Scientific
Research
Publishing

Marginal Distribution Plots for Proportional Hazards Models with Time-Dependent Covariates or Time-Varying Regression Coefficients

Qiqing Yu[1], Junyi Dong[1], George Wong[2]

[1]Department of Mathematical Sciences, SUNY, Binghamton, NY, USA
[2]Department of Integrative Medicine, Mount Sinai Beth Israel, New York, NY, USA
Email: qyu@math.binghamton.edu, jdong9@binghamton.edu, drgeorgewong@yahoo.com

Abstract

Given a sample of regression data from (Y, Z), a new diagnostic plotting method is proposed for checking the hypothesis H_0: the data are from a given Cox model with the time-dependent covariates Z. It compares two estimates of the marginal distribution F_Y of Y. One is an estimate of the modified expression of F_Y under H_0, based on a consistent estimate of the parameter under H_0, and based on the baseline distribution of the data. The other is the Kaplan-Meier-estimator of F_Y, together with its confidence band. The new plot, called the marginal distribution plot, can be viewed as a test for testing H_0. The main advantage of the test over the existing residual tests is in the case that the data do not satisfy any Cox model or the Cox model is mis-specified. Then the new test is still valid, but not the residual tests and the residual tests often make type II error with a very large probability.

Keywords

Cox's Model, Time-Dependent Covariate, Semi-Parametric Set-Up, Diagnostic Plot

1. Introduction

In this paper, we propose a new diagnostic plotting method for the proportional hazards (PH) model [1] with continuous survival time Y, which may be right censored, and with possible time-dependent covariates Z or time-varying regression coefficients β. The new method has some advantages over the existing

methods in the literature, especially when the data do not satisfy any PH model or when the PH model is mis-specified. It provides an alternative diagnostic plot for model checking.

Denote the conditional hazard function and survival function of Y for given Z by $h_{Y|Z}(\cdot|\cdot)$ and $S_{Y|Z}(\cdot|\cdot)$, respectively, or simply $h(\cdot|\cdot)$ and $S(\cdot|\cdot)$. Let $(Y_1, Z_1), \cdots, (Y_n, Z_n)$, be i.i.d. copies of (Y, Z) with the distribution function $F_{Y,Z}$. The PH model was first defined by $h(t|z) = \exp(\beta'z)h_o(t)$, where h_o is the baseline hazard function, z is a $p \times 1$ covariate vector, β is a $p \times 1$ parameter vector, h_o and β are unknown, and p does not depend on n. The model is referred as the time-independent covariate PH (TIPH) model. This model has been extended in two ways: 1) the covariate is time-dependent, *i.e.*, $z = z(t)$ is a function of time t; 2) the regression coefficient is time-varying, *i.e.*, $\beta = \beta(t)$ is a function of time t. For the time-dependent covariates PH (TDPH) model, Kalbfleisch and Prentice [2] distinguish the external ones from the internal ones.

It is often that the external time-dependent covariate Z_i can be written as $Z_i = Z_i(t) = \mathcal{B}(U_i, g_i(t))$, where $\mathcal{B}(\cdot, \cdot)$ is a function, $U_1, \cdots, U_n$ are i.i.d. copies from the time-independent random vector U and $g_i(\cdot)$ is a function of time *t*. A simple example of $\mathcal{B}(U, g(t))$ is $w(U)g(t)$, where $w(\cdot)$ is a function not depending on time t. Without loss of generality (WLOG), we can assume $w(U) = U$. Two simple examples of $g_i(t)$ are (a) $g_i(t) = 1(t \geq a_i)$ and (b) $g_i(t) = 1(t \geq a_i)(t - a_i)$, where a_i is a constant but may depend on subject i [1]. The PH model in case (b) is referred as the linearly-dependent PH model (LDPH model) hereafter. We also consider other forms of $g_i(\cdot)$ for the external case. For the time-varying regression coefficients PH model, in addition to the assumptions made on the TDPH model, one further assumes $h(t|z(t)) = e^{\beta(t)'z(t)}h_o(t)$. A special case of this model is the piecewise PH (PWPH) model with k cut-points $a_1, \cdots, a_k$ [3].

An important step in the data analysis under the PH model is to check whether the model is indeed appropriate for the data. To this end, it is desirable to have some diagnostic plotting methods for the PH model. In the literature, some diagnostic plotting methods under the semi-parametric set-up are designed to inspect whether the data follow the TIPH model $h(t|z) = h_o(t)\exp(\beta'z)$, (*i.e.*, $g(t)$ is constant). One well-known diagnostic method for the PH model is the log-minus-log plots (log-log plots).

Several other graphical methods using residuals to check the PH model assumption have been proposed in the literature [4]. By plotting the estimator of the cumulative hazard functions evaluated at each Cox-Snell residuals against the residuals, one expects a straight line pattern if the data set satisfies the PH model. However, as mentioned by Baltazar-Aban and Pena [5] this plotting method has serious defect. This method may fail to identify the poor fitting in many circumstances. Martingale residuals and its cumulative sums are introduced to examine the functional form of the covariate [4] [6] [7]. By plotting the Martingale residuals without a covariate, say Z_1, against this missing

covariate, one can observe the functional form of Z_1. And by plotting the score process versus follow-up time, one can examine the violation of the proportional hazards assumption. Scheike and Martinussen [8] later extend this method to the time-varying model. Deviance residuals [9] are then introduced. These methods provide more symmetric plots and are useful in identifying the outliers. Moreover, one may use Schoenfeld residuals or the cumulative sums form [10]. The residual plotting methods also induce several residual tests for the PH models.

We provide a new diagnostic plotting method for the PH model. The main idea is to plot the Kaplan-Meier estimator (KME) of S_Y against a proper estimator of the marginal distribution of Y under the selected model. Thus it is called the marginal distribution (MD) plot. The MD plot can be described as a 5-step procedure: 1) Fit the Cox model you have in mind to obtain the regression coefficients. 2) Choose a reference value for the covariate Z, say z_o, such that there exist many observations in its neighborhood, say $\mathcal{N}(z_o)$. 3) Estimate the survival function of Y for $Z = z_o$ using $\mathcal{N}(z_o)$. 4) Use the estimator in 3) to estimate the marginal survival function of Y. 5) Compare the estimator of the marginal survival function in 4) with the KME of S_Y.

The paper is organized as follows. In Section 2, we propose the MD plot and other supplementary diagnostic plots. In Section 3, we present simulation results on the performance of the plot. We also compare the MD plot to the current residual plots. In Section 4, we apply the new diagnostic plot to the long-term breast cancer follow-up data analyzed in Wong *et al.* [11]. Section 5 is a concluding remark, where we explain that the MD plot can also be served as a naive hypothesis test for the PH models.

2. The Main Results

The assumption and notations are given in § 2.1. The idea of the marginal approach is introduced in § 2.2. The method is explained in § 2.3 and § 2.4.

2.1. Assumptions and Notations

Let Θ_{ph} be the collection of all PH models specified by

$$h_{Y|Z}\left(t\mid z(t)\right) = e^{\beta(t)'z(t)}h_o(t) \tag{1}$$

where $\beta(t)$ and $z(t)$ are now possible vectors of functions of t [12]. For model checking under the semi-parametric set-up, the null hypothesis is

$$H_0 : \text{the data are from Model}(1) \text{ with } Z(t) = \mathcal{B}\left(U, g(t)\right) \text{ and given } \mathcal{B}(\cdot, \cdot) \text{ and } g(\cdot), \tag{2}$$

where U is a p-dimensional random covariate vector and the baseline hazard function $h_o(\cdot)$ is unknown. Let Θ be the collection of all possible joint distribution functions $F_{Y,U}$ of (Y,U). Notice that $F_{Y,U}$ does not need to belong to Θ_{ph}. Abusing notation, by $z(t) = ug(t)$, we mean that

$$ug(t) = \left(u_1 g_1(t), \cdots, u_p g_p(t)\right) = D\left(g_1(t), \cdots, g_p(t)\right)\left(u_1, \cdots, u_p\right)', \text{ where}$$

$D\left(g_1(t), \cdots, g_p(t)\right)$ is a $p \times p$ diagonal matrix with diagonal elements

$g_i(t)$'s.

Our method involves the mode, say a vector $c \in \mathbb{R}^p$, of the distribution of the random vector U. That is, $\forall \epsilon > 0$ and $\forall \eta \in \mathbb{R}^p$, $pr(\|U - c\| < \epsilon) \geq pr(\|U - \eta\| < \epsilon)$, where $\|\cdot\|$ is a norm, e.g., $\|U\| = \max_i |U_i|$, where $U = (U_1, \cdots, U_p)$.

Proposition 1. *If* (Y, U) *satisfies Model* (1), *then for each* $c \in \mathbb{R}^p$, $h_{Y|W}(t|w(t)) = h_1(t)\exp\{\beta(t)'w(t)\}$, *where* $h_1(t) = h_o(t)\exp\{\beta(t)'c\}$ *and* $W(t) = Z(t) - c$.

In view of Proposition 1, hereafter, WLOG, we can assume that

AS1. The zero vector 0 is a mode of U and it satisfies that $\|\mathcal{B}(0, g(t))\| = 0$.

