30. August 2010

**Parts of Chapters 5 & 7**

*Stratified and Piecewise Cox Ph Models*

**Example 1:** Recall the Kleinbaum **crossing hazards** figure:
In this study that compares surgery to no surgery, we might expect to see hazard functions for each group as follows:

![Hazard Functions](image)

Before 2 days, \( \text{HR}(1|0) < 1 \), whereas later, \( \text{HR}(1|0) > 1 \). The PH assumption is violated, since HR must be constant over the follow-up time.
Example 2: Crossing survivor curves:
VA Cooperative Trial No. 345
This was a prospective randomized study conducted between March 1992 and August 1994. Patients were randomly assigned to either unsupplemented general anesthesia and postoperative analgesia (USGA) or epidural plus light general anesthesia and postoperative epidural morphine (ESGA). procedures.

A researcher began a retrospective look ca. June 2003

- It is well established that the epidural protects certain aspects of the immune function, and to block the stress response to surgical trauma.

- The epidural protocol has been common practice.

- Therefore, it was hypothesized that cancer surgery patients should benefit from ESGA. The research hypothesis is depicted in the following graph:
In this example we study thesubset of patients in the VA trial who had had surgery for colon cancer. Of the 247 patients identified in that study, we have survival data on 246:

<table>
<thead>
<tr>
<th></th>
<th>ESGA</th>
<th>USGA</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>METAST</td>
<td>42</td>
<td>48</td>
<td>90</td>
</tr>
<tr>
<td>NO MET</td>
<td>79</td>
<td>77</td>
<td>156</td>
</tr>
<tr>
<td></td>
<td>121</td>
<td>125</td>
<td>246</td>
</tr>
</tbody>
</table>
What time reveals for the No MET group:

Return to the 3.5-year mark:
The Cox PH structure imposes restrictions on the behavior of survivor curves.

- With just one exposure variable $x = 0, 1$, the relationship is
  \[ h(t|1) = h(t|0) \cdot \exp(\beta). \]

- Let TRT = 0 if ESGA, 1 if USGA. Then
  \[ S(t|1) = (S(t|0))^\exp(\beta). \]

  Cannot possibly model the crossing curves.

- Consider the results from the Cox PH fit to No Met Data only

  \begin{verbatim}
  > coxph(Surv(TIME,CENSOR)~TRT
    coef  exp(coef) se(coef)     z    p
   TRT -0.42  0.657  0.211 -1.99 0.046
  Likelihood ratio test=4.03 on 1 df, p=0.0447 n=156
  \end{verbatim}
Scaled Schoenfeld residual plot, and the Grambsch-Therneau (1994) test for PH assumption. The residual plot clearly displays that TRT varies with time.

> PH.test
  rho  chisq     p
TRT  -0.174  2.86  0.0906
Example 3: Divergent survivor curves
Australian study of heroin addicts, Caplehorn, et al. (1991)

- two methadone treatment clinics
- $T = \text{days remaining in treatment}$
  ($= \text{days until drop out of clinic}$)
- clinics differ in overall treatment policies
- 238 patients in the study

Description of ADDICTS data set

<table>
<thead>
<tr>
<th>Data set: ADDICTS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Column 1: Subject ID</td>
</tr>
<tr>
<td>Column 2: Clinic (1 or 2) ← exposure variable</td>
</tr>
<tr>
<td>Column 3: Survival status</td>
</tr>
<tr>
<td>0 = censored</td>
</tr>
<tr>
<td>1 = departed clinic</td>
</tr>
<tr>
<td>Column 4: Survival time in days</td>
</tr>
<tr>
<td>Column 5: Prison record ← covariate</td>
</tr>
<tr>
<td>0 = none, 1 = any</td>
</tr>
<tr>
<td>Column 6: Maximum methadone dose (mg/day) ← covariate</td>
</tr>
</tbody>
</table>
Part I: The following is R code, along with modified output, that fits two K-M curves not adjusted for any covariates to the survival data.

```r
> addict.fit <- survfit(Surv(Days.survival,Status)~Clinic, data = ADDICTS)
> addict.fit
   n  events mean   se(mean) median 0.95LCL 0.95UCL
Clinic=1 163    122 432   22.4    428   348   514
Clinic=2  75     28 732   50.5     NA   661     NA
> survdiff(Surv(Days.survival,Status)~Clinic,data = ADDICTS)
   N Observed  Expected (O-E)^2/E (O-E)^2/V
Clinic=1 163    122   90.9     10.6    27.9
Clinic=2  75     28   59.1     16.4    27.9
Chisq= 27.9 on 1 degrees of freedom, p= 1.28e-007
> plot(addict.fit, lwd = 3,col = 1,type = "l",lty=c(1, 3), cex=2,lab=c(10,10,7),...)
```

![K-M curves for ADDICTS not adjusted for covariates.](image)

Figure 1. K-M curves for ADDICTS not adjusted for covariates.
Results:

- The log-rank test is highly significant with $p$-value $= 1.28 \times 10^{-7}$.

