Chapter 6

Assessment of Model Adequacy

Model Assessment

- How can we judge the adequacy or goodness of fit of the Cox proportional hazards model?
- In the linear regression setting, many model assessments are performed using residuals.
  - The residual is the difference between the observed outcome and the outcome predicted by the model:

$$\hat{e}_i = y_i - \hat{y}_i$$

where $\hat{y}_i = \hat{\alpha} + \hat{\beta} x_i$.

- But how can we define a similar residual for a Cox model? What is “observed”? What is “expected”?
  - There are several possible residuals for a Cox model.

Model Assessment

- Two broad classes of residuals:
  - Individual-wise residuals: each subject has one.
    - Martingale residuals
    - Deviance residuals *
  - Covariate-wise residuals: each subject has several, one for each covariate.
    - Schoenfeld and scaled Schoenfeld residuals
    - Score and scaled score residuals

* Not discussed in textbook

Individual-wise Residuals

Martingale Residuals

- The derivation of the martingale residual comes from counting process theory.
  - Let N(t) be the number of events that have been observed up to time *t*
  - H(t) is the cumulative hazard function and can be thought of as the expected number of events up to time *t*
    - As the cumulative hazard increases, we expect to see more events occur.

- A residual-type quantity could therefore be

$$(\#\text{ observed events}) - \left(\begin{matrix}\#\text{ expected events according to}\\ \text{cumulative hazard}\end{matrix}\right) = N(t) - H(t)$$

  - The martingale residual is M(t) = N(t) – H(t)

Martingale Residuals

- How can we estimate a martingale residual for each subject?
  - For a particular individual, let *c* be the censoring indicator
    - c = 1 if the individual had the event and c = 0 if censored
    - c_j is the <u>observed</u> number of events for the j^{th} person by time t_j.
  - $\widehat{H}(t_j|\hat{\beta}, x_j)$ is our model-based estimate of the cumulative hazard function for the j^{th} subject at time t_j.
    - $\widehat{H}(t_j|\hat{\beta}, x_j)$ is an estimate of the <u>expected</u> value of c_j by time t_j.

- The martingale residual is therefore

$$\widehat{M}_j = c_j - \widehat{H}(t_j|\hat{\beta}, x_j)$$

Martingale Residuals

- Like in standard linear regression, the martingale residuals will sum to zero.

- Under the correct model, the martingale residuals are uncorrelated and have mean zero.

- Interpretation:
  - $\widehat{M}_j < 0$: subject survived longer than expected
  - $\widehat{M}_j > 0$: subject had the event sooner than expected

Deviance Residuals

- Deviance residuals are found by transforming the martingale residuals:

$$\widehat{D}_j = sign(\widehat{M}_j) \cdot \sqrt{-2\left(\widehat{M}_j + c_j \log(c_j - \widehat{M}_j)\right)}$$

- Deviance residuals can be somewhat easier to interpret than martingale residuals
  - Martingale residuals have a heavily skewed distribution.
  - Range for martingale residuals: $-\infty < M_j \leq 1$
  - Deviance residuals are more symmetric around zero, and are approximately normally distributed.

Covariate-wise Residuals

Schoenfeld Residuals

- The Schoenfeld residual is a covariate-wise residual, so each subject will have a residual for each of the *p* covariates in the model.

$$\hat{r}_i{}^T = \left(\hat{r}_{i1}, \hat{r}_{i2}, \ldots \hat{r}_{ip}\right)$$

- Schoenfeld residuals derived by taking the partial derivative of the log of the partial likelihood:

$$\hat{r}_{ik} = c_i \left(x_{ik} - \frac{\sum_{j \in R(t_i)} x_{jk} e^{\beta x_j}}{\sum_{j \in R(t_i)} e^{\beta x_j}} \right)$$

Schoenfeld Residuals

- Schoenfeld residuals only defined for subjects with observed event times.
  - $\hat{r}_i$ is missing if subject *i* was censored
- Schoenfeld residuals will sum to zero.
- Why is this quantity considered to be a residual?
  - The Schoenfeld residual is the difference between the actual covariate value and the expected covariate value

Scaled Schoenfeld Residuals

- Can also scale the Schoenfeld residuals by their variance to aid in interpretation.

$$\hat{r}_i^* = \left(\widehat{Var}(\hat{r}_i)\right)^{-1} \cdot \hat{r}_i$$

where $\left(\widehat{Var}(\hat{r}_i)\right)^{-1} = m\widehat{Var}(\hat{\beta})$ and *m* is the number of observed events.

