h_0(t)\exp(\beta z) \). The residual method assumes \( h(t|z) = h_0(t)\exp(\beta z + \theta zt) \) 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

\[
P(H_1|H_0)\begin{array}{cccccccc}
T_1 & T_2 & T_3 & T_4 & T_5 & T_6 & \text{residual (1)} & \text{residual (2)} \\
n=50 & 0.036 & 0.035 & 0.024 & 0.029 & 0.027 & 0.034 & 0.056 & 0.004 \\
n=100 & 0.040 & 0.040 & 0.036 & 0.042 & 0.036 & 0.046 & 0.054 & 0.004 \\
n=200 & 0.039 & 0.041 & 0.025 & 0.029 & 0.029 & 0.032 & 0.058 & 0.004 \\
\end{array}
\]

Case 2

**RC data.**

\[
h_{Y|Z}(t|z) = h_0(t) \exp(\beta z + \theta z^2 t),
\]

where \(P(Z = -1) = P(Z = -2) = 0.5, h_0(t) = 1(t > 0), \beta = -0.5, \theta = 1, \) 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 \(H_0^r: \theta = 0 \text{ v.s. } H_1^r: \theta \neq 0.\) 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 3

**Complete data.**

\[
h_{Y|Z}(t|z) = h_0(t) \exp(\beta z^2).\)

\(h_0(t) = 1(t > 0), \beta = 1\) and \(Z \sim \text{unif}(-3, 3).\)

**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 \(H_0^r: \theta = 0 \text{ v.s. } H_1^r: \theta \neq 0.\) 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.
### CASE 4  
Complete data. \( h_{Y|Z}(t|z) = h_0(t) \exp(2z \mathbb{1}(t \geq 0.2) + 2zt \mathbb{1}(t \geq 0.6)) \), where \( h_0(t) = 1(t > 0) \) and \( Z \sim \text{unif}(0,1) \). \( H_0 : h(t|z) = h_0(t) \exp(\beta z \mathbb{1}(t \geq 0.1)) \). The residual method assumes \( h(t|z) = h_0(t) \exp((\beta z + \theta zt) \mathbb{1}(t \geq 0.1)) \). 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 \( T_6 \).

| 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 5  
RC data. \( h_{Y|Z}(t|z) = h_0(t) \exp(z \mathbb{1}(t \geq 1)) \), where \( h_0(t) = 1(t > 0) \), \( C \sim \text{unif}(0,2) \), and \( Z \sim \text{bin}(2,0.5) \). \( H_0 : h(t|z) = h_0(t) \exp(\beta z) \). The residual method assumes \( h(t|z) = h_0(t) \exp(\beta z + \theta zt \mathbb{1}(t \geq 1)) \) 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, \( \text{cox.zph} \), 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 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**

\[
P(H_0|H_1) 
\begin{array}{ccccccc}
T_1 & T_2 & T_3 & T_4 & T_5 & T_6 & \text{residual (2)} \\
n=100 & 0.84 & 0.86 & 0.64 & 0.71 & 0.64 & 0.69 & 0.594 \\
n=200 & 0.82 & 0.84 & 0.60 & 0.58 & 0.58 & 0.59 & 0.249 \\
n=400 & 1 & 1 & 0.1 & 0.17 & 0.13 & 0.30 & 0.029 \\
\end{array}
\]