Otherwise, let U_o be a mode of U and define $W(t) = Z(t) - c$, where $c = Z(U_o)$. Then by Proposition 1, Model (1) is equivalent to another PH model, where h_1 takes place of the role of the baseline hazard function h_o and $h_1(t) = h_{Y|W}(t|0)$ is also unknown.

Since $S_{Y|U}$ may not satisfy the PH model in the null hypothesis H_0 (see (2)), one can define a new conditional survival function of a new response variable, say Y^*, for given Z such that $S_{Y^*|Z}$ satisfies the PH model in the null hypothesis H_0. Correspondingly, one can define the new marginal survival function of Y^*, say S_{Y^*}. That is,

$$S_{Y^*}(t) = E\left(S_{Y^*|U}(t|U)\right), \quad \text{where} \quad S_{Y^*|U}(t|u) = \exp\left\{-\int_0^t e^{\beta(x)'\mathcal{B}(u, g(x))}h_o(x)\,\mathrm{d}x\right\}. \quad (3)$$

Notice that h_o and S_o are equivalent if h_o exists. Abusing notation, write $S_{Y^*|U}(t|u) = S(t, u, S_o, \beta)$. Notice that S_{Y^*} is a function of the unknown parameter β. Given the distribution function $F_{Y,U}$, if β is not a function of t, then β is the almost sure limit of the maximum partial likelihood estimator (MPLE) of β under H_0 (see Example 1 below), otherwise, it is conceivable that $\beta(t)$ is some limiting point of the estimator of $\beta(t)$ [12].

Example 1. *Assume that* $F_{Y,Z}$ *is a uniform distribution in 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* $x = 4$. *Then the family of distributions* $\{S_{Y|Z}(\cdot|z) : z \in (-1,1) \cup (3,4)\}$ *does not satisfy the PH model, and* $S_Y(t) = S_o(t) = 1 - t$ *for* $t \in [0,1]$. *If one fits the TIPH model* $H_0 : h_{Y|Z}(t|z) = h_o(t)\exp(\beta z)$, *with data from* $F_{Y,Z}$ *without knowing* $F_{Y,Z}$, *then* $S_{Y^*}(t) = E\left((1-t)^{e^{\beta Z}}\right)$ *by* (3), *where* $\beta \approx -0.045$, *which is the limit of the MPLE based on the random sample from* $F_{Y,Z}$ *under* H_0. S_{Y^*} *and* $F_{Y^*,Z}$ *are uniquely determined by* $F_{Y,Z}$ *and* H_0.

Lemma 1. *If* $S_{Y|U}$ *does not follow the model defined in* H_0 *(see* (2)), *then* (a) $S_{Y^*|U} \neq S_{Y|U}$, *and* (b) $S_o(t) = S_{Y^*|U}(t|0) = S_{Y|U}(t|0)$. *Otherwise,* (a) $S_{Y^*|U} = S_{Y|U}$, (b) $S_o(t) = S_{Y^*|U}(t|0) = S_{Y|U}(t|0)$ *and* (c) $S_{Y^*} = S_Y$.

The proof of Lemma 1 is trivial and is skipped.

2.2. The Marginal Distribution Plot

Motivated by Lemma 1, the new plotting method we propose here is to plot an estimator of S_{Y^*}, say $\hat{S}_{Y^*}$, against $\hat{S}_Y$ (the KME of S_Y) together with the

95% confidence band of $\hat{S}_Y$, and

to check whether the graphs of $y = \hat{S}_Y(x)$ and $y = \hat{S}_{Y^*}(x)$ are close

$\left(\text{e.g., whether } y = \hat{S}_{Y^*}(x) \text{ is within or outside the 95\% confidence band of } \hat{S}_Y \right)$,

where $\hat{S}_{Y^*}(t) = \sum_{i=1}^{n} \hat{S}_{Y^*|U}(t|u_i)/n$, $\hat{S}_{Y^*|U}(t|u) = S(t, u, \hat{S}_o, \hat{\beta})$ (see (3)), $\hat{\beta}$ is a consistent estimator of β under H_0, and $\hat{S}_o$ is a consistent estimator of $S_o = S_{Y|U}(\cdot|0)$ under the assumption that $F_{Y,U} \in \Theta$, even if the data do not satisfy the pre-assumed PH model in H_0 (see Remark 1). In the latter case, it is conceivable that $\hat{S}_{Y^*}$ gives a close image of S_{Y^*} If the graphs of $\hat{S}_Y$ and $\hat{S}_{Y^*}$ are close, then it suggests that H_0 in (2) is true. We thus call this plot the marginal distribution plot.

Since the main issue of the MD plot is the estimator $\hat{S}_{Y^*}$ and it is not trivial to find $\hat{S}_{Y^*}$ due to $h_o(\cdot)$ in (3), the main focus of the paper is to introduce how to construct $\hat{S}_{Y^*}$. We shall explain in details how to obtain $\hat{S}_{Y^*}$ through various $g(t)$. We also give $\hat{S}_{Y^*}$ for the general piecewise continuous $g(t)$ or $\mathcal{B}(\cdot, g(t))$.

For simplicity, we shall first explain our method when U or Z is a univariate covariate and $z(t) = \mathcal{B}(u, g(t)) = ug(t)$. We introduce the generalization to the case of a covariate vector (or matrix) in Remark 3 and to the time-dependent model for general $\mathcal{B}(u, g(t))$ in § 2.3.4.

Suppose that $(Y_1, U_1, C_1), \cdots, (Y_n, U_n, C_n)$ are i.i.d. copies from a random vector (Y, U, C), where Y may be subject to right censoring by censoring variable C. The success of the MD plot relies on the proper estimators of S_o, $S_{Y^*|U}(t|u)$ and S_{Y^*}. Denote them by $\hat{S}_o$, $\hat{S}_{Y^*|U}(t|u)$ and $\hat{S}_{Y^*}$, respectively. Since 0 is a mode of the covariate U by assumption AS1, a consistent estimator of S_o is the KME, denoted by $\hat{S}_o$, based on the data satisfying $|U_i| < \epsilon_n$, where $\epsilon_n = rn^{-1/k_o}$ with $k_o > 1$ (*e.g.*, $\epsilon_n = rn^{-1/(3p)}$), and r is a given positive number (e.g., the inter-quartile-range or the standard deviation of U_i's). WLOG, let the first n_o observations be all the observations satisfying $|U_i| < \epsilon_n$. If Y_i's are right censored, then the KME $\hat{S}_o$ is based on $(U_i, Y_i \wedge C_i, \delta_i)$, $i = 1, \cdots, n_o$, where $\delta_i = 1(Y_i \leq C_i)$. For ease of explanation, we only consider the case of complete data in this section. The extension to the right censored data is straightforward and we present simulation results with right-censored data in Section 3. For the complete data, the KME of S_o based on the first n_o observations is

$$\hat{S}_o(t) = \frac{1}{n_o} \sum_{i=1}^{n_o} 1(Y_i > t) = \frac{\sum_{i=1}^{n} 1(Y_i > t) 1\left(|U_i| < rn^{-1/(3p)}\right)}{\sum_{j=1}^{n} 1\left(|U_j| < rn^{-1/(3p)}\right)} \tag{4}$$

$\hat{S}_{Y^*|U}(t|u)$ will be introduced in details in § 2.3 and $\hat{S}_{Y^*}(t) = \sum_{i=1}^{n} \hat{S}_{Y^*|U}(t|u_i)/n$.

Remark 1. Since $\hat{S}_o(t)$ is a kernel estimator, it is well known that under certain regularity conditions, $\hat{S}_o(t)$ converges to $S_o(t)$ in probability and

$\frac{1}{n}\sum_{i=1}^{n}S\left(t,U_i,\hat{S}_o,\beta\right)$ converges to $S_{Y^*}(t)$ in probability for each given β, and $\forall F_{Y,U} \in \Theta$. If the given PH model holds, then $S_Y(t) = S_{Y^*}(t)$ and we expect that the graphs of $\hat{S}_Y(t)$ and $\hat{S}_{Y^*}(t)$ are close, as $\hat{S}_Y$ and $\hat{S}_{Y^*}$ are consistent estimators of S_Y and S_{Y^*}, respectively. Otherwise it is likely that $S_Y(t) \neq S_{Y^*}(t)$ and two curves $y = \hat{S}_Y(x)$ and $y = \hat{S}_{Y^*}(x)$ are apart.

Remark 2. One may wonder whether $\hat{S}_o$ in (4) can be replaced by the existing estimators of the baseline survival function under the PH model, denoted by $\tilde{S}_o$. For instance, under the TIPH model, several consistent estimators of S_o, say $\tilde{S}_o$, can be obtained from the standard statistical packages. For example, the Breslow estimator of baseline survival function can be obtained by applying survfit.coxph( ) in R. However, if the given TIPH model does not satisfy the data, $\tilde{S}_o$ is inconsistent, whereas $\hat{S}_o$ given in (4) is still consistent. In fact, given a joint distribution of a random vector (Y,U), if it does not satisfy the TIPH model, there exists at least one pair of survival function S_1 and vector b satisfying $E\left[\left\{S_1(t)\right\}^{\exp(bU)}\right] = S_Y(t)$, e.g., $(S_1,b) = (S_Y,0)$. It means that the estimator $\tilde{S}_Y$ using $\tilde{S}_0$ will always suggest that the given TIPH model fits the regression data even though the data set may not satisfy the TIPH model, and hence $\tilde{S}_0$ is not a proper choice. Simulation study in § 3.1 suggests that under the assumption in Example 1, $\tilde{S}_Y$ converges to $S_Y(t)$ in probability and using $\tilde{S}_Y$ fails to identify the wrong model assumption.