- Scaled and unscaled Schoenfeld residuals used in the same way, but there is some evidence that scaled Schoenfeld residuals are a better diagnostic

Score Residuals

- The score residuals are also derived using counting process theory and are a function of the martingale residuals:

$$L_{ik} = \sum_{j=1}^{n} \left(x_{ik} - \underbrace{\frac{\sum_{h \in R(t_j)} x_{hk} e^{\beta x_h}}{\sum_{h \in R(t_j)} e^{\beta x_h}}}_{\text{weighted average of covariate values for subjects still at risk at time } t_j} \right) \cdot dM_i(t_j)$$

Change in martingale residual at time t_j

Score Residuals

- $\hat{L}_i{}^T = (\hat{L}_{i1}, \hat{L}_{i2}, \dots, \hat{L}_{ip})$ is the vector of estimated score residuals.

- Score residuals are a measure of the difference between the actual covariate value and the expected covariate value, weighted by the martingale residual

- Score residuals are defined for both censored and non-censored subjects.

Scaled Score Residuals

- Can also scale the score residuals

$$\hat{L}_i^* = \widehat{Var}(\hat{\beta}) \cdot \hat{L}_i$$

Steps of Model Assessment

1. Check whether model assumptions are reasonable.
2. Identify any poorly fit or overly influential subjects
3. Compute summary measures of goodness of fit

Steps of Model Assessment

1. Check whether model assumptions are reasonable.
   - Proportional hazards assumption
   - Linearity assumption

Checking for Proportional Hazards

- Proportional hazards assumption
  - The hazard for each subject is proportional to the baseline hazard, $h_0(t)$.
    - $h(t \mid \beta, x) = h_0(t) \cdot e^{\beta x}$
  - When we assume proportional hazards, we are assuming that the effect of a covariate is constant over time.
    - $HR(t| x = a \text{ vs } x = b) = e^{\beta(a-b)}$
  - In other words, if the proportional hazards assumption is reasonable, there should be no significant interaction between the effect of the covariate and time.

Checking for Proportional Hazards

- To test the proportional hazards assumption for a covariate *x*:
  - Fit the Cox model including an interaction term between x and time

$$\log\big(h(t|\beta, x)\big) = \log(h_0(t)) + \beta x + \gamma x t$$

  - Test the null hypothesis H_0: $\gamma = 0$ vs H_1: $\gamma \neq 0$.
    - If the null hypothesis is rejected, the proportional hazards assumption is not reasonable for this covariate.
  - Note: *xt* is a time-varying covariate. Its value must be recalculated at each observed survival time for each subject still at risk (more on this in Chapter 7).

Checking for Proportional Hazards

- Schoenfeld residuals can also be used to test the proportional hazards assumption.
  - Plot the Schoenfeld residuals for a particular covariate against time
  - If the residuals do not look random, then the proportional hazards assumption may not be appropriate for that covariate.
- Can also calculate the correlation between the Schoenfeld residuals and survival time. Correlation should be nonsignificant.

- May want to consider multiple functions of survival time (e.g., log, square) as well as raw survival time.

