Diagnostics for the Cox Model

Yishu Xue, Elizabeth D. Schifano

Presented by Tianjian Shi

Dec. 4th 2018

Introduction

- Proportional hazards model proposed by Cox (1972), the most popular regression model for the analysis of time-to-event data

- Key features

1. No specification regarding to the form of the baseline hazard function
2. The proportional hazards assumption: when the predictor variables do not vary over time, the hazard ratio comparing any two observations is constant with respect to time

- A few questions

1. Is the proportional hazards assumption satisfied?
2. Are the functional forms of the variables appropriate?
3. Are there any outliers or influential observations?

Preliminaries

- The Cox model (1972), also called the proportional hazards model.
- The hazard for individual i

$$\lambda_i(t) = \lambda_0(t)\exp(\beta' X_i),$$

- The hazard ratio for two subjects with covariate vectors X_i and X_j is

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t)\exp(\beta' X_i)}{\lambda_0(t)\exp(\beta' X_j)} = \exp\{\beta'(X_i - X_j)\},$$

- To incorporate time-dependent covariates, we use the notation $X_i(t)$ to allow for the possibility of time-dependent covariates

$$\frac{\lambda_i(t)}{\lambda_j(t)} = \frac{\lambda_0(t)\exp\{\beta' X_i(t)\}}{\lambda_0(t)\exp\{\beta' X_j(t)\}} = \exp\{\beta'(X_i(t) - X_j(t))\},$$

Preliminaries

- Estimation of β is based on the partial likelihood introduced by Cox (1972) and later formulated by Cox (1975):

$$PL(\beta) = \prod_{i=1}^{n} \left\{ \frac{\exp\{\beta' X_i(t)\}}{\sum_{j=1}^{n} \exp\{\beta' X_j(t)\} Y_j(t)} \right\}^{\delta_i},$$

- Using the counting process formulation of Fleming and Harrington (1991), let $N_i(t)$ be the number of events for subject *i* at time *t*, and define

$$dN_i(t) = I(T_i \in [t, t + \Delta t), \delta_i = 1),$$

- The partial likelihood can be alternatively written as

$$PL(\beta) = \prod_{i=1}^{n} \prod_{t \geq 0} \left\{ \frac{Y_i(t)\gamma_i(\beta, t)}{\sum_j Y_j(t)\gamma_j(\beta, t)} \right\}^{dN_i(t)},$$

where $\gamma_i(\beta, t)$ is the risk score for subject i, $\gamma_i(\beta, t) = \exp\{\beta' X_i(t)\} \equiv \gamma_i(t)$.

Preliminaries

- Take the logarithm,

$$\ell(\beta) = \sum_{i=1}^{n} \int_0^\infty \left[Y_i(t)\beta' X_i(t) - \log \sum_j Y_j(t)\gamma_j(t) \right] dN_i(t).$$

- Differentiate with respect to β

$$U(\beta) = \sum_{i=1}^{n} \int_0^\infty [X_i(s) - \bar{x}(\beta, s)] dN_i(s),$$

where $\bar{x}(\beta, s)$ is a weighted average of X over those observations still at risk at time s with weights $Y_i(s)\gamma_i(s)$:

$$\bar{x}(\beta, s) = \frac{\sum_{i=1}^{n} Y_i(s)\gamma_i(s)X_i(s)}{\sum_{i=1}^{n} Y_i(s)\gamma_i(s)}.$$

Preliminaries

- The $p \times p$ information matrix is given by

$$\mathcal{I}(\beta) = \sum_{i=1}^{n} \int_0^{\infty} V(\beta, s) dN_i(s),$$

where $V(\beta, s)$ is the weighted variance of X at time s:

$$V(\beta, s) = \frac{\sum_i Y_i(s)\gamma_i(s)[X_i(s) - \bar{x}(\beta, s)]'[X_i(s) - \bar{x}(\beta, s)]}{\sum_i Y_i(s)\gamma_i(s)}.$$

- The expected information matrix is unknown, the observed information matrix is available as

$$\mathcal{I}(\hat{\beta}) = -\left.\frac{\partial^2 \ell(\beta)}{\partial\beta\partial\beta'}\right|_{\beta=\hat{\beta}}.$$

Diagnostics for the Cox model

- Proportional hazards assumption
- Function forms
- Outlying observations
- Influential observations

A Case Study

- The dental restoration longevity data, provided by the University of Iowa College of Dentistry's Geriatric and Special Needs (SPEC) Clinic
- 5-year period from 1/1/1995–12/31/1999
- 697 unique patients who went to the SPEC Clinic to treat their molars upon their first visit and received restoration
- Among the 697 patients, 228 experienced an event during the follow-up, giving a censoring rate of 67.3%.
- They considered the following covariates: Gender, Age at receiving restoration, Occupation (Faculty, Non-faculty) and Size (Small, Medium, Large).

Proportional Hazards Assumption

- Schoenfeld (1980) proposed a chi-squared goodness-of-fit test statistic
- Schoenfeld residual is of the form *Observed – Expected*

$$r_k\left(\hat{\beta}\right) = X_{(k)} - E\left(X_{(k)}|R_k\right), \quad k = 1, \ldots, d$$

- In practice, we replace β with $\hat{\beta}$ and obtain $\widehat{r_k}$.
- If the proportional hazards assumption holds, a plot of Schoenfeld residuals against event times will approximately scatter around 0.