2.3. Estimation of $S(t,z,S_o,\beta)$

We shall first illustrate the main idea through three typical cases when β is a constant: 1) $g(t) = 1$ *i.e.*, the TIPH model, 2) $g(t) = \left(1(t < a), 1(t \geq a)\right)$ *i.e.*, the PWPH model, 3) $g(t) = (t-a)1(t \geq a)$ *i.e.*, the LDPH model. We also discuss the general case of $g(t)$.

1) Case of $g(t) = 1$, *i.e.*, the TIPH model. Since $S(t,u,S_o,\beta) = \left\{S_o(t)\right\}^{\exp(\beta u)}$ by (3), define

$$\hat{S}_{Y^*|U}\left(t|u\right) = S\left(t,u,\hat{S}_o,\hat{\beta}\right) = \left(\hat{S}_o(t)\right)^{\exp(\hat{\beta}u)} \quad \text{and} \quad \hat{S}_{Y^*}(t) = \sum_{i=1}^{n}\left(\hat{S}_o(t)\right)^{\exp(\hat{\beta}U_i)}\bigg/n, \quad (5)$$

where $\hat{\beta}$ is a consistent estimator of β under the selected PH model, *e.g*, the MPLE.

Remark 3. Even though U in (4) and (5) is a random variable, it is easy to extend to the case that U is a vector. Assume $U = \left(U_1, \cdots, U_p\right)$ is a p-dimensional random vector, n is much larger than p and U is bounded. We can define $\|U\| = \max_{1 \leq i \leq p}|U_i|$ or $\sqrt{\sum_{i=1}^{p}U_i^2}$. In the simulation study, we know the mode of U. In applications, the sample mode is not well defined. We choose a "sample mode" of U_i's, say $c \in \mathbb{R}^p$ as follows: 1) Select a proper radius r and choose points q such that in a neighborhood of q with rasius r, $\mathcal{N}_r(q)$, there are more than 20 (or $n^{1-\frac{1}{3p}}$) observations. Of course, we do not want r to be too large. 2) Among these points q, choose the one, say c, with the largest number of observations in $\mathcal{N}_r(c)$ among all $\mathcal{N}_r(q)$. 3) Then set $U_i^* = U_i - c$, $i = 1, \cdots, n$, (in view of Proposition 1).

2) Case of $g(t) = (1(t < a), 1(t \geq a))$, *i.e.*, the PWPH model with one cut point. Now $\beta = (\beta_1, \beta_2)'$ and

$$S(t, u, v, S_o, \beta) = (S_o(a))^{\{\exp(\beta_1 u) - \exp(\beta_2 v)\} 1(t \geq a)} (S_o(t))^{\exp\{\beta_1 u 1(t < a) + \beta_2 v 1(t \geq a)\}}.$$

The covariate $z(t)$ in (2) is $z(t) = D(1(t < a), 1(t \geq a))(u, v)'$, where D is a 2×2 diagonal matrix. Then $\hat{S}_{Y|U,V}(t|u,v) = S(t, u, v, \hat{S}_o, \hat{\beta})$ and

$$\hat{S}_{Y^*}(t) = \frac{1}{n} \sum_{i=1}^{n} (\hat{S}_o(a_i))^{\{\exp(\hat{\beta}_1 U_i) - \exp(\hat{\beta}_2 V_i)\} 1(t \geq a_i)} (\hat{S}_o(t))^{\exp(\{\hat{\beta}_1 U_i 1(t < a_i) + \hat{\beta}_2 V_i 1(t \geq a_i)\})}, \quad (6)$$

where a_i may depend on the i-th observation. And $g(t) = (1(t < a), 1(t \geq a))'$ corresponds to a special case of the PWPH model with one cut-point. For the PWPH model with more than one cut-point, the derivation of $\hat{S}_{Y^*}$ is similar. Typically for the PWPH model with two cut-points,

$$h(t|u, r, v) = \exp\{\beta' z(t)\} h_o(t) = \exp\{\beta_1 u 1(t < a) + \beta_2 r 1(t \in [a, b)) + \beta_3 v 1(t \geq b)\} h_o(t),$$

where $\beta = (\beta_1, \beta_2, \beta_3)'$, the covariate $z(t) = D(1(t < a), 1(a \leq t < b), 1(t \geq b))(u, r, v)'$, and D is a 3×3 diagonal matrix. Then (3) yields

$$S(t, u, r, v, S_o, \beta) = \begin{cases} (S_o(t))^{\exp(\beta_1 u)} & \text{if } t \in (-\infty, a) \\ (S_o(a))^{\exp(\beta_1 u)} \left(\dfrac{S_o(t)}{S_o(a)} \right)^{\exp(\beta_2 r)} & \text{if } t \in [a, b) \\ (S_o(a))^{\exp(\beta_1 u)} \left(\dfrac{S_o(b)}{S_o(a)} \right)^{\exp(\beta_2 r)} \left(\dfrac{S_o(t)}{S_o(b)} \right)^{\exp(\beta_3 v)} & \text{if } t \in [b, \infty), \end{cases}$$

$$\hat{S}_{Y^*}(t) = \sum_{i=1}^{n} S(t, U_i, R_i, V_i, \hat{S}_o, \hat{\beta}) / n,$$

where $\hat{\beta}$ is an estimator of β.

3) Case of $g(t) = 1(t \geq a)(t - a)$, *i.e.*, the LDPH model. In this situation, $S(t, u, S_o, \beta)$ is not a simple form in terms of $S_o(t)$. Let $\hat{S}_o$ be defined as in (4) and let $a = b_0 < b_1 < \cdots < b_k$ be the discontinuous points of $\hat{S}_o(t)$ for $t > a$, we propose to estimate $S(t, u, S_o, \beta)$ and $S_{Y^*}(t)$ by

$$\hat{S}_{Y^*|U}(t|u) = \begin{cases} \hat{S}_o(t) & \text{if } t < b_1 \\ \hat{S}_o(a) \prod_{i=1}^{j} \left(\dfrac{\hat{S}_o(b_i)}{\hat{S}_o(b_{i-1})} \right)^{e^{(b_i - a)u\hat{\beta}}} & \text{if } t \in [b_j, b_{j+1}) \end{cases} \quad (7)$$

and $\hat{S}_{Y^*}(t) = \sum_{i=1}^{n} \hat{S}_{Y^*|U}(t|U_i) / n$. The reason is as follows. Notice that $\hat{S}_o$ defined in (4) is a step function. It has the same consistency property as

$$\breve{S}_o(t) = \begin{cases} \hat{S}_o(t) & \text{if } t \leq a \\ \hat{S}_o(a) \exp\left(-\int_0^t \breve{h}_o(x) \mathrm{d}x \right) & \text{if } t > a, \end{cases} \quad (8)$$

where $\breve{h}_o(t) = 1(t \in (a_i, b_i]) h_i$, $h_i = \dfrac{1}{\epsilon} \ln \dfrac{\hat{S}_o(b_{i-1})}{\hat{S}_o(b_i)}$ for $t \in (b_{i-1}, b_i]$,

$a_i = b_i - \epsilon$, $i = 1, 2, \cdots, k$, $\epsilon = \min\left\{\left|b_j - b_{j-1}\right| : j = 1, \cdots, k\right\}/n$, and $\breve{S}_o(t) = \hat{S}_o(t)$ at the discontinuous points of $\hat{S}_o(t)$. Since

$$S(t, u, S_o, \beta) = \begin{cases} S_o(t) & \text{if } u = 0 \text{ or } t \le a \\ S_o(a)\exp\left(-\int_a^t h_o(x) e^{(x-a)\beta u}\mathrm{d}x\right) & \text{if } u \ne 0, t > a \end{cases} \quad (\text{by}(3)), \quad (9)$$

$S(t, u, \breve{S}_o, \hat{\beta})$

$$= \begin{cases} \hat{S}_o(t) \\ \qquad \text{if } t \le a_1 \text{ or } u = 0 \\ \hat{S}_o(a)\exp\left(-\sum_{i=1}^j \breve{h}_i \dfrac{\exp\left\{(a_i - a)u\hat{\beta}\right\}}{u\hat{\beta}}\left[\exp\left\{u\hat{\beta}(b_i - a_i)\right\} - 1\right]\right) \\ \qquad \text{if } u \ne 0, t \in \left[b_j, a_{j+1}\right), j \le k \\ \hat{S}_o(a)\exp\left(-\sum_{i=1}^j \breve{h}_i \dfrac{\exp\left\{(a_i - a)u\hat{\beta}\right\}}{u\hat{\beta}}\left[\exp\left\{u\hat{\beta}(t - a_i)\right\} - 1\right]\right) \\ \qquad \text{if } u \ne 0, t \in \left[a_j, b_j\right), j \le k. \end{cases}$$