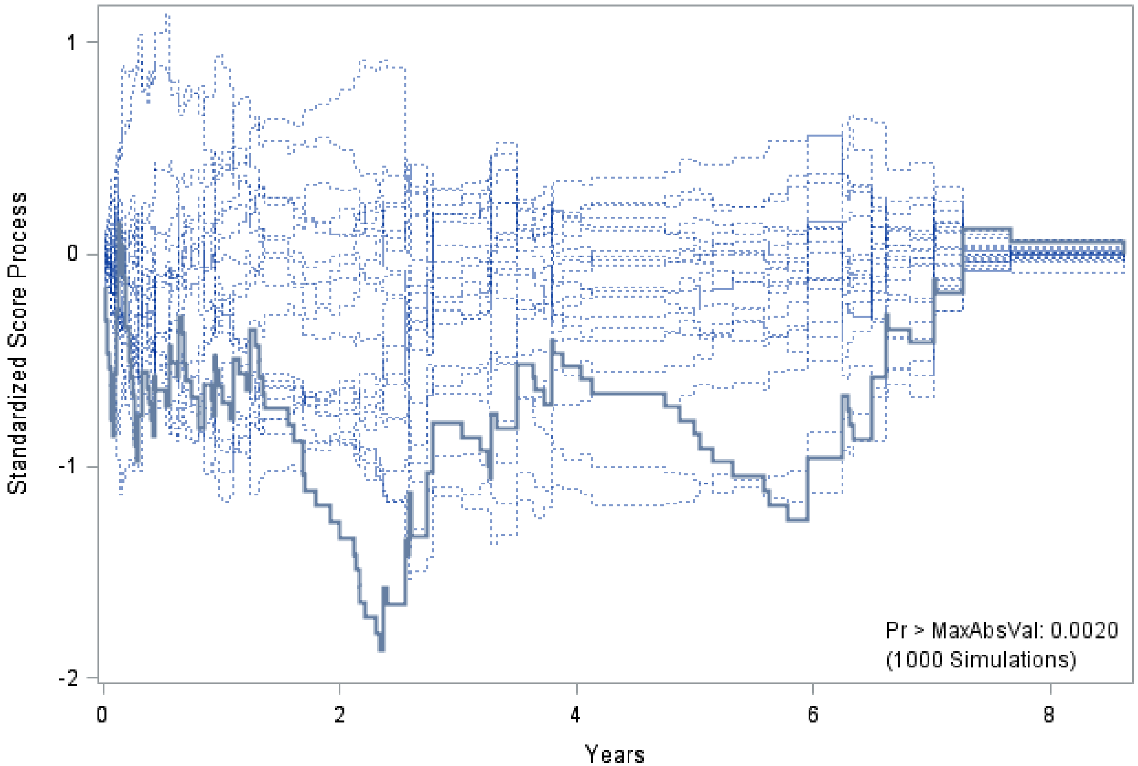
Checking for Proportional Hazards

- A plot of the empirical score process can also be used to check the proportional hazards assumption.
  - The empirical score process is derived from the cumulative sums of the martingale residuals.
- Tests the null hypothesis of proportional hazards
  - A large number (1000 or more) of processes are generated under the null hypothesis.
  - The observed empirical score process is compared to these simulated processes using a Kolmogorov-type supremum test.

Checking for Proportional Hazards

- If the observed empirical score process is not similar to the simulated processes, then the proportional hazards assumption may not be reasonable.

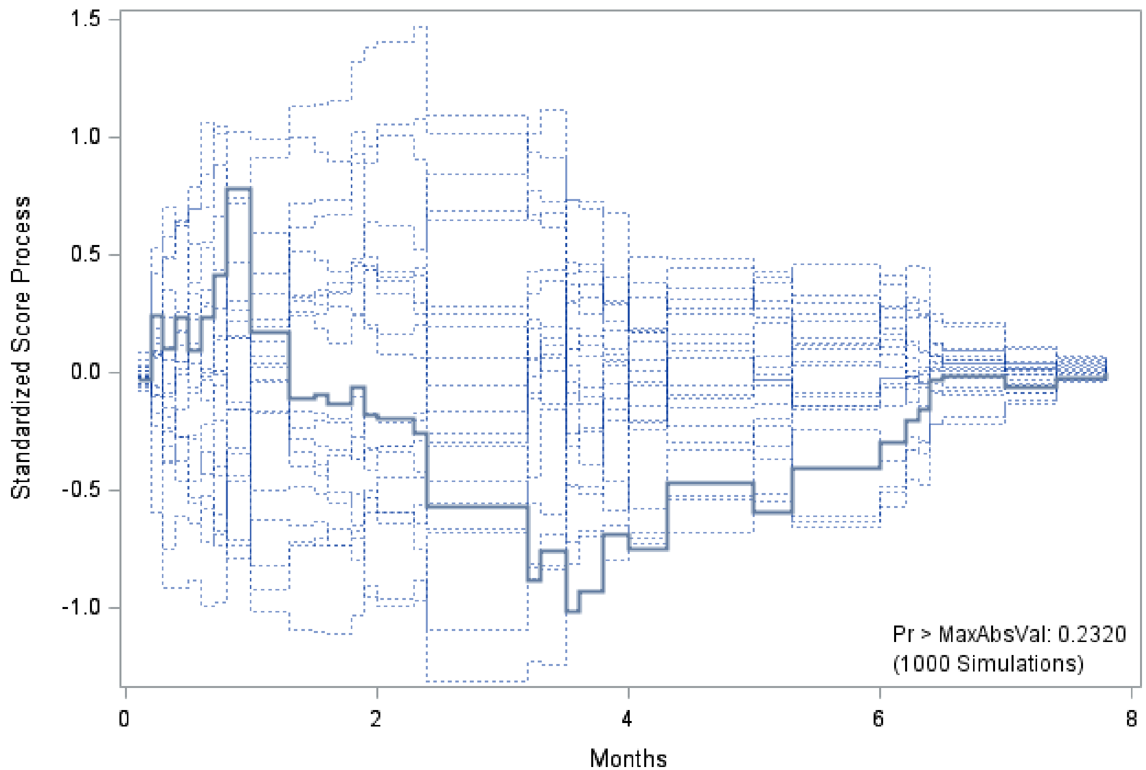
Proportional hazards assumption may not reasonable