**Case 6** Complete data. \( h_{Y|Z}(t|z) = h_0(t) \exp(\beta z + \theta z^2 t) \), where \( h_0(t) = 1(t > 0), \beta = -0.5, \theta = 1, \) 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 \( H_0^t: \theta = 0 \text{ v.s. } H_1^t: \theta \neq 0. \) \( H_0 \) 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 \( h_0(t) = 1(t > 0), \beta = -1, \theta = 5 \) and \( Z \sim \text{unif}(0,4). \) \( H_0: h(t|z) = h_0(t) \exp(\beta z). \) The residual method assumes \( h(t|z) = h_0(t) \exp(\beta z + \theta z t) \) and tests \( H_0^t: \theta = 0 \text{ v.s. } H_1^t: \theta \neq 0. \) 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 \( h_0(t) = 1(t > 0) \), \( Z \sim pois(1) \) 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 \( H'_0: \theta = 0 \) v.s. \( H'_1: \theta \neq 0 \). Both \( H_0 \) and \( H'_0 \) are correct. We compare \( (H_1|H_0) \) of these two methods. Here, we only apply existing code, \textit{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. \( h_{Y|Z}(t|z) = h_0(t) \exp(\beta z \mathbf{1}(t \geq 1)) \). \( h_0(t) = 1(t > 0) \), \( \beta = 1 \) 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 \( H'_0: \theta = 0 \) v.s. \( H'_1: \theta \neq 0 \). 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 \( h_0(t) = 1(t > 0) \), \( \beta = -2 \), \( \theta = 2 \) and \( Z \sim N(0,1) \). \( H_0: h(t|z) = h_0(t) \exp(\beta z) \). \( H'_0: \theta = 0 \) v.s. \( H'_1: \theta \neq 0 \). 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. \(h_{Y|Z}(t|z) = h_0(t) \exp(\beta z^2 + \theta z t)\). \(h_0(t) = 1(t > 0)\), let \(\beta = 5, \theta = 1\) and \(Z \sim \text{N}(0,1)\). \(H_0: h(t|z) = h_0(t) \exp(\beta z)\). The residual method assumes \(h(t|z) = h_0(t) \exp(\beta z + \theta z t)\) and tests \(H_0^\ast: \theta = 0\) v.s. \(H_1^\ast: \theta \neq 0\). 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 \(\geq 0.77\).

<table>
<thead>
<tr>
<th></th>
<th>type II error</th>
<th>pseudo type II error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Case 11 (P(H_0</td>
<td>H_1))</td>
<td>(T_1)</td>
</tr>
<tr>
<td>n=50</td>
<td>0.608</td>
<td>0.604</td>
</tr>
<tr>
<td>n=100</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>n=200</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**CASE 12** Complete data. \(h_{Y|Z}(t|z) = h_0(t) \exp(\beta z + \theta z_1(t \geq 0.5))\), where \(h_0(t) = 1(t > 0)\), \(\beta = 1, \theta = 5, Z \sim \text{unif}(-2,2)\). \(H_0: h(t|z) = h_0(t) \exp(\beta z)\). \(H_0^\ast: \theta = 0\) v.s. \(H_1^\ast: \theta \neq 0\). 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.
**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.

| Case 4 $P(H_0|H_1)$ | $T_1$ | $T_2$ | $T_3$ | $T_4$ | $T_5$ | $T_6$ | residual (1) | residual (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           |

<table>
<thead>
<tr>
<th>PH model</th>
<th>type II error</th>
<th>pseudo type II error</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$T_1$</td>
<td>$T_2$</td>
</tr>
<tr>
<td>n= 50</td>
<td>0.982</td>
<td>0.986</td>
</tr>
<tr>
<td>n=100</td>
<td>1.000</td>
<td>1.000</td>
</tr>
<tr>
<td>n=200</td>
<td>1.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>
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 $\Omega_1$ be the event that $\hat{F}_{Y,Z}(t,z) = \frac{1}{n} \sum_{i=1}^{n} 1(Y_i \leq t, Z_i \leq z) \rightarrow F_{Y,Z}(t,z)$ and $\hat{F}_Y(t) = \frac{1}{n} \sum_{i=1}^{n} 1(Y_i \leq t) \rightarrow F_Y(t)$ and let $\Omega_2$ be the event that $\hat{G}(s) = \frac{1}{n} \sum_{i=1}^{n} 1(Z_i \leq s) \rightarrow F_Z(s)$, then by the SLLN, $P(\Omega_2) = 1$ and $P(\Omega_1) = 1$. Let $\Omega_{1h} = \{ \omega \in \Omega, \hat{h}_o(t)(\omega) \rightarrow h_o(t) \}$ and $\Omega_{1s} = \{ \omega \in \Omega : \sup |\hat{h}_o(t)(\omega) - h_o(t)| \rightarrow 0 \}$ and $\Omega_2 = \Omega_{1h} \cap \Omega_{1s} \cap \Omega_1 \cap \Omega_2$, then by Lemma 2 and 3, $P(\Omega_2) = 1$.