Write $\breve{S}_{Y^*|U}(t|u) = S(t, u, \breve{S}_o, \hat{\beta})$. At the discontinuous points t of $\hat{S}_o$,

$$\breve{S}_{Y^*|U}(t|u) = \hat{S}_o(a)\exp\left(-\sum_{i=1}^j \breve{h}_i \frac{\exp\left\{(a_i - a)u\hat{\beta}\right\}}{u\hat{\beta}}\left[\exp\left\{u\hat{\beta}(b_i - a_i)\right\} - 1\right]\right)$$

$$\approx \hat{S}_o(a)\exp\left(-\sum_{i=1}^j \breve{h}_i \exp\left\{(b_i - a)u\hat{\beta}\right\}(b_i - a_i)\right)$$

$$= \hat{S}_o(a)\prod_{i=1}^j\left(\frac{\hat{S}_o(b_i)}{\hat{S}_o(b_{i-1})}\right)^{\exp\left\{(b_i - a)u\hat{\beta}\right\}} \quad \text{if } u \ne 0, t = b_j, j \le k (\text{by}(8)).$$

Define $\hat{S}_{Y^*|U}(t|u) = \begin{cases} \hat{S}_o(t) & \text{if } t < b_1 \\ \hat{S}_o(a)\prod_{i=1}^j\left(\dfrac{\breve{S}_o(b_i)}{\breve{S}_o(b_{i-1})}\right)^{\exp\left\{(b_i - a)u\hat{\beta}\right\}} & \text{if } t \in \left[b_j, b_{j+1}\right). \end{cases}$ It

leads to (7). $\hat{S}_{Y^*|U}(t|u)$ is simpler than $\breve{S}_{Y^*|U}(t|u)$ in implementation and they have the same asymptotic properties. Simulation results in section 3 suggest that $\hat{S}_{Y^*}$ is consistent under the selected PH model.

4) The case of other forms of $g(t)$ or $\mathcal{B}(\cdot, g(t))$. It can be shown that the estimator $\hat{S}_{Y^*|U}(t|u)$ in the previous three cases of $g(t)$ are all in the form that it is a step function with discontinuous points $b_1, \cdots, b_m$, which are as the same as those of $\hat{S}_o$, and

$$\hat{S}_{Y^*|U}(t|u) = \begin{cases} \hat{S}_o(t) & \text{if } t < b_1 \\ \hat{S}_o(a)\prod_{i=1}^j\left(\dfrac{\breve{S}_o(b_i)}{\breve{S}_o(b_{i-1})}\right)^{\exp\left\{\mathcal{B}(u, g(b_i))\hat{\beta}\right\}} & \text{if } t \in \left[b_j, b_{j+1}\right), \end{cases} \quad (10)$$

where b_j's are defined in the case of $g(t) = (t - a)\mathbf{1}(t \ge a)$. It can be shown

that if $\mathcal{B}\left(u, g\left(t\right)\right)$ is piece-wise continuous in t, then it also leads to a consistent estimator of $S_{Y^*}\left(t\right)$.

2.4. Supplementary Diagnostic Plots

The MD plot needs to know $g\left(t\right)$. One possible way to conjecture the form of the function $g\left(t\right)$ for a time-dependent covariate is to extend the PWPH plotting method in Wong *et al.* [11] as follows:

$$\text{Plot}\left(-\ln\left(\hat{S}_1\left(t\right)\right), -\ln\left(\hat{S}_o\left(t\right)\right)\right) \text{ for } t > 0 \text{ and check whether it is piecewise linear,} \quad (11)$$

$$\text{where } \hat{S}_1\left(t\right) = \frac{\sum_{i=1}^{n} 1\left(Y_i > t, \left|Z_i - z_1\right| < rn^{-1/(3p)}\right)}{\sum_{i=1}^{n} 1\left(\left|Z_i - z_1\right| < rn^{-1/(3p)}\right)} \quad (12)$$

r is a positive constant and z_1 belongs to the support of Z. For instance, for a PWPH model defined in (6),

$$\ln \hat{S}_1\left(t\right) = \begin{cases} \exp\left(\hat{\beta}_1 u\right) \ln \hat{S}_o\left(t\right) & \text{if } t < a \\ \ln\left(\hat{S}_o\left(a\right)\right)^{\exp\left(\hat{\beta}_1 u\right) - \exp\left(\hat{\beta}_2 v\right)} + \exp\left(\hat{\beta}_2 v\right) \ln \hat{S}_o\left(t\right) & \text{if } t \geq a, \end{cases}$$

which corresponds to two lines: $y = b_1 x$ and $y = a_2 + b_2 x$, and the cut-points can be determined from the PWPH plot (see **Figure 3** in Section 3).

Let $b_1 < \cdots < b_m$ be all the distinct exact observations. In view of the expression in (7) under the PH model with $g\left(t\right) = \left(t - a\right)1\left(t \geq a\right)$, if $b_i \geq a$, then

$$\ln \frac{S_{Y^*|U}\left(b_{i+1}|z\right)}{S_{Y^*|U}\left(b_i|z\right)} \approx \exp\left\{\left(b_{i+1} - a\right)z\beta\right\} \ln \frac{S_o\left(b_{i+1}\right)}{S_o\left(b_i\right)},$$

$$\ln \frac{\hat{S}_1\left(b_{i+1}\right)}{\hat{S}_1\left(b_i\right)} \approx \exp\left\{\left(b_{i+1} - a\right)z_1\hat{\beta}\right\} \ln \frac{\hat{S}_o\left(b_{i+1}\right)}{\hat{S}_o\left(b_i\right)},$$

$$\hat{G}_i = \ln\left[\left\{\ln \frac{\hat{S}_1\left(b_{i+1}\right)}{\hat{S}_1\left(b_i\right)}\right\} \middle/ \left\{\ln \frac{\hat{S}_o\left(b_{i+1}\right)}{\hat{S}_o\left(b_i\right)}\right\}\right] \begin{cases} = 0 & \text{if } b_i < a \\ \approx \left(b_{i+1} - a\right)z_1\hat{\beta} & \text{if } b_i \geq a, \end{cases}$$

where $\hat{S}_1$ and z_1 are given in (11) and (12). Thus another diagnostic plotting method is

$$\text{to plot}\left(b_i, G_i\right), \text{ for } i \in \left\{1, \cdots, m\right\} \quad (13)$$

and check whether it appears as a two-piecewise-linear curve: one is $y = 0$ and another one is $y = \left(x - a\right)b$. If so, it is likely that $g\left(t\right) = \left(t - a\right)1\left(t \geq a\right)$, where a is the intersection of the two line segments. We shall call the plotting method the LDPH plot, as $g\left(t\right) = \left(t - a\right)1\left(t \geq a\right)$ corresponding to the LDPH model. The advantage of the PWPH plot and the LDPH plot is that they provide clues on the cut-points, which are needed in the MD plot, unless the cut-point is given.

Remark 4. If the cut-points in the PWPH or TDPH model vary from observation to observation, then the PWPH plot as in (11) and LDPH plot as in (13) do not work. However the cut-points a_i are also observations in such cases, in addition to $\left(Y_i, U_i\right)$'s (in the case of complete data). Thus we do not

need to guess the cut-points, and one can replace $\hat{S}_j$ in (12) by

$$\hat{S}_j(t) = \frac{\sum_{i=1}^{n} 1\left(Y_i > t, |U_i - z_j| < r_1 n^{-1/3}, |a_i - a| < r_2 n^{-1/(3p)}\right)}{\sum_{l=1}^{n} 1\left(|U_l - z_j| < r_1 n^{-1/3}, |a_l - a| < r_2 n^{-1/(3p)}\right)},$$

where a is a predetermined reference cut-point and r_1 and r_2 are positive constants.

3. Simulation Study

In order to compare the MD plot to the other plots, we present three sets of simulation results in §3.1, §3.2 and §3.3, respectively. The mode is 0 as assumed in AS1. S_o can be uniform or exponential distributions. The covariate can be discrete or continuous. The data do not need to be from a PH model. There is no unit in time t due to the simulation.

3.1. Simulation Study under the Assumptions in Example 1

Two random samples of $n = 30$ and $n = 300$ pseudo random numbers Y_i are generated from $U(0,1)$ distribution. For each $Y_i = y_i$, generate Z_i from $U(y_i - 1, y_i)$ with probability 0.5 and from $U(3, 4)$ with probability 0.5. These (Y_i, Z_i) satisfy the assumptions given in Example 1. Let $W_i = 1(Z_i \geq 3)$. Then the family of distributions $\left\{ S_{Y_i|Z_i}(\cdot|z) : z \in (-1,1) \cup (3,4) \right\}$ does not satisfy the null hypothesis $H_0 : h(t|z) = h_o(t)\exp(\beta z)$, but it can be shown that $\left\{ S_{Y_i|W_i}(\cdot|z) : z \in \{0,1\} \right\}$ does.

The sample of size $n = 300$ is only used for the MD plots in panels (1,3) and (3,3) of **Figure 1** with data (Y_i, Z_i)'s and data (Y_i, W_i)'s respectively.