Proportional Hazards Assumption

- Grambsch and Therneau (1994) generalized the approach in Schoenfeld (1982) to test the proportional hazards assumption. Assuming the true hazard function is of the time-varying form:

$$\lambda_i(t) = \lambda_0(t) \exp\left[\{\beta + G(t)\theta\}' X_i(t)\right],$$

where $G(t)$ is a diagonal matrix with jj element $g_j(t)$, they showed that the test statistic

$$T(G) = \left(\sum G_k \hat{r}_k\right)^T D^{-1} \left(\sum G_k \hat{r}_k\right)$$

with

$$D = \sum G_k \hat{V}_k G_k^T - \left(\sum G_k \hat{V}_k\right)\left(\sum \hat{V}_k\right)^{-1} \left(\sum G_k \hat{V}_k\right)^T,$$

Proportional Hazards Assumption: Case Study

Table 1: Articles and their functional forms of $G(t)$ falling under the framework of Grambsch and Therneau (1994)

Article	$g(t)$
Cox (1972), Gill and Schumacher (1987), Chappell (1992)	A specified function of time
Schoenfeld (1980), Moreau *et al.* (1985), O'Quigley and Pessione (1989)	Piecewise constant on non-overlapping time intervals with the constants and intervals predetermined
Harrell (1986)	$g(t) = \bar{N}(t-)$, tests the correlation between the rank of the event times and the Schoenfeld residuals
Lin (1991)	The proposed test is equivalent to $g(t) = t$ when the maximizer of a weighted partial likelihood, $\hat{\beta}_w$, is based on a one-step Newton-Raphson algorithm staring from $\hat{\beta}$
Nagelkerke *et al.* (1984)	Let $g_j(t_1) = 0$ and $g_j(k+1) = a_j^2 \hat{r}_{jk}$, $j = 1, \ldots, p$ to test for the serial correlation of the Schoenfeld residuals, where a_j is the weight of the j^{th} covariate;

Proportional Hazards Assumption: Case Study

Table 2: Proportionality test results for Model 1 and Model 2

	Model 1		Model 2	
	χ^2 Stat	p-value	χ^2 Stat	p-value
Male	0.586	0.444	0.429	0.513
Age	4.029	0.045	3.555	0.059
Age^2	-	-	0.638	0.424
Non-Faculty	0.429	0.513	0.558	0.455
SizeMedium	1.560	0.212	1.711	0.191
SizeSmall	0.298	0.585	0.416	0.519
Global	6.788	0.237	6.932	0.327

Table 3: Cox regression results for tooth restoration failure for the Model 2

	Estimate	exp(Estimate)	Standard error	Z Stat	p-value
Male	−0.221	0.802	0.137	−1.612	0.107
Age	0.206	1.228	0.0076	2.709	0.007
Age^2	−0.092	0.912	0.085	−1.075	0.282
Non-Faculty	0.116	1.123	0.146	0.795	0.427
SizeMedium	−0.140	0.869	0.165	−0.850	0.395
SizeSmall	−0.510	0.601	0.169	−3.018	0.003