**Lemma 4.** Let $(X, \mathcal{F}, P)$ 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, $\Omega_a = \{ \omega \in X : \mu_n(\cdot, \omega) \rightarrow \mu(\cdot, \omega) \text{ set-wisely} \}$, $\Omega_b = \{ \omega \in X : f_n(t,\omega) \rightarrow f(t,\omega) \text{ point-wisely in } t \}$, and $\Omega_c = \{ \omega \in X : g_n(t,\omega) \rightarrow g(t,\omega) \text{ point-wisely in } t \}$. If $P(\Omega_a) = P(\Omega_b) = P(\Omega_c) = 1$, $|f_n| \leq g_n$, and $\int g_n \, d\mu_n \rightarrow \int g \, d\mu < \infty$ almost surely, then $\int f_n \, d\mu_n \rightarrow \int f \, d\mu$ almost surely.

**Proof.** Let $\Omega = \Omega_a \cap \Omega_b \cap \Omega_c$, then $P(\Omega) = 1$. For each $\omega \in \Omega$, $\mu_n(\cdot, \omega) \rightarrow \mu(\cdot, \omega)$ set-wisely, $f_n(t,\omega) \rightarrow f(t,\omega)$ point-wisely in $t$, and $f_n(t,\omega) \rightarrow f(t,\omega)$ point-wisely in $t$. Since $|f_n| \leq g_n$ and $\int g_n \, d\mu_n \rightarrow \int g \, d\mu < \infty$ 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 $P(\Omega) = 1$, $\int f_n \, d\mu_n \rightarrow \int f \, d\mu$ almost surely.

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

**Proof.** Notice that $h_n : = \frac{\sum_{i=1}^{n} 1(Y_i > \hat{\eta}, ||Z_i|| < c_n)}{\sum_{j=1}^{n} 1(||Z_j|| < c_n)} = \frac{\sum_{i=1}^{n} 1(Y_i > \hat{\eta} \land \eta > \hat{\eta}) + 1(\eta > Y_i > \hat{\eta})1(||Z_i|| < c_n)}{\sum_{i=1}^{n} 1(||Z_i|| < c_n)}$. Then

$$\frac{\sum_{i=1}^{n} 1(Y_i > \eta, ||Z_i|| < c_n)}{\sum_{j=1}^{n} 1(||Z_j|| < c_n)} - \frac{\sum_{i=1}^{n} 1(\eta > Y_i > \hat{\eta}, ||Z_i|| < c_n)}{\sum_{j=1}^{n} 1(||Z_j|| < c_n)} \leq h_n \leq \frac{\sum_{i=1}^{n} 1(Y_i > \eta, ||Z_i|| < c_n)}{\sum_{j=1}^{n} 1(||Z_j|| < c_n)} + \frac{\sum_{i=1}^{n} 1(\eta > Y_i > \hat{\eta}, ||Z_i|| < c_n)}{\sum_{j=1}^{n} 1(||Z_j|| < c_n)}.$$

The first terms of the upper bound and the lower bound converge almost surely to $P(Y > \eta | Z = 0)$ by Stute [1986]. Let $\epsilon > 0$, when $n$ is large enough, since $\hat{\eta} \rightarrow \eta$ almost surely, then $|\eta - \hat{\eta}| < \epsilon$ almost
surely. The second term in the upper bound 

\[
\frac{\sum_{i=1}^{n} 1(Y_i \geq \hat{\eta} || Z_i || < c_n)}{\sum_{j=1}^{n} 1(|| Z_j || < c_n)} \leq \frac{\sum_{i=1}^{n} 1(Y_i > \eta - \epsilon, || Z_i || < c_n)}{\sum_{j=1}^{n} 1(|| Z_j || < c_n)} 
\]

\[\rightarrow P(\eta - \epsilon < Y \leq \eta | Z = 0) = S_o(\eta - \epsilon) - S_o(\eta) \text{ almost surely by Stute (1986)}. \]

Since it is true for any \( \epsilon \) 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, 

\[
\frac{\sum_{i=1}^{n} 1(Y_i > \hat{\eta} || Z_i || < c_n)}{\sum_{j=1}^{n} 1(|| Z_j || < c_n)} \rightarrow P(Y > \eta | Z = 0) \text{ almost surely.}
\]