3.1.1. The MD Plots Successfully Identify the Violation of the TIPH Model Assumption When Data Are (Y_i, Z_i)'s

Since $S_{Y|Z}$ does not satisfy the PH model, a proper estimate of S_{Y^*} is expected to deviate from $\hat{S}_Y$. It is seen from the two MD plots with data (Y_i, Z_i)'s in panels (1,2) and (1,3) of **Figure 1** that (a) when $n = 30$, $\hat{S}_{Y^*}$ lies most of the time outside or around the edge of the 95% confidence band of $\hat{S}_Y$; (b) when $n = 300$, $\hat{S}_{Y^*}$ is totally outside the confidence band of $\hat{S}_Y$. The MD plots suggest that our estimator $\hat{S}_{Y^*}$ appears quite different from $\hat{S}_Y$ even when $n = 30$ and they suggest that the data (Y_i, Z_i), $i = 1, 2, \cdots, n$ are not from the TIPH Model, as expected.

3.1.2. The MD Plots Correctly Support the TIPH Model for (Y_i, W_i)'s

Since $S_{Y|W}$ satisfies the PH model, a proper estimate of S_{Y^*} should be close to $\hat{S}_Y$. It is seen from both of the MD plot with data (Y_i, W_i)'s in panels (3,2) and (3,3) of **Figure 1** that the curves of $\hat{S}_{Y^*}$ corresponding to $n = 30$ and $n = 300$ both lie within the confidence bands of $\hat{S}_Y$. It is a strong evidence that the data (Y_i, W_i)'s are from the TIPH model, as expected.

3.1.3. About $\tilde{S}_{Y^*}$ Defined in Remark 2

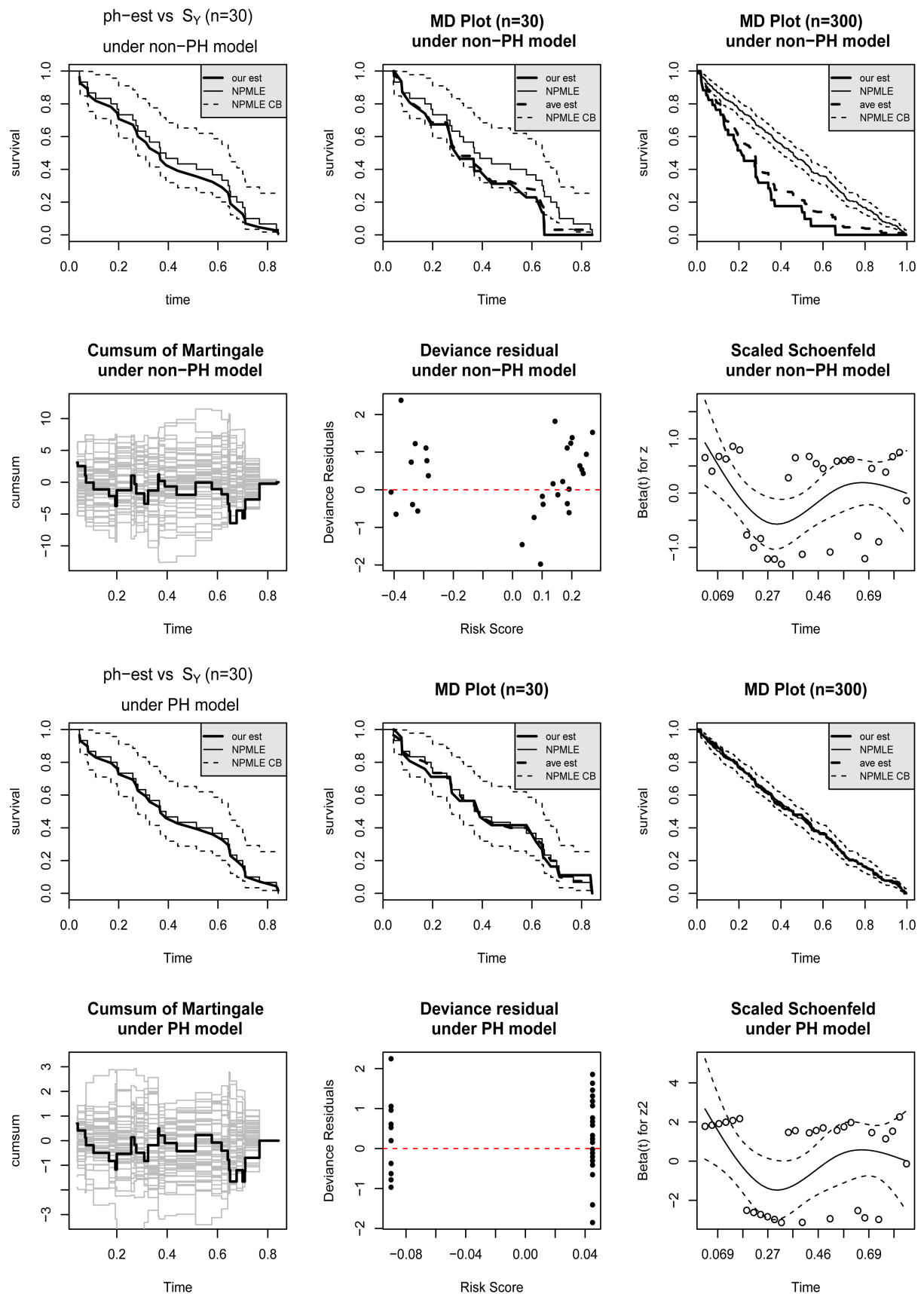To show that the consistent estimator of S_o is the key in the MD approach, we

Figure 1. Diagnostic plots based on Example 1.

present in panels (1,1) and (3,1) of **Figure 1** the plots of the curves of "ph est" ($\tilde{S}_{Y^*}$) and the curves of $\hat{S}_Y$, together with its confidence band, with data (Y_i, Z_i)'s and (Y_i, W_i)'s, respectively. It is seen that the curves of "ph est" and $\hat{S}_Y$ almost coincide for both data sets even when $n = 300$. These plots suggest that the curve of "ph est" is always close to the curve of the KME, regardless of whether the pre-assumed model fits the data. Thus $\tilde{S}_Y$ is not a proper choice for the diagnostic plot.

3.1.4. About Residual Plots

In panels (2,1) and (4,1) of **Figure 1**, the residual plots using the score process by the cumulative martingale residuals method are plotted with data (Y_i, Z_i)'s and (Y_i, W_i)'s, respectively. There is no big difference in pattern between these two plots. Notice that $S_{Y|Z}$ does not satisfy the TIPH model but $S_{Y|W}$ does. Thus this residual plot can not tell whether the data satisfy the TIPH model in this example.

In panels (2,2) and (4,2) of **Figure 1**, the deviance residuals are plotted with data (Y_i, Z_i)'s and (Y_i, W_i)'s, respectively. In panels (2,3) and (4,3) of **Figure 1**, the scaled Schoenfeld residuals are plotted with data (Y_i, Z_i)'s and (Y_i, W_i)'s, respectively. There is no big difference in pattern between these two pairs of plots. Thus these two types of residual plots can not tell whether the data satisfy the TIPH model in this example.

3.1.5. About Residual Tests

We also carried out simulation study on the testing $H_0 : h(t|z) = h_o(t) \exp(\beta' z)$ with data (Y_i, Z_i)'s and using the residual test in the existing R package. Our simulation study suggests that for $n = 50$, 100 or 200 and with a replication of 5000, the residual test does not reject the incorrect H_0 for more than 93% of the time. Thus, it is not surprised that the residual plots do not work well.

3.2. Simulation Based on Data from the TIPH Model

A sample of complete data with $n = 300$ is generated from the TIPH model: $h(t|z) = \exp(\beta z^2) h_o(t)$, where $\beta = 1$, $h_o(t) = 1$ and $Z \sim \text{Norm}(0,1)$. Panels (1,1), (1,2), (1,3) and (3,1) in **Figure 2** presents four different plots by fitting the data set into the TIPH model $h(t|z) = \exp(\beta z) h_o(t)$. Thus the covariate is mis-specified.

The second sample of complete data with size $n = 300$ is generated from the model, $h(t|z) = \exp(\beta z) h_o(t)$, where $\beta = 1$, $Z \sim \text{Norm}(0,1)$ and $h_o(t) = 1$, $t > 0$. Panels (2,1), (2,2), (2,3) and (4,1) in **Figure 2** presents 4 plots by fitting the data set into the TIPH model $h(t|z) = \exp(\beta z) h_o(t)$. Thus the covariate is correctly specified.

