Supplementary Material for Diagnostic Plotting Methods for Proportional Hazards Models With Time-dependent Covariates or Time-varying Regression Coefficients

BY QIQING YU, JUNYI DONG
Department of Mathematical Sciences, SUNY, Binghamton, NY, 13902
qyu @ math.binghamton.edu

AND GEORGE WONG
Department of Integrative Medicine, Mount Sinai Beth Israel, New York, NY 10003

1 Introduction

In section 2, we introduce the residuals plots for checking the PH model assumptions. In section 3, we present the proofs of example 1 and example 2 (Yu, Dong, & Wong, 2016).

2 Residuals in PH model

Under the time-independent proportional hazards model $h(t|z) = h_o(t)\exp(\beta' z)$, where $h_o(t)$ is the baseline hazard function, β is a $p \times 1$ parameter vector, and z is a $p \times 1$ covariate vector. Assume (X, Z) follows the TIPH model. Let C be a censoring variable and let $Y = \min(X, C)$ and $\delta = \mathbb{1}(X \leq C)$. Assume (Y_1, Z_1, δ_1), (Y_2, Z_2, δ_2), ..., (Y_n, Z_n, δ_n) are i.i.d. copies fo (Y, Z, δ). As defined by Lin (1993), let $N_i(t) = \delta_i \mathbb{1}(Y_i \leq t)$ be a counting process, let $h_o(t)$ and $H_o(t)$ be the hazard function and the cumulative hazard function, respectively. Then the log partial likelihood (Cox, 1972) is

$$\sum_{i=1}^{n} \int_0^{\infty} \delta_i \Big[\mathbb{1}(Y_i \geq t)\beta' Z_i - \log\Big\{ \sum_{j=1}^{n} \mathbb{1}(Y_j \geq t) e^{\beta Z_j} \Big\} \Big] dN_i(t) \tag{1}$$

Define

$$S^{(r)}(\beta, t) = \sum_{j=1}^{n} \mathbb{1}(Y_j \geq t)\exp(\beta' Z_j) Z_j^{\otimes r} \tag{2}$$

$r = 0, 1, 2$, where $Z_j^{\otimes 0} = 1$, $Z_j^{\otimes 1} = Z_j$ and $Z_j^{\otimes 2} = Z_j Z_j'$. Then then partial likelihood score function is

$$\sum_{i=1}^{n} \delta_i \Big\{ Z_i - \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \Big\}$$

One can estimate β by maximizing partial likelihood in (1), denote by $\hat{\beta}$.

2.1 Cox-Snell residuals

The Cox-Snell residual (Cox and Snell, 1986) is defined as

$$cs_j = \hat{H}_o(Y_j)\exp(\hat{\beta}' Z_j), \quad j = 1, 2, ..., n \tag{3}$$

where $\hat{H}_o(t)$ is the Breslow estimator of the cumulative hazard function and $\hat{\beta}$ is the MPLE. The reason is as follows. Assume the data set is complete and Y is continuous, then $H(t|z) = H_o(t)\exp(\beta' Z_j)$. Also $H(Y|z)$ follows exponential distribution with mean 1 as $H(Y|z) = -\log S(Y|z)$ and $S(Y|z) \sim unif(0, 1)$.

Hence if the Cox model assumptions are satisfied, then the plot of $\hat{H}_o(cs_j)\exp(\beta' z_j)$ against cs_j is roughly linear.

2.2 Martingale residuals

The Martingale residuals are defined as

$$\hat{M}_i(t) = N_i(t) - \int_0^t \mathbb{1}(Y_i \geq u)\exp(\hat{\beta}' Z_i)\,d\hat{H}_o(u) \quad i = 1,2,\dots,n$$
$$\text{where} \quad \hat{H}_o(t) = \int_0^t \frac{\sum_{j=1}^n dN_j(u)}{S^{(0)}(\hat{\beta},u)} \tag{4}$$

Also, since Z_i is independent of time t, one can simplify (4) by

$$\hat{M}_i(t) = \delta_i - \hat{H}_o(Y_i)\exp(\hat{\beta}' Z_i) \tag{5}$$

The Martingale residual is the difference at time t between the observed and expected number of events for the $i-th$ subject (Lin, Wei & Ying, 1993). Also Lin, Wei & Ying (1993) list several properties of the Martingale residuals. (a) For any t, $\sum_{i=1}^n \hat{M}_i(t) = 0$. (b) For n large, $\mathbb{E}[\hat{M}_i(t)] \approx 0$ and $Cov[\hat{M}_i(t), \hat{M}_j(t)] \approx 0$, for $i \neq j$.