Proportional Hazards Assumption: Case Study

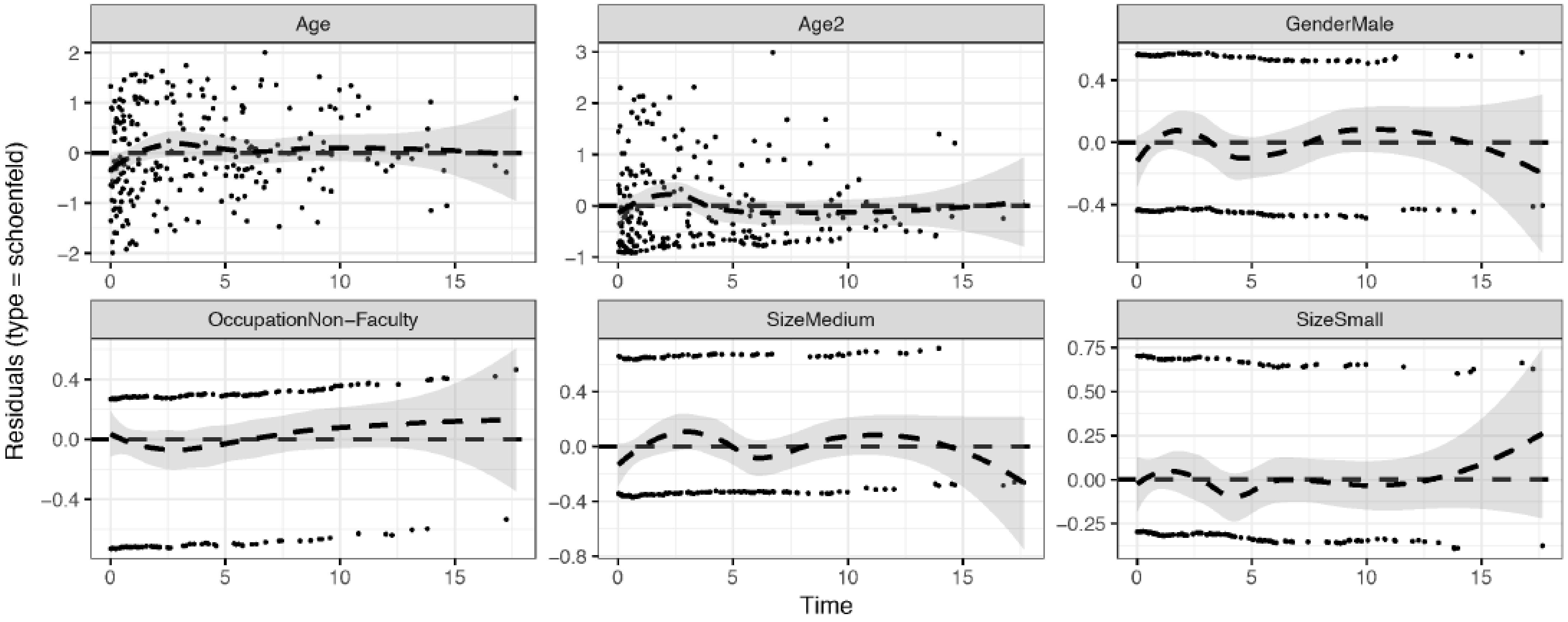

Figure 2: *Schoenfeld residuals for each covariate against survival time.*

Proportional Hazards Assumption: Graphical Methods

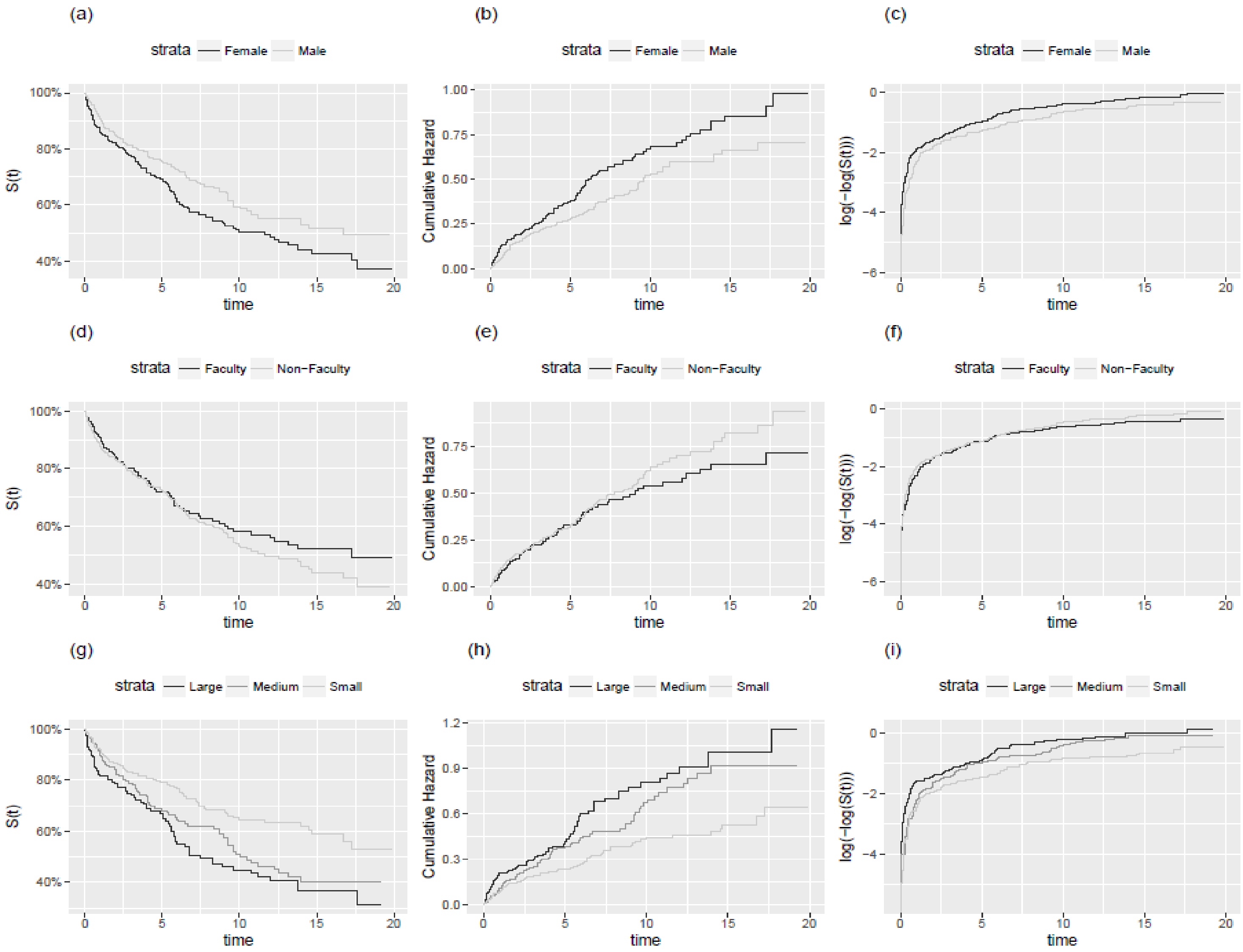

Figure 3: *Estimated survival curves, cumulative hazards and log-log transformed survival curves for categorical covariates. The first row is for Gender, the second row is for Occupation, and the third row is for Size.*

