Evaluating the Proportional Hazards Assumption*

Gebrenegus Ghilagaber

March 4, 2008

1 Proportional Hazards (PH) Model

To make matters simple, consider a Cox PH model with just one fixed covariate:

$$
\begin{array}{rcll}
h(t|x) & = & h_0(t) \exp(\beta x) & \\
 & & \quad h_0(t) & * \quad \exp(\beta x) \\
 & = & \quad \text{Baseline hazard} & * \quad \text{Relative Hazard} \\
 & & \quad (\text{Function of } t \text{ but not of } x) & \quad\;\; (\text{Function of } x \text{ but not of } t)
\end{array}
$$

In other words, the Cox PH model specifies the hazard at time t for an indiviudal with covariate x as a product of a baseline hazard (which is a

*Most of the material in this section is based on Chapter 4 of D. G. Kleinbaum (1996), "Survival Analysis - a self learning text" function of t) and a relative hazard (which is a constant - independent of the time t). Suppose, the covariate x is binary taking values 0 and 1. Then,

$$
\begin{aligned}
h(t|x &= 0) = h_0(t)\,\mathrm{exp}(\beta * 0) \\
&= h_0(t)
\end{aligned}
$$

while

$$
\begin{aligned}
h(t|x &= 1) = h_0(t)\,\mathrm{exp}(\beta * 1) \\
&= h_0(t)\,\mathrm{exp}(\beta) \\
&= h(t|x = 0) * \mathrm{exp}(\beta)
\end{aligned}
$$

and, thus

$$
\frac{h(t|x = 1)}{h(t|x = 0)} = \mathrm{exp}(\beta) = \text{cons tan t over } t.
$$

Since the hazard $h(t|x = 1)$ can be expressed as a product of the other hazard, $h(t|x = 0)$, times a constant the two hazards are proportional to each other - that is the assumption in the Cox PH model. Under this assumption (if the hazards are proportional), the corresponding estimated survivor functions are given by:

$$\widehat{S}(t|x = 1) = \left[\widehat{S}_0(t)\right]^{\exp(\beta)} = \left[\widehat{S}_0(t)\right]^{e^{\beta}} \tag{1}$$

where

$$\widehat{S}_0(t) = \widehat{S}(t|x = 0)$$

is the baseline survivor function.

2 Checking the Proportional Hazards Assumption

How do we test if the PH assumption is violated? Three approaches are outlined below:

2.1 Graphical Approach

2.1.1 Plots of estimated hazards

Plot the estimated hazard functions of the different levels of the covariate and examine if they are proportional (don't cross)

2.1.2 Log-Log Plots of Estimated Survivor Functions, $\widehat{S}(t)$

- Plot $-ln(-ln(\widehat{S}(t)))$ for each level of the covariate of interest.

- Parallel curves indicate that PH assumption is satisfied since in that case (denoting $\widehat{S}(t|x=1) = \widehat{S}_1(t)$

$$
\begin{aligned}
\widehat{S}_1(t) &= \left[\widehat{S}_0(t)\right]^{e^{\beta}} && (2)\\
\Longrightarrow &\ \ \ln\left[\widehat{S}_1(t)\right] = \ln\left(\left[\widehat{S}_0(t)\right]\right) = e^{\beta}\ln\left[\widehat{S}_0(t)\right] && (3)\\
\Longrightarrow &\ \ -\ln\left\{-\ln\left[\widehat{S}_1(t)\right]\right\} = -\ln\left\{-e^{\beta}\ln\left[\widehat{S}_0(t)\right]\right\} && (4)\\
&= \beta - \ln\left\{-\ln\left[\widehat{S}_0(t)\right]\right\} && (5)
\end{aligned}
$$

Thus, under the PH assumption the two curves, $-\ln\left\{-\ln\left[\widehat{S}_1(t)\right]\right\}$ and $-\ln\left\{-\ln\left[\widehat{S}_0(t)\right]\right\}$, should be parallel (separated by a constant, β) throughout the time (for all values of t).

2.1.3 Observed and Expected Survivor curves

- An alternative graphical approach is to compare observed and predicted survivor curves, where

- Observed curves are derived for levels of the covariate under examination without putting it in a PH model, while

- Predicted curves are derived with the covariate in a PH model.

- If observed and predicted curves are close, the the PH assumption is satisfied.

2.1.4 Drawback of the graphical methods:

- Subjective (how parallel is parallel?)

2.2 Time-dependent variables

- When time-dependent variables are used to assess the PH assumption for a time-independent variable, the ordinary Cox PH model is extended to contain a product term (interaction term) involving the time-independent variable being assessed and some function of time.

- Thus, if we want to test if GNEDER fulfills the PH assumption, then we may extend the usual model:

$$h(t|x) = h_0(t)\exp(\beta_1 x_1)$$

where $x_1 =$ GENDER to include an intercation term between GENDER and TIME:

$$h(t|x) = h_0(t)\exp(\beta_1 x_1 + \beta_2 x_2)$$

where $x_2 =$ GENDER*TIME or GENDER*log(TIME), etc

- If β_2 is significant, then the PH assumption is violated for the variable in question (GENDER).

2.3 Goodness-of-fit-test

3 Examples