Lecture thirteen: Model Checking (IV)

Testing the PH assumption

1. **Plotting method**

   One method to check the PH assumption, as mentioned many times before, is to examine it visually (graphical checks),

   categorical variable:

   Log-cumulative hazard plot: A plot of log of the negative log of the KM estimate of survival function ($log(-log(\hat{S}(t)))$) against the logarithm of the survival time. Parallel curves if the PH assumption holds.

   continuous covariate:

   Stratify the covariates if the categorization is meaningful, then do the same as in the categorical variable case.

   Example 4.9: Multiple myeloma study

   (a) Log-cumulative hazard plot

   (b) SAS program

```
options ls=80 nodate;
libname fu '../../sdata';
data fu.smyeloma;
        set fu.myeloma;
        if hb <=7 then hbcat = 0;
        else if 7 < hb <= 10 then hbcat = 1;
        else if 10 < hb <= 13 then hbcat = 2;
        else hbcat = 3;
proc freq;
        tables hbcat;
filename gsasfile 'ex49.gsf';
goptions reset=all gunit=pct border ftext=swissb htitle=6 htext=2.5
gaccess=gsasfile ROTATE=LANDSCAPE gsfmode=replace device=ps;
symbol1 interpol=join h=1 l=1 v=square  c=blue;
symbol2 interpol=join h=1 l=2 v=diamond c=black;
symbol3 interpol=join h=1 l=3 v=circle  c=red;
symbol4 interpol=join h=1 l=4 v=triangle c=brown;
```

```
proc lifetest plots=(lls);
        time survt * censor(0);
        strata hbcat;
run;
```

2. **Adding time-dependent variable**

   (a) **Time-dependent covariates and Cox model** (chapter 8, page 252) One of the extensions of proportional hazards model (Cox, 1972) is the introduction of time-dependent covariates. The model becomes

   $$h_i(t) = exp\{\sum_{j=1}^{p} \beta_j x_{ji}(t)\} h_0(t),$$

   which is no longer a proportinal hazards model.

   (b) **Time-dependent covariates and PH assumption**

   Suppose there is only one covariate, say, the treatment indicator variable X_1 (0, or 1), (it can be continous covariate). The proportional hazards function is

   $$h_i(t) = h_0(t) exp(\beta_1 x_{1i}).$$

   When the PH assumption is not satisfied, and the interest centers on the covariate, whose relative risk change over time. one approach is to introduce a new time-dependent covariate as follow

   $$x_2 = x_1 g(t),$$

   and add it into above equation

   $$h_i(t) = h_0(t) exp(\beta_1 x_{1i} + \beta_2 x_{2i}).$$

   i. The above appoach is also a way to test the PH assumption. If β_2 is significantly different from zero, then PH assumption is violated.

   ii. Choice of $g(t)$: difficult: estimate from data? Common choices are $log\, t$, t, and step function.

   iii. Interpretation of β_2:

A. if $\beta_2 < 0$, the relative risk decreases with time ($g(t)$).

B. if $\beta_2 > 0$, the relative risk increases with time ($g(t)$).

C. when $\beta_2 \neq 0$, treatment effect change over time.

D. if $\beta_2 = 0$, the relative risk is constant (i.e. PH assumption is satisfied)

iv. Example 4.12: Infection in patients on dialysis

A. SAS output

1) No time-dependent covaiate:

	Without Covariates	With Covariates
-2 LOG L	40.945	34.468

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
AGE	1	0.03037	0.02624	1.3400	0.2470	1.031
SEX	1	-2.71076	1.09590	6.1184	0.0134	0.066

2) Time-dependent covariate: age * t

	Without Covariates	With Covariates
-2 LOG L	40.945	32.006

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq	Hazard Ratio
AGE	1	-0.01716	0.04220	0.1653	0.6844	0.983
SEX	1	-2.02044	1.10725	3.3297	0.0680	0.133
tage	1	0.0004471	0.0003213	1.9366	0.1640	1.000

3) Time-dependent covariate: sex * t

Criterion	Without Covariates	With Covariates
-2 LOG L	40.945	34.104

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr>ChiSq	Hazard Ratio
AGE	1	0.03318	0.02686	1.5262	0.2167	1.034
SEX	1	-1.28076	2.53323	0.2556	0.6131	0.278
tsex	1	-0.07422	0.12190	0.3707	0.5426	0.928

B. The reduction of $-2log\,L$ were 2.462 ($p = 0.117$), and 0.364 ($p = 0.546$), respectively.

C. SAS program

```
options ls=80;
libname fu '../../sdata';
data work;
        set fu.dialysis;
proc phreg;
        model infectt*censor(0)= age sex;
proc phreg;
        model infectt*censor(0)= age sex tage;
                tage = age * infectt;
proc phreg;
        model infectt*censor(0)= age sex tsex;
                tsex = sex * infectt;
run;
```

3. **Time-varying coefficient model (VCM) and PH assumption** (Grambsch & Therneau, 1994, Biometrika, 81: 515-526).

   Most of the common alternative to proportional hazards can be cast in terms of a *time-varying coefficient* model. That is, we assume that

$$h_i(t) = exp\{\sum_{j=1}^{p} \beta_j(t) x_{ji}\} h_0(t).$$

The PH assumption is then a test for $\beta_j(t) = \beta_j$, which is a test for zero slope in the appropriate plot of $\hat{\beta}(t)$ on t.

(a)

$$\beta(t) = \beta + \theta g(t).$$

where $g(t)$ is some function (transformation of survival time).