Therneau, Grambsch & Fleming (1990) recommended using residuals to assess the functional form of the covariate and model adequacy with respect to the PH assumptions. Specifically, assume the PH model is of the form

$$h(t,z,x) = h_o(t)\exp(\beta' z + f(x)) \tag{6}$$

where $X \in \mathscr{R}$, $f(\cdot)$ is a function but unknown. To check the functional form of the covariate X, one can plot the Martingale residuals $\tilde{M}_i(t)$ from fitting the model without the covariate X against X. If the plot shows a linear trend, then the covariate is linear in the link function; otherwise, transformation is necessary.

Remark 1. ***Connection between Cox-Snell residuals and Martingale residuals*** *Cox Snell residuals are defined as* $cs_i = \hat{H}_o(Y_i)\exp(\beta' Z_i)$. *In the time-independent situation, the Martingale residuals are* $\hat{M}_i(t) = \delta_i - \hat{H}_o(Y_i)\exp(\hat{\beta}' Z_i)$. *Hence, in the TIPH model,* $\hat{M}_i(t) + cs_i = \delta_i$.

2.3 Deviance Residuals

As noted by Therneau, Grambsch & Fleming (1990), Martingale residuals are skewed in some circumstances. It creates difficulties in interpreting the plot. The Deviance residuals is a function of Martingale residuals and it is less skewed and more normally distributed. The Deviance residuals are defined by

$$d_i = sgn(\hat{M}_i)\Big[-2\big(\hat{M}_i + \delta_i \log(\delta_i - \hat{M}_i)\big)\Big]^{1/2} \tag{7}$$

One can plot the Deviance residuals against the linear predictor $\epsilon = \hat{\beta}' Z$

2.4 Cumsum of Martingale Residuals

Lin, Wei, and Ying (1993) recommends using cumulative sums of Martingale residuals to examining the functional form of a covariate. Let $\hat{M}_i$ be the Martingale residuals as defined in (4), the partial sum of $\hat{M}_i$

is defined as

$$W_j(x) = \sum_{i=1}^{n} \mathbb{1}(Z_{ij \le x}) \hat{M}_i \quad j = 1, 2, ..., p \tag{8}$$

They showed that the null distribution of $W_j(x)$ can be estimated by

$$\begin{aligned}\hat{W}_j(x) = &\sum_{l=1}^{n} \delta_l \Big\{ \mathbb{1}(Z_{lj} \le x) - \frac{\sum_{k=1}^{n} \mathbb{1}(Y_k \ge Y_l) e^{\hat{\beta}' Z_k} \mathbb{1}(Z_{kj} \le x)}{\sum_{k=1}^{n} \mathbb{1}(Y_k \ge Y_l) e^{\hat{\beta}' Z_k}} \Big\} G_l \\ &- \sum_{k=1}^{n} \int_0^t \mathbb{1}(Y_k \ge s) e^{\hat{\beta}' Z_k} \mathbb{1}(Z_{kj} \le x) \Big\{ Z_k - \overline{Z}(\hat{\beta}, s) \Big\}' d\hat{H}_o(s) \\ &\times \mathscr{F}^{-1}(\hat{\beta}) \sum_{l=1}^{n} \delta_l \Big\{ Z_l - \overline{Z}(\hat{\beta}, Y_l) \Big\} G_l \end{aligned} \tag{9}$$

where G_l, $l = 1, 2, ..., n$ are i.i.d sample from standard normal distribution, $\hat{\beta}$ is the PLE of β, $\overline{Z}(\beta, t)$ is $S^{(1)}(\beta, t)/S^{(0)}(\beta, t)$, and $\mathscr{F}^{-1}$ is the estimated covariance matrix of β.