Checking for Proportional Hazards

- If the observed empirical score process is not similar to the simulated processes, then the proportional hazards assumption may not be reasonable.

No evidence that proportional hazards assumption is not reasonable



Checking for Proportional Hazards

- Recall: for the Cox proportional hazards model,

$$S(t|\beta, x) = [S_0(t)]^{e^{\beta x}}$$

- Rearranging this equation, we get:

$$log(-\log S(t|\beta, x)) = \beta x + log(-\log S_0(t))$$

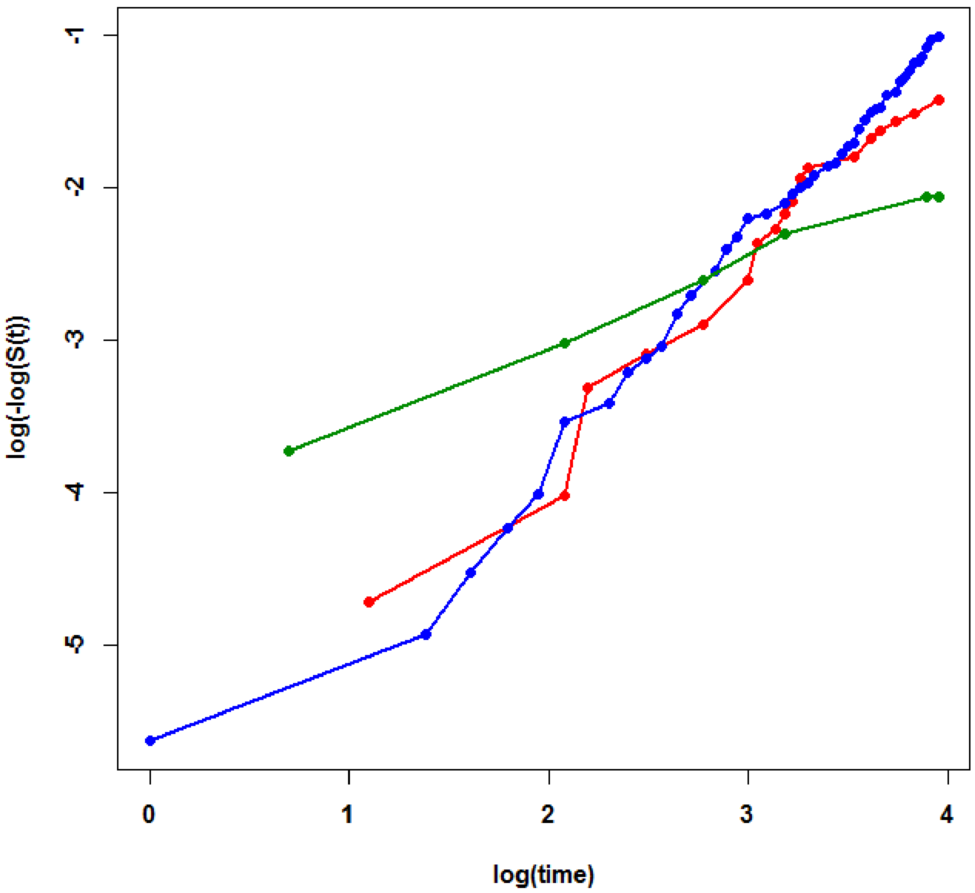
or

$$log\left(H(t|\beta, x)\right) = \beta x + log(H_0(t))$$

- So, if the proportional hazards assumption is reasonable, the log cumulative hazards for different levels of a covariate should be parallel.