Proportional Hazards Assumption: Martingale Residuals

- Diagnostics for proportional hazards assumption can also be based on martingale residuals.
- First discussed by Lagakos (1981) and later by Barlow and Prentice (1988). Further work was done by Therneau *et al.* (1990).
- The martingale residual process is defined as

$$\hat{M}_i(t) = N_i(t) - \int_0^t Y_i(s) \exp\left\{\hat{\beta}' X_i(s)\right\} d\hat{\Lambda}_0(s), \quad i = 1, \ldots, n,$$

where $\hat{\Lambda}_0$ is the estimated cumulative baseline hazard, which can be obtained using the method of Breslow (1974) as

$$\hat{\Lambda}_0(t) = \sum_{i=1}^{n} \int_0^t \frac{dN_i(s)}{\sum_{j=1}^{n} Y_j(s) \exp\left\{\hat{\beta}' X_j\right\}}.$$

Proportional Hazards Assumption: Martingale Residuals

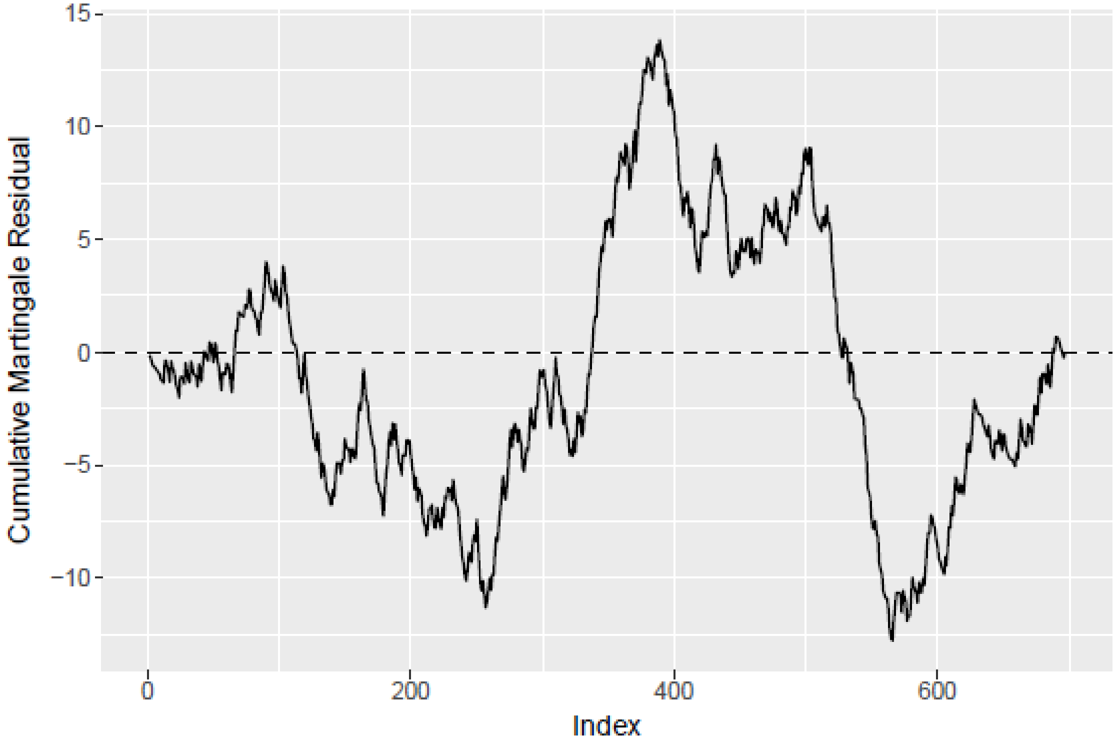

Figure 4: *Cumulative sum of martingale residuals of Model 2, ordered by Age.*

Functional Forms

- Martingale residuals play an important role in functional form diagnostics.
- Therneau *et al.* (1990) discussed the usage of martingale residuals in investigating the functional form of covariates.
- To examine a particular covariate, fit a proportional hazards model omitting that covariate and computing the martingale residuals.
- A smoothed plot versus the omitted covariate often gives approximately the correct functional form of the covariate (e.g., linear, quadratic)
- It requires the covariate of interest to be uncorrelated with other covariates in the model.

Functional Forms

- Age is the only time continuous covariate whose form needs to be assessed.
- Fit a model excluding Age and obtain its martingale residuals.