**Lemma 6.** If \( f_n(x) \rightarrow f(x) \) 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) \text{ uniformly on \([a,b]\).}
\]

**Proof.** If \( f_n(x) \rightarrow f(x) \) 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) \text{ 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, 

\[
ZZ^T - ZZ^T \to E[ZZ^T] - E[Z]E[Z^T] = \Sigma_Z \text{ almost surely and } ZY - Z \tilde{Y} \to E[ZY] - E[Z]E[Y] \text{ almost surely, then } \hat{\beta} \to \beta \text{ almost surely. Thus statement (a) holds.}
\]

Now assume \( F_{Y,Z} \in \Theta_{lr} \) and \( \epsilon \perp Z \), \( \hat{S}^*(t; \beta|x) = \hat{S}_o(t - \beta^T x) \to S_o(t - \beta^T x) \text{ almost surely and } S_{Y^*|Z}(t; \beta|x) = S_o(t - \beta^T x) \). Thus statement (b) holds. Moreover, 

\[
\hat{S}^*(t; \hat{\beta}|x) = \hat{S}_o(t - \hat{\beta}^T x) = \frac{\sum_{i=1}^{n} 1(Y_i \geq t - \hat{\beta}^T x, || Z_i || < c_n)}{\sum_{i=1}^{n} 1(|| Z_i || < c_n)} \text{. Let } \hat{\eta} = t - \hat{\beta}^T x \text{ and } \eta = t - \beta^T x. \text{ Since } \hat{\beta} \to \beta \text{ almost surely by Lemma 2(a), } \hat{\eta} \to \eta \text{ almost surely. Since } S_o \text{ is continuous, by Lemma 5, } \hat{S}^*(t; \hat{\beta}|x) \to P(Y > t - \beta^T x | Z = 0) = S_o(t - \beta^T x) = S_{Y^*|Z}(t; \beta|x) \text{ almost surely. Thus statement (c) holds.}
\]

**A.4. Proof of Lemma 3.**

**Proof of (a).** Let 

\[
L_n(\alpha) = \prod_{i=1}^{n} \frac{\exp(\alpha^T G(Y_i)Z_i)}{\sum_{k=1}^{n} \exp(\alpha^T G(Y_i)Z_k)} \text{ and } \mathcal{L}_n(\alpha) = \frac{1}{n} \ln L_n(\alpha) + \ln n, \]

then 

\[
\mathcal{L}_n(\alpha) = \frac{1}{n} \sum_{i=1}^{n} \alpha^T G(Y_i)Z_i - \frac{1}{n} \sum_{i=1}^{n} \ln \left[ \frac{1}{n} \sum_{k=1}^{n} e^{\alpha^T G(Y_i)Z_k} 1(Y_k \geq Y_i) \right].
\]

We shall show that its limit is
\[ \mathcal{L}(\alpha) = E[\alpha^T G(Y_1) Z_1 - \ln p(\alpha, Y_1)], \] where \( p(\alpha, Y_1) = E[e^{\alpha^T G(Y_1) Z_2} 1(Y_2 \geq Y_1) | Y_1] \),

Let \( p_{n}(\alpha, Y_1) = \frac{1}{n} \sum_{k=1}^{n} e^{\alpha^T G(Y_1) Z_k} 1(Y_k \geq Y_1), \) \( f_{n}(\alpha, Y_1) = \ln p_{n}(\alpha, Y_1) \) and \( f(\alpha, Y_1) = \ln p(\alpha, Y_1) \).

By assumption, \( ||G(Y_1) Z_1|| \leq M < \infty. \) If \( \alpha \) is finite, then \( |\alpha^T G(Y_1) Z_1| \) is bounded by some real number \( K \) and \( e^{-K} \leq \exp(\alpha^T G(Y_1) Z_1) \leq e^K. \) Also \( \frac{1}{n} \sum_{i=1}^{n} \alpha^T G(Y_i) Z_i \rightarrow E[\alpha^T G(Y_1) Z_1] \) almost surely by the SLLN.