3.2.1. Residuals Plots

Panels (1,1) and (2,1) in **Figure 2** display the residual plots using the score process by the cumulative martingale residuals method by fitting data from the first and the second TIPH model respectively into $h(t|z) = \exp(\beta z) h_o(t)$. It is
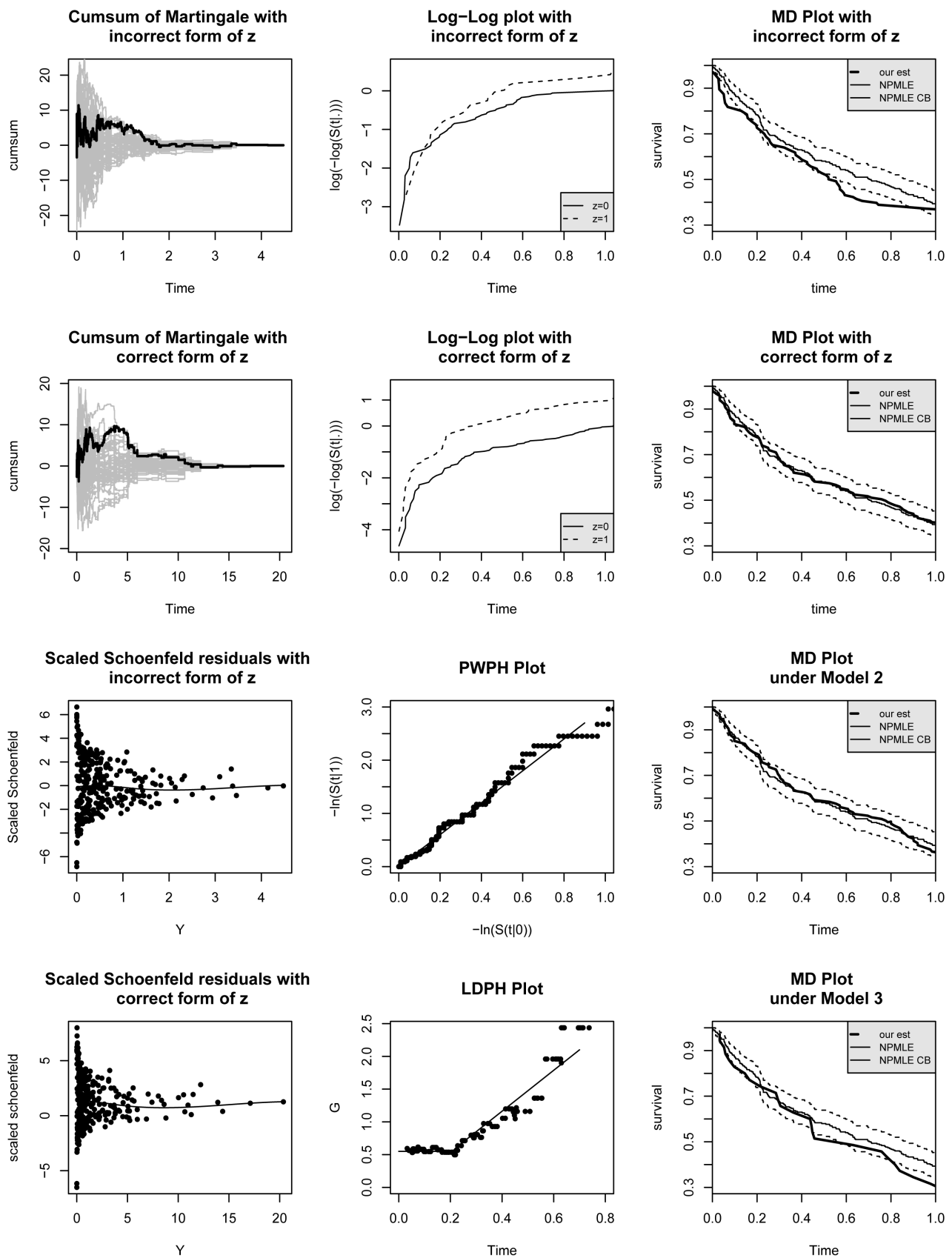
Figure 2. Diagnostic plots based on data from Model: $h(t|z) = h_o(t)e^{\beta z}$ or $h(t|z) = h_o(t)e^{\beta z^2}$.

not easy to distinguish the mis-specified model from the correct one by this pair of residual plots.

Similarly, in **Figure 2**, Panels (3,1) and (4,1) display the scaled Schoenfeld residual plots for fitting the first and second samples, respectively, into the TIPH model $h(t|z) = \exp(\beta z) h_o(t)$. Moreover, Panels (1,2) and (2,2) present the log-log plots in a similar pattern. Again, those methods do not provide enough information about the overall fitting of the models.

3.2.2. Residuals Tests

In fact, with the data from the first mis-specified model and with a moderate sample sizes $n \geq 50$, our simulation study with a replication of 5000 suggests that the residual test in the existing R package (e.g., cox.zph( )) would not reject the mis-specified TIPH model for more than 70% of the time. Thus it is not surprised that the residual plots cannot detect the mis-specified TIPH model.

3.2.3. MD Plots

The MD plot in panel (1,3) fits the mis-specified TIPH model with data from the first sample. It successfully identifies that the functional form of the covariates Z is mis-specified for the first data set, as $\hat{S}_{Y^*}$ is almost totally outside the 95% confidence band of $\hat{S}_Y$. In other words, the MD approach suggests that the first data set does not follow the PH model $h(t|z) = \exp(\beta z) h_o(t)$. On the other hand, the MD plot in panel (2,3) successfully identifies that the functional form of the covariates Z is correct for the second data set, as $\hat{S}_{Y^*}$ is totally inside the 95% confidence band of $\hat{S}_Y$.

Based on the second sample, the modified PWPH plot and the LDPH plot are displayed in panels (3,2) and (4,2); the MD plots under the PWPH and LDPH Models are displayed in panels (3,3) and (4,3). The PWPH plot in panel (3,2) suggests that the data are either from a TIPH Model, or from a PWPH model with one cut-point at $a \approx 0.7$. The MPLE of the regression coefficient under the TIPH Model is $\hat{\beta} = 1.26$ and the MPLE of the regression coefficients under the PWPH model with $k = 1$ is $(\hat{\beta}_1, \hat{\beta}_2) = (1.30, 1.20)$ with $SE = (0.170, 0.136)$. Both $\hat{\beta}_1$ and $\hat{\beta}_2$ are not significantly different from $\hat{\beta}$, as their differences from $\hat{\beta}$ are within two SEs. Both MD plots in panels (2,3) and (3,3) suggest that the PWPH model with at most one cut-point fits the data, as expected, as both curves of $\hat{S}_{Y^*}$ are totally inside the 95% confidence band of $\hat{S}_Y$.

The LDPH plot in panel (4,2) suggests that the LDPH Model may fit the data, but it is seen from panel (4,3) that even within the interval [0,1], only less than 30% of the curve of $\hat{S}_{Y^*}$ lies inside the confidence band of $\hat{S}_Y$. Thus the MD plot suggests that the data are not from the LDPH Model, as expected. It is seen that the LDPH plot performs not as good as the MD plot.

3.3. Simulation on the LDPH Model

A sample of $n = 300$ right-censored data is generated under the LDPH Model, where $h(t|z) = \exp(\beta u g(t)) h_o(t)$, $g(t) = (t-a)1(t \geq a)$, $a = 0.2$, and

$h_o(t) = e^{-t}$. U has a Poisson distribution with mean 1, and $\beta = 10$. It is subject to right censoring and the right censoring variable $C \sim U(1,2)$.

The modified PWPH plots and the LDPH plot are given in panels (1,1), (2,1) and (3,1) in **Figure 3**. The MD plots and the qqplots for the TIPH, PWPH and LDPH Models are given in the second and the third columns of **Figure 3**.