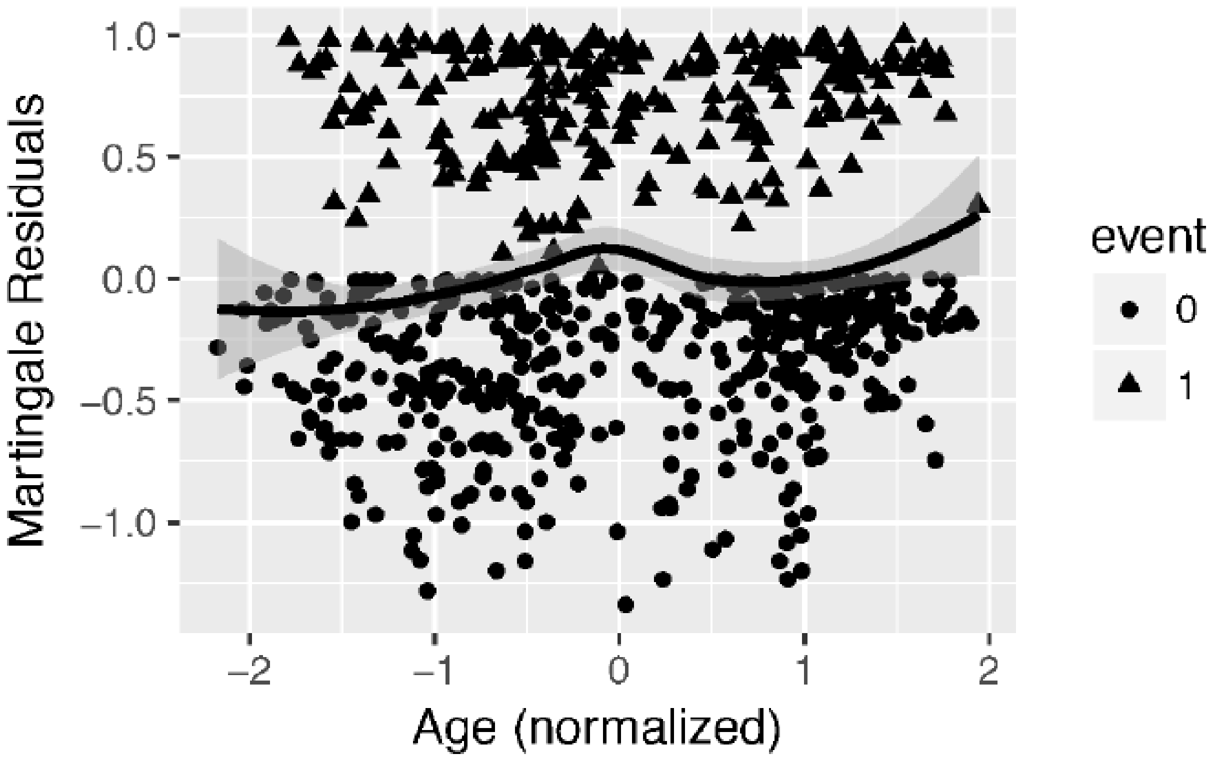

Figure 1: *Plot of martingale residual of the model excluding Age against Age.*

Outlying Observations

- Therneau *et al.* (1990) introduced the deviance residual for a Cox model:

$$d_i = \operatorname{sgn}\left(\hat{M}_i\right)\left[-2\left\{\hat{M}_i + \delta_i \log\left(\delta_i - \hat{M}_i\right)\right\}\right]^{\frac{1}{2}},$$

- From the functional form it is apparent that the deviance residual is essentially a transformation of the martingale residual.
- Plotting *di* against $\widehat{\beta}' X_i$ or $\exp(\widehat{\beta}' X_i)$ will help identify potential outliers which have deviance residuals with too large absolute values.
- Nardi and Schemper (1999) proposed two new types of residuals:

1. The log-odds residual $L_i = \log\left[S_i(T_i)/\{1 - S_i(T_i)\}\right]$
2. Normal deviate residual $\eta_i = \Phi^{-1}\{S_i(T_i)\}, i = 1, \ldots, n,$