\[ e^K \geq p_{n}(\alpha, Y_1) \geq e^{-K} \frac{1}{n} \sum_{k=1}^{n} 1(Y_k \geq Y_1) \Rightarrow K \geq f_{n}(\alpha, Y_1) \geq g_{n}(\alpha, Y_1), \]

where \( g_{n}(\alpha, Y_1) = -K + \ln \frac{1}{n} \sum_{k=1}^{n} 1(Y_k \geq Y_1) \). Notice that

\[ \int g_{n}(\alpha, t) d\hat{F}_{Y_1}(t) = \frac{1}{n} \sum_{j=1}^{n} \left[ \ln \frac{1}{n} \sum_{k=1}^{n} 1(Y_k \geq Y_j) - K \right] = \frac{n-1}{n} \frac{1}{n-1} \sum_{i=1}^{n-1} \ln \left( \frac{i}{n(n-1)} + 1 - \frac{i}{n-1} \right) - K. \]

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

\[ R_{n-1} = \frac{1}{n-1} \sum_{i=1}^{n-1} \ln \left( \frac{i}{n(n-1)} + 1 - \frac{i}{n-1} \right) \text{underestimates } \int_{0}^{1} h_{1}(x) dx. \]

Hence \( \int \ln \left[ \frac{\hat{F}_{Y_1}(t)}{n} + (1 - \hat{F}_{Y}(t)) \right] d\hat{F}_{Y}(t) \leq \int_{0}^{1} \ln \left( \frac{x}{n} + 1 - x \right) dx < \infty \) if \( n \geq 2. \)

Let \( h_{2}(x) = \ln \left( \frac{x}{n} + 1 - x \right), \) \( \frac{1}{n-1} < x < 1 + \frac{1}{n-1}, \) then \( h'_{2}(x) < 0. \) Using left-endpoints estimation,

\[ L_{n-1} = \frac{1}{n-1} \sum_{i=0}^{n-2} \ln \left( \frac{i+1}{n(n-1)} + 1 - \frac{i+1}{n-1} \right) = \frac{1}{n-1} \sum_{i=1}^{n-1} \ln \left( \frac{i}{n(n-1)} + 1 - \frac{i}{n-1} \right) \text{overestimates } \int_{1}^{\frac{1}{n-1}} h_{2}(x) dx. \]

Then \( \int \ln \left[ \frac{\hat{F}_{Y_1}(t)}{n} + (1 - \hat{F}_{Y}(t)) \right] d\hat{F}_{Y}(t) \geq \int_{\frac{1}{n-1}}^{1} \ln \left( \frac{x}{n} + 1 - x \right) dx > -\infty \) if \( n \geq 3. \) Also, since

\[ \int \ln \left( \frac{x}{n} + 1 - x \right) dx \rightarrow \int \ln(1-x) dx = -1, \]

\[ \int \ln \left[ \frac{\hat{F}_{Y_1}(t)}{n} + (1 - \hat{F}_{Y}(t)) \right] d\hat{F}_{Y}(t) \rightarrow \int \ln(1-x) dx = -1. \]

Then \( \lim \int g_{n}(\alpha, t) d\hat{F}_{Y_1}(t) \) is finite. Since \( f_{n} \) is bounded by the integrable function \( g_{n} \) and \( f_{n} \rightarrow f \) almost surely, by Lemma 4,

\[ \int f_{n} d\hat{F}_{Y}(t) = \frac{1}{n} \sum_{i=1}^{n} f_{n}(Y_i) \rightarrow \int f dF_{Y}(t) = E[\ln E[e^{\alpha^T G(Y_1) Z_2} 1(Y_2 \geq Y_1) | Y_1]] \]
almost surely and \( L_n(\alpha) \to L(\alpha) \) almost surely, for each \( \alpha \in \mathbb{R}^p \).