- The graph in Figure 1 glaringly confirms this difference.

- This graph shows curve for clinic 2 is always above curve for clinic 1.

- Curves diverge, with clinic 2 being dramatically better after about one year in retention of patients in its treatment program.

- Lastly, this suggests the PH assumption is not satisfied.
Part II: The Cox PH model We now fit a Cox PH model which adjusts for the three predictor variables. This hazard model is

\[ h(t|\mathbf{x}) = h_0 \exp(\beta_1 \text{Clinic} + \beta_2 \text{Prison} + \beta_3 \text{Dose}). \]

A summary of the R output is:

```r
> fit1 <- coxph(Surv(Days.survival,Status) ~ Clinic+Prison+Dose,data = ADDICTS,x = T) # Fits a Cox PH model
> fit1

  coef     exp(coef)   se(coef)   z      p
Clinic -1.0098   0.364 0.21488 -4.70 2.6e-006
Prison  0.3265   1.386 0.16722  1.95 5.1e-002
Dose   -0.0354   0.965 0.00638 -5.54 2.9e-008

Likelihood ratio test=64.6 on 3 df, p=6.23e-014 n= 238
> testph <- cox.zph(fit1) # Tests the proportional hazards assumption
> print(testph) # Prints the results

   rho  chisq      p
Clinic -0.2578 11.19 0.000824
Prison -0.0382  0.22 0.639324
Dose   0.0724  0.70 0.402755
GLOBAL NA 12.62 0.005546
> par(mfrow = c(2, 2))
> plot(testph) # Plots the scaled Schoenfeld residuals.
```
Figure 2. Diagnostic plots of the constancy of the coefficients in the fit1 model. Each plot is of a component of $\hat{\beta}(t)$ against ordered time. A spline smoother is shown, together with $\pm 2$ standard deviation bands.
Results:

• The GLOBAL test (a LRT) for non-PH is highly statistically significant with $p$-value = 0.005546.

• The $p$-values for Prison and Dose are very large, supporting that these variables are time-independent.

• The Grambsch-Therneau test has a $p$-value = 0.000824 for Clinic. This provides strong evidence that the variable Clinic violates the PH assumption and confirms what the graph in Figure 1 suggests.

• The plot of $\hat{\beta}_1(t)$, the coefficient for Clinic, against ordered time in Figure 2 provides further supporting evidence of this violation.

• We recommend finding a function $g(t)$ to multiply Clinic by; that is, create a defined time-dependent variable, and then fit an extended Cox model.
Since the Cox PH model is inappropriate, the following strategies are employed:

- analyze by **stratifying** on the exposure variable; that is, do not fit any regression model, and, instead obtain the Kaplan-Meier curve for each group separately;

- to adjust for other significant factor effects, use Cox model **stratified on exposure variable** $E$.
  
  > coxph(Surv(time,status)~X1+X2+···+strata(E))

- fit a Cox PH model that includes a time-dependent variable which measures the interaction of exposure with time. This model is called an **extended Cox model**. We try to find the point in time $t_0$ where the hazard rates change. Then fit a **piecewise** Cox PH model over these two time intervals.
Part III: Stratified Cox model

Suppose we have \( j = 1, 2, \ldots, s \), i.e., \( s \) strata. For each stratum we assume the Cox PH model:

\[
h_j(t|x) = h_{0j}(t) \exp(x'\beta), \quad j = 1, \ldots, s.
\]

The regression coefficients are assumed to be the same in each stratum although the baseline hazard functions may ne different and completely unrelated. Then using only the data for those subjects in the \( j \)th stratum, we have:

Let \( t_{(1j)}, \ldots, t_{(rj)} \) denote the \( r \leq n_j \) ordered (uncensored) death times, so that \( t_{(kj)} \) is the \( k \)th ordered death time. Let \( x_{(kj)} \) denote the vector of covariates associated with the individual who dies at \( t_{(kj)} \).