Outlying Observations: Martingale and Deviance Residuals

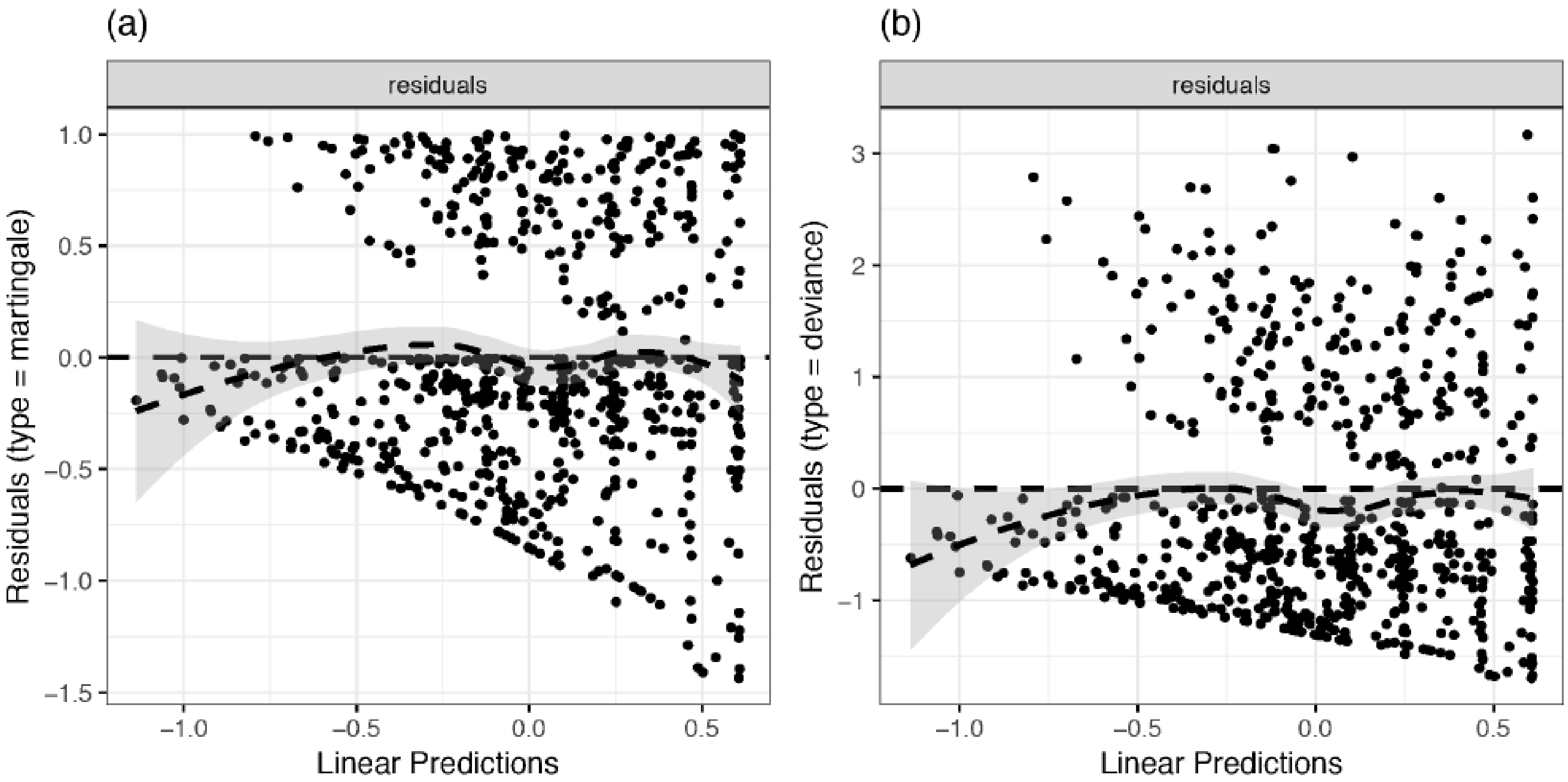

Figure 5: *Plot of (a) martingale and (b) deviance residuals of Model 2.*

Outlying Observations: Log-Odds and Normal Deviate Residuals

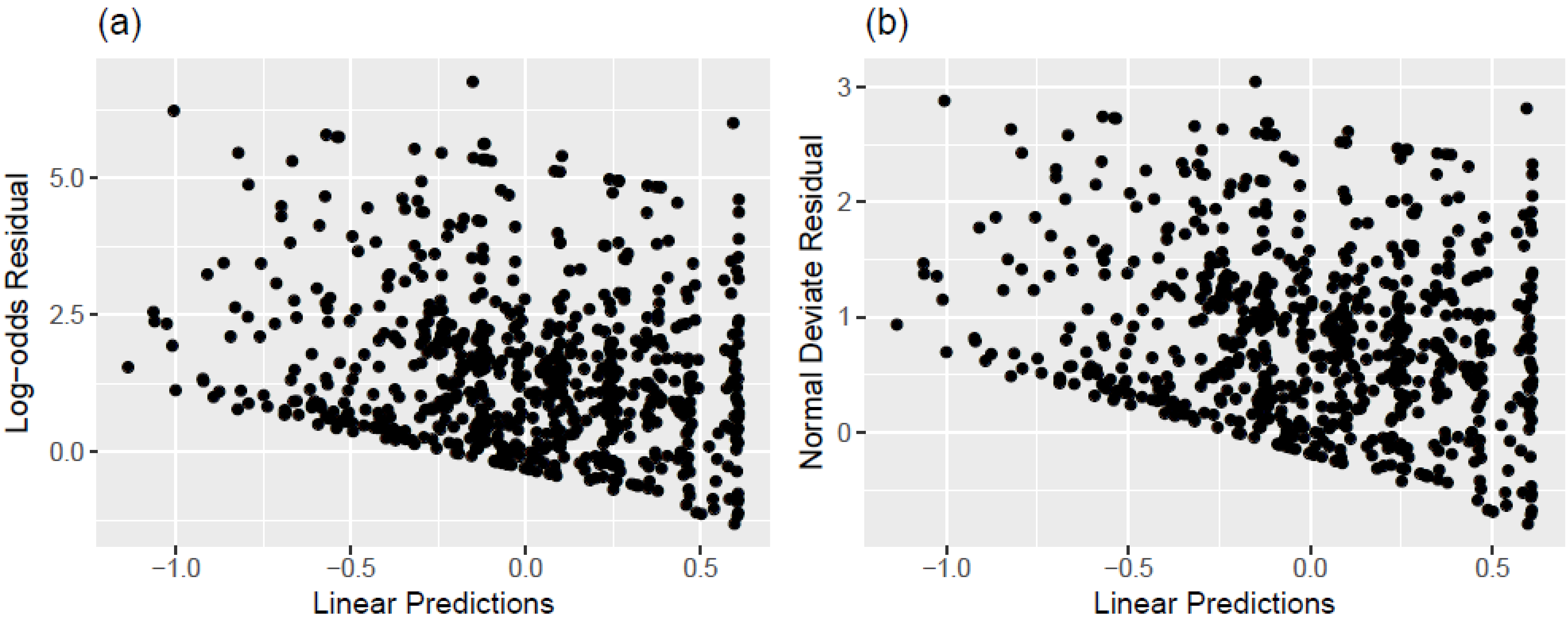

Figure 6: *Plot of (a) log-odds and (b) normal deviate residuals of Model 2.*

Influential Observations

- In studying the influence of one observation, a general practice is to delete that observation, fit the model again, and compare the parameter estimates with those of the model fit on the complete data.
- The Cox model is conceptually different from linear or generalized linear models in that it involves both parametric and nonparametric estimation. Therefore, an observation could be influential in terms of more than just regression coefficients.