Checking for Proportional Hazards

- To check the proportional hazards assumption for a categorical covariate, estimate the survival function for each level using Kaplan-Meier.
- Plot $log\left(-\log \hat{S}_i(t)\right)$ against t or $\log t$.
  - It's standard to use $\log t$.
- If lines are not parallel, the proportional hazards assumption may not be reasonable.



Checking the Linearity Assumption

- Linearity assumption
  - There is a linear relationship between the covariates and the log hazard function.
    - $\log h(t \mid \beta, x) = \log h_0(t) + \beta x$

Checking the Linearity Assumption

- Martingale residuals can be used to check the linearity assumption for continuous covariates.
  - Fit the model without the covariate
  - Plot the martingale residuals against the covariate
  - There should appear to be a linear relationship between the residuals and the covariate; otherwise, the linearity assumption may not be appropriate.
- Deviance residuals can also be used in the same way to check the linearity assumption.

Steps of Model Assessment

2. Identify any poorly fit or influential subjects
   - Outliers
   - Leverage
   - Influence

Identifying poorly fit subjects

- An outlier is an observation whose survival time is unusual for a person with those covariates.

- Can use deviance residuals to check for outliers
  - Very large ($\widehat{D}_j > 3$) or very small ($\widehat{D}_j < -3$) may indicate that an observation is an outlier.

Identifying poorly fit subjects

- Leverage is a model diagnostic that measures how unusual the covariate values are for an individual.

- Score residuals can be used to determine which observations have leverage.
  - Large positive or negative score residuals indicate that an observation has high leverage.
  - Plot the score residuals against the covariate and look for points that don’t “fit in”.

Identifying influential subjects

- Influence is a model diagnostic that refers to how much the results of a model depend on a single observation.

- To determine the influence of an observation:
  - Fit the model using the full data set ($\hat{\beta}$)
  - Delete the ith observation and refit the model, call this estimate $\hat{\beta}_{(-i)}$
  - If $\hat{\beta}$ - $\hat{\beta}_{(-i)}$ is close to zero, the observation has little influence

Identifying influential observations

- The change in parameter estimate, $\hat{\beta}$ - $\hat{\beta}_{(-i)}$, is approximately equal to the vector of scaled score residuals for that subject

- So, scaled score residuals can be used to determine which observations are influential.
  - Plot the scaled score residuals against the covariate and look for points that don't "fit in".

Other Diagnostic Measures

- Likelihood displacement is another measure of influence.
  - The likelihood displacement for an observation is the change in partial likelihood when that observation is deleted.
  - $\widehat{LD}_i = 2 \cdot logL_p\left(\hat{\beta}\right) - 2 \cdot logL_p(\hat{\beta}_{(-i)})$
- Influential observations will have large values of $\widehat{LD}_i$.
- Plot likelihood displacement against the martingale residuals.
  - Look for extreme values along the sides of the plot.

Steps of Model Assessment

3. Compute summary measures of goodness of fit
   - For Cox proportional hazards models, there is no exact analogue to R^2 in linear regression (the proportion of variation explained).
   - One proposed measure:

$$R_p^2 = 1 - e^{\frac{2\left(\log L_0 - \log L_p\right)}{n}}$$

where L_p is the partial likelihood for the model and L_0 is the partial likelihood for a model that does not include any covariates.

   - Caveat: having a large number of censored observations can lead to small R_p^2, even if the model fits the data well.
     - For this reason, R_p^2 is rarely used.

Summary of Residuals and Diagnostics

Name	Type	Use	SAS keyword	Example
Martingale	Individual	Test linearity assumption	resmart	Figure 5.2a (p. 149)
Deviance	Individual	Test linearity assumption Check for outliers	resdev	(Not discussed in book)
Schoenfeld	Covariate	Test proportional hazards assumption	ressch	(None; Book recommends using scaled Schoenfeld)
Scaled Schoenfeld	Covariate	Test proportional hazards assumption	wtsch	Figure 6.1 (p. 181) Figure 6.2 (p. 182)
Score	Covariate	Assess leverage	ressco	Figure 6.4 (p. 186) Figure 6.5 (p. 186)
Scaled Score	Covariate	Assess influence	dfbeta	Figure 6.6 (p. 187) Figure 6.7 (p. 188)
Likelihood Displacement	Individual	Assess influence	ld	Figure 6.8 (p. 190)