Cox’s partial likelihood function for the \( j \)th stratum:

\[
L_{cj}(\beta) = \prod_{kj=1}^{r} \frac{\exp(x'_{(kj)}\beta)}{\sum_{l \in R(t_{(kj)})} \exp(x'_{l}\beta)}.
\]

Then estimation and testing methods are as before, where the partial log likelihood to be maximized is given by

\[
LL_c(\beta) = \sum_{j=1}^{s} LL_{cj}(\beta).
\]
We now stratify on the exposure variable Clinic and fit a Cox PH model to adjust for the two time-independent covariates Prison and Dose. Modified R output and a plot of the two adjusted K-M survival curves follow.

```r
> fit2 <- coxph(Surv(Days.survival,Status) ~ strata(Clinic)+Prison+Dose,data=ADDICTS)
> fit2

               coef exp(coef) se(coef)  z      p
Prison 0.3896  1.476     0.16893 2.31 2.1e-02
Dose -0.0351  0.965     0.00646-5.43 5.6e-08

Likelihood ratio test=33.9 on 2 df, p=4.32e-008  n= 238
> survfit(fit2)

        n  events  mean  se(mean) median  .95LCL  .95UCL
Clinic=1 162    122  434   22.0    434    358    517
Clinic=2  74     28  624   38.1    878    661  NA

# Note that each stratum has one less observation.
# This tells us that the shortest observed retention time in each clinic is censored.
> plot(survfit(fit2),lwd=3,col=1,type="l",lty=c(1,3),
     cex=2,lab=c(10,10,7),...)
> abline(v = 366,lty=3,lwd=2)
```
Results:

- Figure 3 provides same pictorial evidence as Figure 1; that is, curve for clinic 2 is always above clinic 1’s curve, with clinic 2 being dramatically better in retention of patients in its treatment program after about one year.

- The estimated coefficients for Prison and Dose do not change much. This gives good evidence that the stratified model does satisfy the PH assumption; hence, this analysis is valid.

- Figure 3 provides a picture of the effect of Clinic on retention of patients. But by stratifying on Clinic, we get no estimate of its effect; i.e., no estimated
\( \beta_1 \) coefficient. Hence, we cannot obtain a hazard ratio for Clinic.

- The exposure variable Clinic must be in the model in order to obtain a hazard for it. For this reason, we look now to the extended Cox model.
Part IV: A Piecewise Cox PH model analysis

Here we use a model that contains two heavyside functions, \( g_1(t) \) and \( g_2(t) \), with \( t_0 \), the change point, to be determined. The hazard model is

\[
h(t|x(t)) = h_0(t) \exp(\beta_1 \text{Prison} + \beta_2 \text{Dose} + \gamma_1 (\text{Clinic} \times g_1(t)) + \gamma_2 (\text{Clinic} \times g_2(t)))
\]

where

\[
g_1(t) = \begin{cases} 
1 & \text{if } t < t_0 \\
0 & \text{if } t \geq t_0
\end{cases} \\
g_2(t) = \begin{cases} 
1 & \text{if } t \geq t_0 \\
0 & \text{if } t < t_0
\end{cases}
\]

and

\[
\text{Clinic} = \begin{cases} 
1 & \text{if Clinic}=1 \\
0 & \text{if Clinic}=2.
\end{cases}
\]  

The hazard ratio for the exposure variable Clinic now varies with time. It assumes two distinct values depending whether time \(< t_0\) days or time \(\geq t_0\) days. The form of the HR is

\[
t < t_0 : \quad \text{HR} = \exp(\gamma_1)
\]
\[
t \geq t_0 : \quad \text{HR} = \exp(\gamma_2).
\]

Time-dependent covariates effect the rate for upcoming events. In order to implement an extended Cox model properly in R using the \texttt{coxph} function, one must use the Anderson-Gill (1982) formulation of the proportional hazards model as a counting process. They treat
each subject as a very slow Poisson process. A censored subject is not viewed as incomplete, but as one whose event count is still zero. For a brief introduction to the counting process approach, see Appendix 2 of Hosmer & Lemeshow (1999) and the online manual S-PLUS 2000, Guide to Statistics, Vol 2, Chapter 10. Klein & Moeschberger (1997, pages 70–79) discuss this counting process formulation. They devote their Chapter 9 to the topic of modelling time-dependent covariates. For a more advanced and thorough treatment of counting processes in survival analysis, see Fleming and Harrington (1991).

The ADDICTS data set must be reformulated to match the Anderson-Gill notation. To illustrate this, consider the following cases: In both cases the $t$ denotes the patient’s recorded survival time, whether censored or not.