Influential Observations

- The matrix of dfbeta residuals

$$D = r_U \mathcal{I}^{-1}\left(\hat{\beta}\right).$$

- r_U *is* the $n \times p$ matrix with the *ith* row being the vector of score residuals for observation *I*

$$r_{Ui}\left(\hat{\beta}\right) = \int_0^\infty \left[X_i - \bar{x}\left(\hat{\beta}, s\right)\right] d\hat{M}_i(s), \quad i = 1, \ldots, n,$$

- Likelihood displacement, proposed by Cook (1986)

$$\mathrm{LD}(w) = 2\left[\ell\left(\hat{\beta}\right) - \ell\left(\hat{\beta}(w)\right)\right],$$

Influential Observations: Dfbeta Residuals

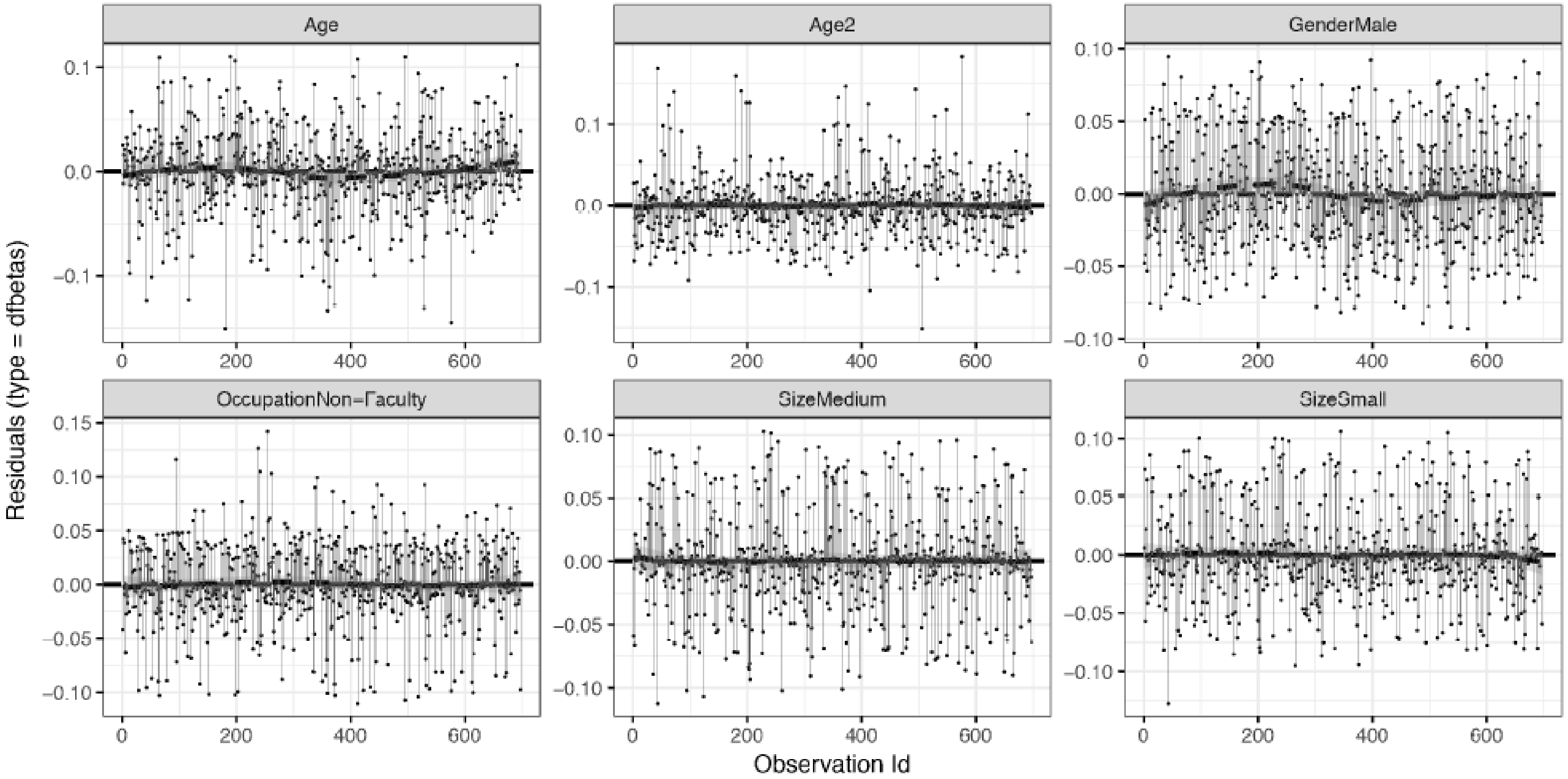

Figure 7: *Dfbetas residuals for covariates of Model 2.*

Influential Observations: Likelihood Displacement

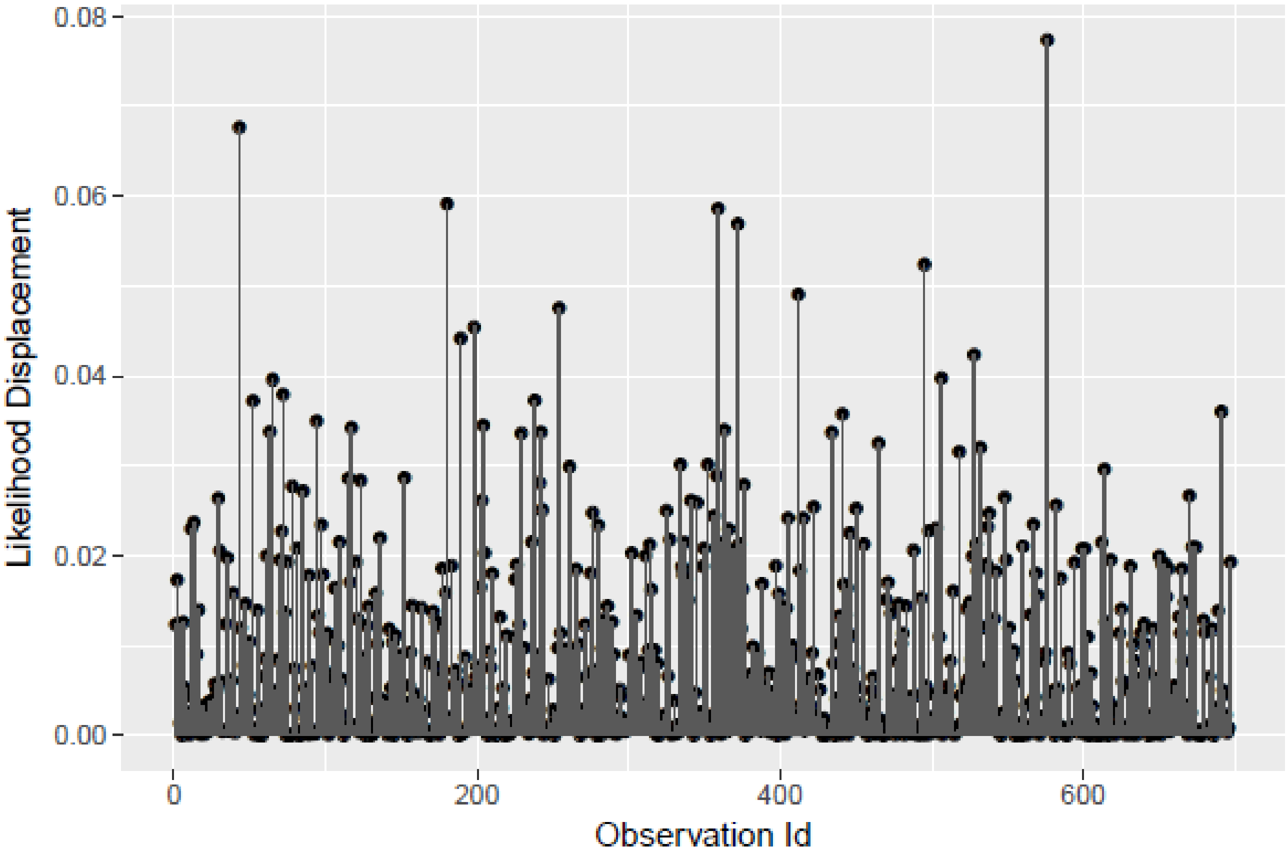

Figure 8: *Likelihood displacement caused by dropping each observation.*

Homework Question (choose 1 of the 2)

1. Write down the Schoenfeld residuals formula and briefly explain how it works. (You can take the assumption that there are no tied event times for simplicity, review lecture note #6 for the answer.)
2. Describe the idea that Therneau et al. (1990) proposed to use martingale residuals in investigating the functional form of continuous covariate.

Reference