By assumption, \( B = \{ \beta : \beta = \text{arg}\sup_{\alpha \in \mathbb{R}^p} L(\alpha) \} \) is a singleton set, \( \beta_0 = \text{arg}\sup_{\alpha \in \mathbb{R}^p} L(\alpha) \) is uniquely determined. Let \( \hat{\beta}_n = \text{arg}\sup_{\alpha \in \mathbb{R}^p} L_n(\alpha) \). Then \( L(\beta_0) = \sup_{\alpha \in \mathbb{R}^p} L(\alpha) \geq L(\alpha) \) and \( L_n(\hat{\beta}_n) = \sup_{\alpha \in \mathbb{R}^p} L_n(\alpha) \geq L_n(\alpha) \) for any \( \alpha \in \mathbb{R}^p \). Since \( L_n(\hat{\beta}_n) \geq L_n(\beta_0) \),

\[
\liminf_{n \to \infty} L_n(\hat{\beta}_n) \geq \liminf_{n \to \infty} L_n(\beta_0) = L(\beta_0) \text{ almost surely} \quad (2)
\]

Let \( \beta^* \) be a limiting point of \( \hat{\beta}_n \) in the sense that there exists a subsequence of \( \hat{\beta}_n(\omega) \), say \( \hat{\beta}_{n_1}(\omega) \), such that \( \hat{\beta}_{n_1}(\omega) \to \beta^* (= \beta^*(\omega)) \). By the assumption in the Lemma 3, \( P(\lim_{n \to \infty} ||\hat{\beta}_n|| < \infty) = 1 \). Let \( \Omega_1 \) be as defined in Remark 1 and let \( \Omega_* = \Omega_1 \cap \lim_{n \to \infty} ||\hat{\beta}_n|| < \infty \). For each \( \omega \in \Omega_* \), \( \beta^*(\omega) \) is finite. Then

\[
L_n(\hat{\beta}_{n_1}(\omega)) = \frac{1}{n} \sum_{i=1}^{n} \hat{\beta}_{n_1}(\omega)^T G(Y_i) Z_i - \frac{1}{n} \sum_{i=1}^{n} \ln \left[ \frac{1}{n} \sum_{k=1}^{n} e^{\hat{\beta}_{n_1}(\omega)^T G(Y_i) Z_k} 1(Y_k \geq Y_i) \right]
\]

Since \( \hat{\beta}_{n_1}(\omega) \to \beta^* \) and \( \beta^* \) is finite, \( \hat{\beta}_{n_1}(\omega) \) is bounded. By the similar argument as above, \( L_n(\hat{\beta}_{n_1}(\omega)) \to L(\beta^*) \) almost surely. Since \( \lim L_n(\hat{\beta}_{n_1}) \geq \liminf_{n \to \infty} L_n(\hat{\beta}_n) \geq L(\beta_0) \) almost surely, we have \( L(\beta^*) \geq L(\beta_0) \) almost surely. Then \( L(\beta^*) = L(\beta_0) \), that is, \( \beta^* = \beta_0 \), as \( B \) is a singleton set. Since every convergent subsequence of \( \hat{\beta}_n \) converges to \( \beta_0, \hat{\beta}_n(\omega) \to \beta_0 \) \( \forall \omega \in \Omega_* \). That is, \( \hat{\beta}_n \to \beta_0 \) almost surely.

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

\[
\Omega_{*'} = \{ \omega \in \Omega : \sup |\hat{S}'_o(t)(\omega) - S_o(t)| < 0 \}. \quad \text{Since} \quad \hat{S}'_o(t) \quad \text{and} \quad \tilde{S}_o(t) \quad \text{has the same asymptotic properties and} \quad P(\Omega_*) = 1 \quad \text{[Yu and Li 1994],} \quad P(\Omega_{*'}) = 1.
\]

Let \( \omega \in \Omega_{*'} \). Since \( h_o \) is a piece-wise constant baseline hazard function, for each \( t \in \{ a_j : j \geq 0 \} \), 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 \( \hat{h}_o(t)(\omega) = 0 \) 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 $\tilde{h}_o(t)(\omega) > 0$, 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 $a_p < Y_{(j-1)} < t \leq Y_{(j)} < a_q$.