2.5 Schoenfeld residuals

Further, let the conditional weighted mean and variance of the covariate vector at time t (Grambsch &Therneau, 1994)be

$$M(\beta, t) = S^{(1)}(\beta, t)/S^{(0)}(\beta, t) \tag{10}$$

$$V(\beta, t) = \frac{S^{(2)}(\beta, t)}{S^{(0)}(\beta, t)} - \Big\{ \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \Big\} \Big\{ \frac{S^{(1)}(\beta, t)}{S^{(0)}(\beta, t)} \Big\}' \tag{11}$$

The Schoenfeld residuals and the scaled Schoenfeld residuals respectively are defined as

$$r_k(\beta) = Z_k - M(\beta, y_k) \;\text{ and }\; r_k^*(\beta) = V_k^{-1} r_k + \beta \tag{12}$$

Notice that $M(\beta, t)$ is the contional expectation of Z_i given R_i, where $R_i = \{j = 1, 2, ..., n : Y_j \ge Y_i\}$ is the indices of observations at risk at time Y_i. The the Schoenfeld residual $r_k(\beta)$ is the difference between the observed covariate Z_k and its contional expectations $\mathbb{E}[Z_i|R_i]$.

One can estimate the Schoenfeld residuals by

$$\hat{r}_k = Z_k - M(\hat{\beta}, t) \tag{13}$$

Grambsch &Therneau (1994)shows that if the TIPH model holds, $\mathbb{E}[\hat{r}_k] \approx 0$ and the plot of $\hat{r}_k$ against failure time will be centered around 0. Grambsch &Therneau (1994) further recommend to replace the estimator of V_k in (12), $\hat{V}_k$, by $\overline{V} = \mathscr{F}/d$ to estimate the scaled Schoenfeld residuals, where $\mathscr{F}$ is the second derivative of the partial likelihood and d is the number of events. Thus, an estimator of the scaled Schoenfeld residuals is

$$\hat{r}_k^* = \overline{V}^{-1} \hat{r}_k + \hat{\beta} = \overline{V}^{-1} (Z_k - M(\hat{\beta}, t)) + \hat{\beta} \tag{14}$$

Therneau, Grambsch & Fleming (1982) recommend plotting the martingale residuals obtained from the model ignoring one covariate, say Z_1, against the missing covariate and check the functional form of Z_1. Lin & Yin (1993) recommend plotting the cumulative sums of the martingale residuals. Schoenfeld (1982) and Lin (1991) suggest plotting the Schoenfeld residuals against failure times. Wei (1984), Therneau, Grambsch & Fleming (1990) recommend plotting the cumulative sums form. And Grambsch & Fleming (1994) also suggest a smoothed scatter plot the the estimate of the scaled Schoenfeld residuals against the failure time.

Schoenfeld (1982) extend it to the model

$$h(t|z,w) = h_o(t)\exp(\beta' z + \theta g(t) w) \tag{15}$$

with $g(t)$ varying about 0. Under this model, the Schoenfeld residual

$$\mathbb{E}[\hat{r}_k] \approx g(t)\Big\{\mathbb{E}[W^2|R_i] - \mathbb{E}[W|R_i]^2\Big\} \tag{16}$$

The sign of $\mathbb{E}[\hat{r}_k]$ depends only on the sign of $g(t)$ as the second term in (14) is non-negative. By plotting Schoenfeld residual $\hat{r}_k$ against failure time Y_k, one can observe the changes in $g(t)$.

3 Proof of Example 1 and Example 2

Proof of the last statement in Example 1. Let

$$S_0(t) = S_{Y|W}(t|0) = \int_{-\infty}^{\log 0.5} S_{Y|Z}(t|x)dF_Z(x)(= \int_0^{0.5}(1-t)^x dx = \frac{(1-t)^{0.5}-1}{\log(1-t)}$$
$$S_1(t) = S_{Y|W}(t|1) = \int_{\log 0.5}^{0} S_{Y|Z}(t|x)dF_Z(x)(= \int_{0.5}^{1}(1-t)^x dx = \frac{1-t-(1-t)^{0.5}}{\log(1-t)}$$