**Case 1:** For $t < t_0$, $g_1(t) = 1$ and $g_2(t) = 0$. Here a patient’s data record is just one row and looks like this:

<table>
<thead>
<tr>
<th>Start</th>
<th>Stop</th>
<th>Status</th>
<th>Dose</th>
<th>Prison</th>
<th>Clinic</th>
<th>Clinic</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$t$</td>
<td>same</td>
<td>same</td>
<td>same</td>
<td>Clinic</td>
<td>0</td>
</tr>
</tbody>
</table>

**Case 2:** For $t \geq t_0$, $g_1(t) = 0$ and $g_2(t) = 1$. Here a patient’s data record is formulated into two rows and looks like this:

<table>
<thead>
<tr>
<th>Start</th>
<th>Stop</th>
<th>Status</th>
<th>Dose</th>
<th>Prison</th>
<th>Clinic</th>
<th>Clinic</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>$t$</td>
<td>same</td>
<td>same</td>
<td>same</td>
<td>Clinic</td>
<td>0</td>
</tr>
</tbody>
</table>
The following R program puts the ADDICTS data set into the counting process form, finds the optimal change point $t_0$; i.e., the time which maximizes the profile log partial likelihood. We then fit the model and report results.

```r
> ADDICTS<-read.table("C://ADDICTS.txt",header=T)
> ADDICTS$Clinic[ADDICTS$Clinic==2]<-0
> names(ADDICTS)
[1] "ID" "Clinic" "Status" "Days.survival"
[5] "Prison" "Dose"
> attach(ADDICTS)
> library(survival)
> optimal.change.point(ADDICTS,Days.survival,Status,Clinic)

<table>
<thead>
<tr>
<th>changepoint</th>
<th>loglik</th>
</tr>
</thead>
<tbody>
<tr>
<td>119</td>
<td>-683.2117</td>
</tr>
</tbody>
</table>

> #Thus, in the survSplit function, let cut = 461.
> #Use the function extcox.1Et to obtain the dataset in the Andersen-Gill counting process format
> AG<-extcox.1Et(ADDICTS,Days.survival,Status,Clinic,461)
> names(AG)
[1] "ID" "Clinic" "Status" "Days.survival"
[5] "Prison" "Dose" "end" "status"
[9] "trt" "start" "ET1" "ET2"
> fit4<-coxph(Surv(start,end,status)~Prison+Dose+ET1+ET2, data=AG)
```
> fit4
Call: coxph(formula = Surv(start, end, status) ~ Prison +
          Dose + ET1 + ET2, data = AG)

          coef exp(coef) se(coef)   z       p
Prison   0.3890    1.476   0.16859  2.31 2.1e-02
Dose   -0.0354    0.965   0.00645 -5.48 4.3e-08
ET1     0.4887    1.630   0.23396  2.09 3.7e-02
ET2     2.3971   10.991   0.52998  4.52 6.1e-06

Likelihood ratio test=79  on 4 df, p=3.33e-16  n= 337
> temp<-cox.zph(fit4)
> temp

             rho  chisq     p
Prison -0.0176  0.0465  0.829
Dose   0.0829  0.9305  0.335
ET1    0.0264  0.1059  0.745
ET2   -0.0089  0.0117  0.914
GLOBAL  NA  1.0595  0.901
> windows()
> par(mfrow=c(2,2))
> plot(temp)
This graph is automatically outputted from the optimal.change.point function.

Profile of the log–partial likelihood for a piecewise Cox PH model

change point (distinct survival times)
95% C.I.’s for the Clinic’s HR’s

\[ t < 461: [1.03, 2.58] \]

\[ t \geq 461: [3.89, 31.06] \]
Results:

• The output shows a significant $\hat{HR} = 1.63$ with $p$-value $= 0.037$ for the effect of Clinic when time $< 461$ days. For $t \geq 461$, the $\hat{HR} = 10.99$ is highly significant with $p$-value $= 6.1 \times 10^{-6}$.

• The table reports confidence intervals for the two HR’s. The general form of these 95% C.I.’s is $\exp(\text{coef} \pm 1.96 \times \text{se(coef)})$. The 95% C.I. for the HR when $t$ precedes 461 is a bit above 1 and is narrow. This supports a significant effect due to clinic during the first year and has good precision. The 95% C.I. for the HR when $t \geq 461$ lies above 1 and is very wide showing a lack of precision.

• These findings support what was displayed in Figure 3. But now it is quantified. There is strong statistical evidence of a large difference in clinic survival times after 461 days in contrast to a small and but still significant difference in clinic survival times prior to 461 days, with clinic 2 always doing better in retention of patients than clinic 1. After 461 days, clinic 2 is nearly 11 times more likely to retain a patient longer than clinic 1. Also, clinic 2 has $\frac{1}{11} \approx 10\%$ the risk of clinic 1 of a patient leaving its methadone treatment program.
• See Kleinbaum (1996, Chapter 6) for further analysis of this data.

• An alternative regression quantile analysis as presented in Chapter 8 may be appropriate when the PH assumption