Both the PWPH plot and the MD plots with corresponding qqplot in panels (1,1), (1,2) and (1,3) suggest that the data are not from the TIPH Model. We
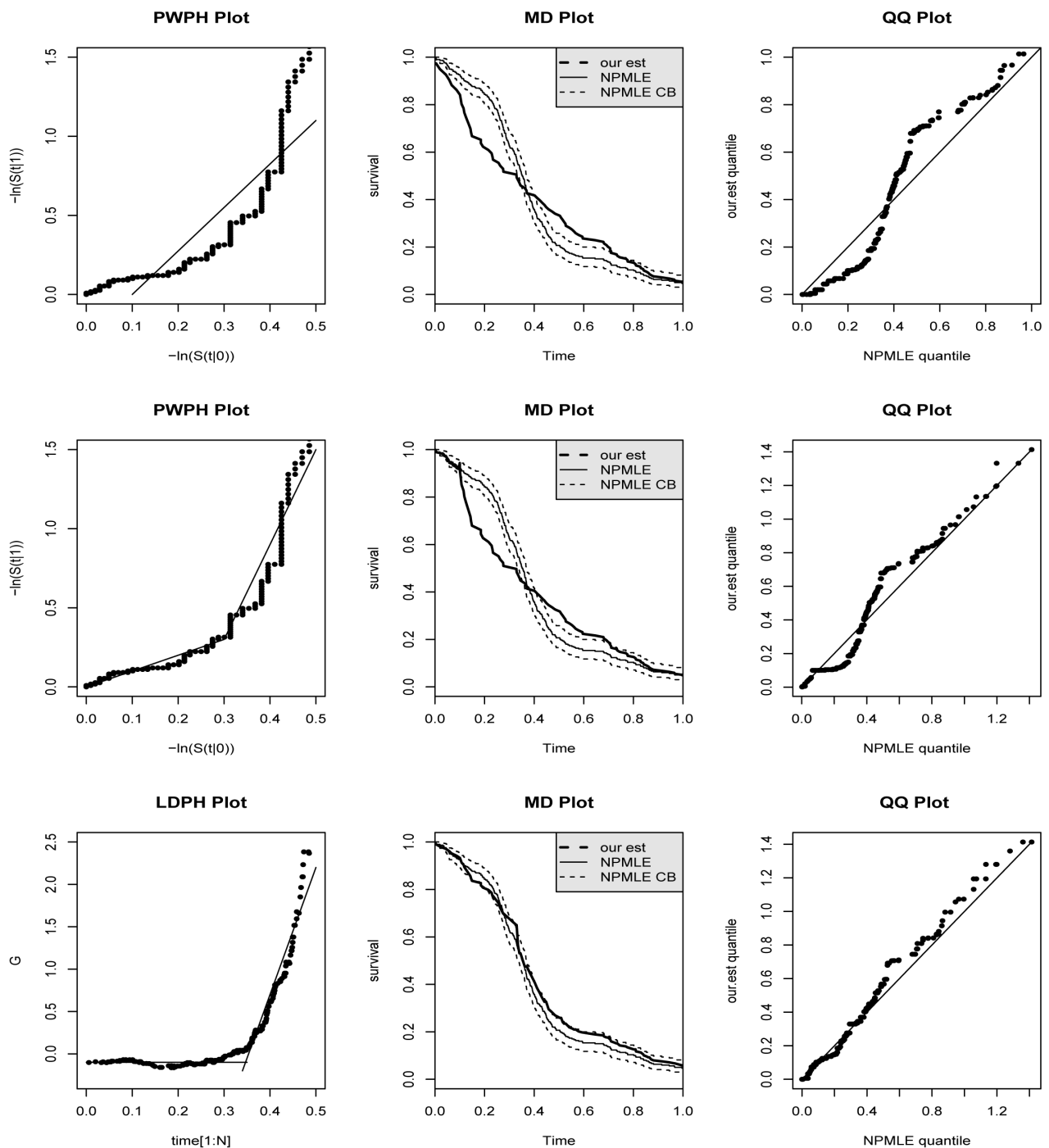
Figure 3. Diagnostic plots based on data from the LDPH Model.

need the information from the PWPH plot to decide the cut-point needed in the MD plots. The PWPH plot in panel (2,1) suggests that the data may be from a PWPH model with a cut-point a satisfying $-\ln S(a|0) \approx 0.05$, that is, $a \approx 0.1$. However, the MD plot in panel (2,2) suggests that the data are not from the PWPH model. This again indicates that the MD plot performs better than the modified PWPH plot.

The LDPH plot in panel (3,1) in **Figure 3** suggests that the data are from the LDPH Model with the cut-point a in $(0.2, 0.4)$. The MD plot and the corresponding qqplot in panels (3,2) and (3,3) suggest that the data are from the LDPH Model, as expected. Thus the LDPH plot has the advantage that it can suggest the cut-point a in the case that a is not given in the LDPH Model. Since U is discrete in this case, the PWPH plot and the LDPH plot perform better than those in the previous simulation studies when U is continuous.

4. Data Analysis

A common situation that will involve the use of the PH model is a long-term clinical follow-up study. In such a study, the impact of a prognostic variable may change at different time periods. This is the case in the breast cancer data analyzed in Wong *et al.* [11]. The data are obtained from an Institutional Review Board approved long-term clinical follow-up study on 371 women with stages I-III unilateral invasive breast cancer treated by surgery at Memorial Sloan-Kettering Cancer Center in New York City between 1985 and 2001. The median follow-up time of the study is 7.4 years (range is 1 month-180 months (14.8 years)), which is the longest among published studies on bone marrow micro-metastasis (BMM).

One objective of the study is to investigate whether tumor diameter is significant in predicting early or late relapse. Then the relapse time Y is the response and the tumor diameter Z is the covariate. Clinical consideration and survival plots suggest late failure can be considered at time greater than 5 years from initial breast cancer surgery. Data analysis based on a PWPH model with two cut-points at 2 years and 5 years with covariate Z is carried out in Wong *et al.* [11]. We apply our diagnostic plotting methods to check whether such a PWPH model fits our data in the case that Z is continuous and Z is discretized as a dichotomized variable (> 2 cm or ≤ 2 cm), as is done in Wong *et al.* [11]. In the latter case, we try the log-log plots as well.

We apply our data to the TIPH model with covariate Z (which is basically continuous), the regression coefficient is $\hat{\beta} = 0.35$ and is significant. We shall only present our new methods for the case that the tumor diameter Z is continuous in panel (2,1) of **Figure 4**. The modified log-log plots in panel (1,1) seems to suggest that the TIPH Model fits the data, but the MD plot in panel (2,1) suggests that the TIPH Model does not fit our data, as the curve of $\hat{S}_{Y^*}$ lies almost totally outside the 95% confidence band of $\hat{S}_Y$. Thus the partial likelihood estimate $\hat{\beta} = 0.35$ is not valid.

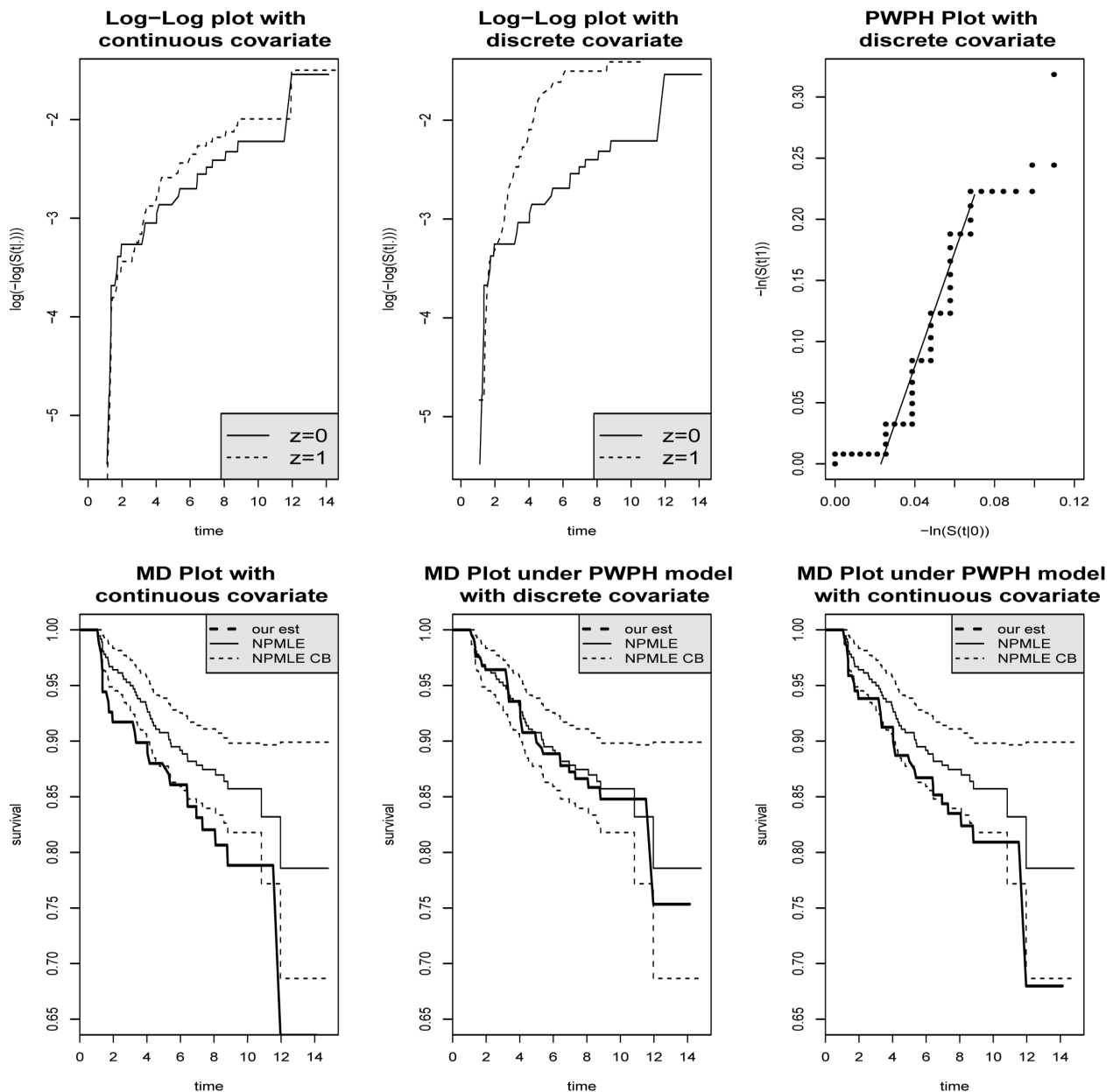Wong *et al.* [11] suggest to do data analysis by dichotomizing Z according

Figure 4. Diagnostic plots for cancer data with the tumor diameter measurements.