If the family of survival functions $\{S_0, S_1\}$ follows the PH model, then $S_1(t) = (S_o(t))^{exp(\beta)}$, $\forall\ t \in (0,1)$. Thus

$$(S_0(t))^{e^\beta} = (\frac{(1-t)^{0.5}-1}{\log(1-t)})^{e^\beta} = \frac{1-t-(1-t)^{0.5}}{\log(1-t)},\ \forall t \in (0,1).$$

It follows that

$$\beta = \log\frac{(1-t-(1-t)^{0.5})/\log(1-t)}{((1-t)^{0.5}-1)/\log(1-t)}$$

is constant in $t \in (0,1)$, but

$$\beta = \log\frac{(1-t-(1-t)^{0.5})/\log(1-t)}{((1-t)^{0.5}-1)/\log(1-t)} \approx \begin{cases} -0.35 & \text{if} \quad t = 0.5 \\ -0.69 & \text{if} \quad t = 0.75 \end{cases}$$

Hence, $\{S_0, S_1 : W \in \{0,1\}\}$ does not follow the PH model. □

Proofs in Example 2. Verify that $S_{Y|Z}(0.5|-0.9) = 0$ and $S_{Y|Z}(0.5|3.5) = 0.5$. Thus $S_{Y|Z}(t|z)$ does not satisfy the PH model. Otherwise, $0.5 = S_{Y|Z}(0.5|3.5) = (S_o(0.5))^{exp(3.5\beta)} = ((S_o(0.5))^{exp(-0.9\beta)})^{exp(0.9\beta+3.5\beta)} = (S_{Y|Z}(0.5|-0.9))^{exp(0.9\beta+3.5\beta)} = 0$.

If one discretizes Z by $W = \mathbf{1}(Z > 3)$, then $F_{Y|W}(t|0) = F_{Y|W}(t|1) = t,\ t \in (0,1)$. Thus the family of survival functions $\{S_{Y|W} : W \in \{0,1\}\}$ satisfies the PH model (with $\beta = 0$).

Moreover, if one generates a random sample of complete data from (Z, Y), and sets

$$\hat{S}_k(t) = \frac{\sum_{i=1}^n \mathbf{1}(Y_i > t, |Z_i - k| < n^{-1/3})}{\sum_{j=1}^n \mathbf{1}(|Z_j - k| < n^{-1/3})}, k \in \{0,3\}$$

then it is easy to check that both converge to $S(t) = (1-t),\ t \in (0,1)$. Thus the graphs of $\hat{S}_0$ and $\hat{S}_3$ will almost coincide if n is large, and their limits satisfy the PH model trivially with the equal hazard function.

References

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

[2] Kalbfleisch, J.D. and Prentice, R. L. (1980). The statistical analysis of failure time data. *Wiley* NY.

[3] Lee, E.T. (1992). *Statistical methods for survival data analysis* (2nd ed.), Wiley, NY.

[4] 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.

[5] G. Y. C. Wong, M. P. Osborne, Q. Diao and Q. Yu (2015). "Piecewise Cox Models With Right-Censored Data". Submitted.

[6] W .E. Barlow, R.L. Prentice (1988). "Residuals for relative risk regression". Biometrika 75, 65-74.

[7] T. M. Therneau, P. M. Grambsch, and T. R. Fleming (1990). "Martingale-based residuals for survival models". Biometrika 77, 147-160.

[8] Therneau, T., and Grambsch, P. (2000). Modeling survival data : extending the Cox model. Springer.

[9] P.M. Grambsch and T.M. Therneau (1994). "Proportional Hazards Tests and Diagnostics Based on Weighted Residuals". Biometrika 81, 515-526.

[10] T.H. Scheike and T. Martinussen (2004). "On estimation and tests of time-varying effects in the proportional hazards model". Scandinavian Journal of Statistics 31, 51-62.

[11] D.Y. Lin, L.J. Wei, Z. Ying (1993). "Checking the Cox model with cumulative sums of the martingale-based residuals". Biometrika 80, 557-572.

[12] Zhou, M. (2001). "Understanding the Cox regression models with time-change covariates," *American Statistician*, 55, 153-155.

[13] Baltazar-Aban, & Pena E. "Properties of hazard-based residuals and implications in model diagnostics". *J Amer Statist Assoc.* 1995;90:185–197.