- Cox DR (1972). Regression models and life-tables (with discussion), *Journal of the Royal Statistical Society Series B (Methodological)*, 34, 187–220.
- Cox DR (1975). Partial likelihood, *Biometrika*, 62, 269–276.
- Cox DR and Snell EJ (1968). A general definition of residuals, *Journal of the Royal Statistical Society Series B (Methodological)*, 30, 248–275.
- Fleming TR and Harrington DP (1991). *Counting Processes and Survival Analysis*, John Wiley & Sons, New York.
- Grambsch PM and Therneau TM (1994). Proportional hazards tests and diagnostics based on weighted residuals, *Biometrika*, 81, 515–526.Nardi A and Schemper M (1999). New residuals for Cox regression and their application to outlier screening, *Biometrics*, 55, 523–529
- Schoenfeld D (1980). Chi-squared goodness-of-fit tests for the proportional hazards regression model, *Biometrika*, 67, 145–153.
- Schoenfeld D (1982). Partial residuals for the proportional hazards regression model, *Biometrika*, 69,239–241.tics based on weighted residuals, *Biometrika*, 81, 515–526.
- Storer BE and Crowley J (1985). A diagnostic for Cox regression and general conditional likelihoods, *Journal of the American Statistical Association*, 80, 139–147.
- Therneau TM, Grambsch PM, and Fleming TR (1990). Martingale-based residuals for survival models, *Biometrika*, 77, 147–160