to $Z > 2 \, \text{cm}$ or $Z \leq 2 \, \text{cm}$. Then the covariate is discrete, and the standard log-log plots, as well as the PWPH plot proposed by Wong *et al.* [11], is applicable. These two plots are given in panels (1,2) and (1,3) of **Figure 4**. In view of the graph, the log-log plots suggest that the TIPH Model does not fit the discretized data. Moreover, it does not present useful information on the other possible PH models. However, the PWPH plot in panel (1,3) does suggest that the PWPH model with two cut-points close to 1.5 years and 6 years fits the discretized data. Wong *et al.* [11] suggest to do data analysis using a PWPH model with the cut-points at 2 and 5 years (which are close to 1.5 and 6). Thus we further compute the partial likelihood MLE for model with

$g(t) = \left(1(t < 2), 1(t \in [2,5)), 1(t \geq 5)\right)$. The estimates $\hat{\beta}_1$, $\hat{\beta}_2$ and $\hat{\beta}_3$ are −1.76,

2.02 and 0.34, with p-value 0.77, 0.001 and 0.52, respectively. In panels (2,2) and (2,3) of **Figure 4**, we provide the MD plots based on the discrete and continuous covariates, respectively, for fitting the PWPH model. In particular, since the curve of $\hat{S}_{Y^*}$ lies totally inside the 95% confidence band of $\hat{S}_Y$ in panel (2,2), the MD plot suggests that the PWPH model considered in Wong *et al.* [11] fits the discretized data. Thus their data analysis is valid. However, the MD plot in panel (2,3) indicates that the original data set with the continuous covariate Z does not fit a PWPH model, as the curve of $\hat{S}_{Y^*}$ lies totally on one edge of the 95% confidence band of $\hat{S}_Y$.

5. Concluding Remarks

Our simulation results and the data analysis suggest that the MD plot has certain advantages over the existing residual plots, especially when the null hypothesis H_0 in (2) is mis-specified or the data are not from any PH model. Our MD plot does not involve residuals studied in the literature, and this is the first difference between the residual approaches and the MD approach.

5.1. Drawback and Remedy of the MD Plot

The MD approach is closely related to H_0 in (2) with $Z(t) = \mathcal{B}(U, g(t))$, where the parameter β is unknown but $\mathcal{B}(\cdot, g(t))$ is given. The MD plot is applicable to all PH models with $Z(t) = \mathcal{B}(U, g(t))$ and with all types of covariates U, provided that $\mathcal{B}(\cdot, g(t))$ is given. The assumption of a given $\mathcal{B}(\cdot, g(t))$ may be viewed as a drawback of the MD plot if one wants to find a certain PH model to fit the data. There are several ways to overcome this drawback.

1) For the case that $\mathcal{B}(u, g(t)) = u g(t)$, we also propose a modified PWPH plot and LDPH plot in §2.4 for inferring the functional form of $g(t)$.

2) One can apply the MD approach to several possible typical models. For instance, in §3.2 and §3.3, given a data set, the MD approach finds which of the 3 semi-parametric PH models fits.

3) Of course, one can also make use of the existing residual approaches in the literature for guessing $g(t)$. There is no harm to inspect all diagnostic plots available based on the data.

On the other hand, the MD plot can further check the validity of the function forms suggested by the existing residual plots. As illustrated in the paper, with the help of the confidence band of the KME $\hat{S}_Y$, it is more reliable and more informative than the residual plots on whether the model suggested by the residual plots is appropriate for the data. Thus the MD plot is a nice complement to the existing diagnostic methods, not a replacement to them.

5.2. A Naive but Valid Model Test

As seen from the simulation results, when the data are not from any PH model (see Example 1 in §3.1) or H_0 is mis-specified (see simulation on the TIPH Model in §3.2), the MD plots can successfully reject H_0, provided that n is large enough $(n \geq 30)$. Thus the MD plot can play a role of a model checking

test ϕ_{md}, that is,

$$
\begin{cases}
\text{reject } H_0 & \text{if } \hat{S}_{Y^*} \text{ lies inside the confidence band} \left(\text{case}(I)\right) \\
\text{do not reject } H_0 & \text{if } \hat{S}_{Y^*} \text{ lies outside the confidence band} \left(\text{case}(II)\right) \\
\text{inconclusive} & \text{otherwise} \left(\text{case}(III)\right).
\end{cases}
$$

On the contrary, in such case, our simulation results suggest that the residual plots often cannot detect that H_0 is incorrect and the residual tests often do not reject H_0. There must be some reasons.

As summarized in Therneau and Grambsch [13] "Interestingly, nearly all of the tests for proportional hazards that have been proposed in the literature are $T(G)$ tests" "and differ only in their choice of the time transform g(t)". The parameter space Θ_r for the residual approach contains all distributions that follow Model (1) with $\gamma z(t) = \beta \mathcal{B}(U,t) + \theta \mathcal{B}_2(U,t)$, where $\gamma = (\beta, \theta)$. For instance, in the case of Example 1, let $\gamma z(t) = \beta u + \theta ut$. So a residual test is to test $H_0^r : \theta = 0$ v.s. $H_1^r : \theta \neq 0$, instead of testing H_0 in (2) v.s. $H_1 : H_0$ is not true. The parameter space Θ_r for testing H_0^r v.s. H_1^r is a subset of the parameter space Θ for testing H_0 v.s. H_1, where Θ is the collection of all $F_{Y,U}$, where Y is continuous. When the data are not from any PH model or Θ_r is mis-specified, the residual tests are not valid. Moreover, in that case, it is somewhat true that $\theta = 0$ most of the time, and thus when H_0 is not rejected, it is likely to make type II error. In real applications, we do not know whether the data are indeed from some PH model, or whether we select a correct sub-model of the PH model.

Remark 5. In summary, if the existing residual tests reject H_0, the decision is likely to be correct. Otherwise, we have no confidence to believe that the test is valid. In this regard, it is interesting to point out that the parameter space for the aforementioned "test" induced by the MD plot is actually Θ. Thus the MD approach is valid even if $F_{Y,U} \notin \Theta_r$. It can also detect the incorrect model assumption Θ_r by testing the new null hypothesis $H_0 : F_{Y,U} \in \Theta_r$ when n is not too small, but the residual approach cannot accomplish this task. This is the second difference between the MD approach and the residual approach. In application, if H_0 is not rejected, the residual tests often make type II error, and we strongly suggest to consider the MD approach then.

Of course, very often, neither case (I) nor (II) happens, then we need some more rigorous model checking tests. The difference of $\hat{S}_Y$ and $\hat{S}_{Y^*}$ is a natural choice for a test statistic, and it is one of our on-going research. The other is to extend the MD approach to other common regression models, such as the linear regression model and the generalized additive model.

Acknowledgements

References

[1]   Cox, D.R. and Oakes, D. (1984) Analysis of Survival Data. Chapman and Hall, New

York.

[2] Kalbfleisch, J.D. and Prentice, R.L. (1980) The Statistical Analysis of Failure Time Data. Wiley, New York.

[3] Zhou, M. (2001) Understanding the Cox Regression Models with Time-Change Covariates. *American Statistician*, **55**, 153-155.
https://doi.org/10.1198/000313001750358491

[4] Therneau, T.M., Grambsch, P.M. and Fleming, T.R. (1990) Martingale-Based Residuals for Survival Models. *Biometrika*, **77**, 147-160.
https://doi.org/10.1093/biomet/77.1.147

[5] Baltazar-Aban, I. and Pena, E. (1995) Properties of Hazard-Based Residuals and Implications in Model Diagnostics. *Journal of the American Statistical Association*, **90**, 185-197. https://doi.org/10.1080/01621459.1995.10476501

[6] Barlow, W.E. and Prentice, R.L. (1998) Residuals for Relative Risk Regression. *Biometrika*, **75**, 65-74. https://doi.org/10.1093/biomet/75.1.65

[7] Lin, D.Y., Wei, L.J. and Ying, Z. (1993) Checking the Cox Model with Cumulative Sums of the Martingale-Based Residuals. *Biometrika*, **81**, 557-572.
https://doi.org/10.1093/biomet/80.3.557

[8] Scheike, T.H. and Martinussen, T. (2004) On Estimation and Tests of Time-Varying Effects in the Proportional Hazards Model. *Scandinavian Journal of Statistics*, **31**, 51-62. https://doi.org/10.1111/j.1467-9469.2004.00372.x

[9] Hastie, T.J. and Tibshirani, R.J. (1990) Generalized Additive Models. Chapman and Hall, New York.

[10] Grambsch, P.M. and Therneau, T.M. (1994) Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. *Biometrika*, **81**, 515-526.
https://doi.org/10.1093/biomet/81.3.515

[11] Wong, G.Y.C., Osborne, M.P., Diao, Q. and Yu, Q. (2016) Piecewise Cox Models With Right-Censored Data. *Communication in Statistics: Simulation and Computation*, Online.

[12] Tian, L., Zucker, D. and Wei, L.J. (2005) On the Cox Model with Time-Varying Regression Coefficients. *Journal of the American Statistical Association*, **100**, 172-183.
https://doi.org/10.1198/016214504000000845

[13] Therneau, T.M. and Grambsch, P.M. (2000) Modeling Survival Data: Extending the Cox Model. Springer, Berlin. https://doi.org/10.1007/978-1-4757-3294-8