When $j < m$, let $\eta_n \to 0$, then there are three cases:

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

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

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

In case (i), notice that $-\ln \tilde{S}_o(t)(\omega) = \int_{s \leq t} \tilde{h}_o(s)(\omega) \, ds$, $-\ln \tilde{S}_o(t)(\omega) + \ln \tilde{S}_o(Y_{(j-1)})(\omega) = \int_{Y_{(j-1)} < s \leq t} \tilde{h}_o(s)(\omega) \, ds = h_j(t - Y_{(j-1)})$. Let $h_n = t - Y_{(j-1)} \geq 0$, since $h_n \leq \eta_n$ and $\eta_n \to 0$, $h_n \to 0$. Then $\tilde{h}_o(t)(\omega) = h_j = -\ln \tilde{S}_o(t)(\omega) + \ln \tilde{h}_o(t)(\omega) \leq \eta_n$. Let $f_n(t) = -\ln \tilde{S}_o(t)(\omega) = -\ln \tilde{S}_o(Y_{j-1})(\omega) + h_j(t - Y_{j-1})$ 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, $\tilde{h}_o(t)(\omega) \to (-\ln S_o(t))' = h_o(t)$.

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

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

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

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

**Proof of (c).** Let $\Omega_s$ be as defined in (b) and let $\Omega_h = \{\omega \in \Omega, \tilde{h}_o(t)(\omega) \to h_o(t)\}$, then by (b), $P(\Omega_h) = P(\Omega_s) = 1$. Let $\omega \in \Omega_h \cap \Omega_s$. Since $\beta$ is finite and $G(t)Z$ is bounded, there exists $0 < \gamma_1 \leq \gamma_2 < \infty$ such that $\gamma_1 \leq \exp(\beta^T G(s)x) \leq \gamma_2$. Then, $\gamma_1 \tilde{h}_o(s)(\omega) \leq \hat{h}_o(s)(\omega) \exp(\beta^T G(s)x) \leq \gamma_2 \hat{h}_o(s)(\omega)$.
Since \(h\) bounded, there exists \(0 < h < \infty\). If \(t = \infty\), then for any \(t\), \(\exp(-\int_0^t \hat{h}_o(s)\omega ds) = \hat{S}_o(t)\omega \to S_o(t) = \exp(-\int_0^t h_o(s)ds) > 0\) and \(\int_0^t \hat{h}_o(s)\omega ds \to \int_0^t h_o(s)ds < \infty\). Since \(f_n(s) = \hat{h}_o(s)\omega \exp(\beta^T G(s)x) \to h_o(s)\exp(\beta^T G(s)x), f_0^t \hat{h}_o(s)\omega \exp(\beta^T G(s)x) ds \to f_0^t h_o(s)\exp(\beta^T G(s)x) ds\) by the dominated convergence theorem. Hence

\[
\hat{S}_o(t; \beta|x) = \exp(-\int_0^t \hat{h}_o(s)\omega ds) \exp(\beta^T G(s)x) ds \to \exp(-\int_0^t h_o(s)\omega ds) \exp(\beta^T G(s)x) ds = S_o(t; \beta|x).
\]

If \(t < \infty\), it is sufficient to show that \(\hat{S}_o(t; \beta|x) \to 0 = S_o(t; \beta|x)\) for \(t \geq b\). Notice that \(S_o(t) = 0\) and \(S_o(t; \beta|x) = 0\) when \(t \geq \tau\). And

\[
\exp(-\int_0^t \hat{h}_o(s)\omega \gamma_1 ds) \geq \exp(-\int_0^t \hat{h}_o(s)\omega ds) \geq \exp(-\int_0^t \hat{h}_o(s)\omega ds) = \hat{S}_o(t; \beta|x) \geq \exp(-\int_0^t h_o(s)\omega \gamma_2 ds).
\]

Since \(\hat{S}_o(t) \to S_o(t)\), \(\exp(-\int_0^t \hat{h}_o(s)\omega \gamma_1 ds) = [\hat{S}_o(t) \omega]^\gamma_1 \to S_o(t) \omega^\gamma_1 = 0, \exp(-\int_0^t \hat{h}_o(s)\omega \gamma_2 ds) = [\hat{S}_o(t) \omega]^\gamma_2 \to S_o(t) \omega^\gamma_2 = 0\) and then \(\hat{S}_o(t; \beta|x) \to 0 = S_o(t; \beta|x)\).

Since \(P(\Omega_h \cap \Omega_s) = 1, \hat{S}_o(t; \beta|x) \to S_o(t; \beta|x)\) almost surely for each \(t\).

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

Since \(P(\Omega_h \cap \Omega_s) = 1, \hat{S}_o(t; \hat{\beta}|x) \to S_o(t; \beta|x)\) almost surely for each \(t\).

References