(b) Implementation in Splus: **cox.zph()**, which supports four common choices: $g(t)$ is identity, log, rank (of survival times) and 1 - Kaplan-Meier.

(c) The choice of $g(t)$ depends on specific case, but no one will be optimal for all situations.

(d) Rescaled Schoenfeld residuals were used for constructing the test. See also ASSESS statement of PROC PHREG in SAS.

(e) Example 4.10: Infection in patients on dialysis.

```
SAS program - time-dependent covariate approach:

options ls=80;
libname fu '../../sdata';
data work;
        set fu.dialysis;
proc phreg;
        model infectt*censor(0)= age sex;
        output out=outp wtressch = wschage wschsex;
data fu.phdial;
        set outp;
        agex = wschage + 0.03;
        sexx = wschsex - 2.711;
filename x1 'phdial.pdf';
goptions reset=all gunit=pct border ftext=swissb htitle=6
htext=2.5 gsfname=x1 ROTATE=LANDSCAPE gsfmode=append device=pdf;
proc gplot;
        plot agex*infectt;
        plot sexx*infectt;
proc reg;
```

```
        model agex = infectt;
proc reg;
        model sexx = infectt;
run;
```

(f) Example 4.11: Infection in patients on dialysis.

```
Splus - Grambsch and Therneau test of proportional hazards:

ex411.s<-function(){
        tmpdf <- importData("../../sdata/disas7bdat")
        fcox <- coxph(Surv(infectt, censor)~age+sex,
                                            data = tmpdf, x=T)
        zph <- cox.zph(fcox)
        motif()
        par(mfrow=c(2,2))
        plot(zph)
        list(fcox, zph)
}

Splus output from ex411.s:

ex411.out[[1]]

Call:
coxph(formula=Surv(infectt,censor)~age+sex, data=tmpdf,x=T)

       coef exp(coef) se(coef)     z     p
age  0.0304    1.0308   0.0262  1.16 0.250
sex -2.7108    0.0665   1.0959 -2.47 0.013

Likelihood ratio test=6.48  on 2 df, p=0.0392  n= 13

ex411.out[[2]]
          rho chisq     p
   age  0.220 0.524 0.469
   sex -0.148 0.302 0.583
GLOBAL     NA 0.571 0.752
```

(g) Example: Renal insufficiency study: catheter placement.
survt: Time to infection, months
censor: Infection indicator (0=no, 1=yes)
cath: Catheter placement (1=surgically, 2=percutaneously)

i. the data: available at the course website.

ii. Plots from KM estimate.

iii. Splus output from coxph() and cox.zph().

```
kidd.out[[1]]

Call:
coxph(formula = Surv(survt, censor) ~ cath, data = kidddf)

       coef exp(coef) se(coef)     z    p
cath -0.613     0.542    0.398 -1.54 0.12

kidd.out[[2]]

Likelihood ratio test=2.41  on 1 df, p=0.121  n= 119
      rho chisq       p
cath -0.6   8.7 0.00319
```

where rho is the correlation coefficient between transformed survival time and the scaled Schoenfeld residuals; p is the two-sided p-value for testing the slope = 0, i.e. $\theta = 0$.

iv. plot of $\beta(t)$ vs $g(t)$.

v. Splus program

```
kidd.s<-function(){
        fcox <- coxph(Surv(survt, censor)~cath, data = kidddf)
        zph <- cox.zph(fcox)
        motif()
        plot(zph)
        list(fcox, zph)
}
```

4. **What to do if PH is violated?**

Modeling Survival Data: Extending the Cox Model (Therneau & Grambsch, 2000) lists several options. One of the choices is to partition the

time axis. The proportional hazards assumption may hold at least approximately over short time periods, although not over the entire study.

(a) Piece-wise Cox model (section 11.2.1, page 385)

(b) Renal insufficiency study (cont.)

i. Suppose to cut the time axis into two intervals by τ. Define two time-dependent covariates $z_2(t)$ and $z_3(t)$ as follow

$$z_2(t) = \begin{cases} cath & \text{if } t > \tau \\ 0 & \text{if } t \leq \tau \end{cases} \tag{1}$$

$$z_3(t) = \begin{cases} cath & \text{if } t \leq \tau \\ 0 & \text{if } t > \tau \end{cases} \tag{2}$$

and the hazard function is

$$h(t) = \begin{cases} h_0(t)e^{\beta_3 cath} & \text{if } t \leq \tau \\ h_0(t)e^{\beta_2 cath} & \text{if } t > \tau \end{cases} \tag{3}$$

ii. Determine the optimal value of "change point" τ
Notice that the likelihood will change values only at an event time, we pick the one with smallest value of $-2log\, L$,

iii. output from SAS with $\tau = 3.5$

Model Fit Statistics

Criterion	Without Covariates	With Covariates
-2 LOG L	208.907	195.002
AIC	208.907	199.002
SBC	208.907	201.518

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Chi-Square	Pr > ChiSq
aftert	1	-2.08891	0.75973	7.5600	0.0060

Event times	$-2log\,L$
0.5	195.756
1.5	200.448
2.5	195.259
3.5	195.002
4.5	198.336
5.5	200.986
6.5	197.711
8.5	200.855
9.5	202.168
10.5	203.335
11.5	204.336
15.5	201.659
16.5	202.954
18.5	204.119
23.5	205.239
26.5	206.457

```
pret          1        1.08175      0.78320       1.9077        0.1672
```

```
                        The PHREG Procedure

                Analysis of Maximum Likelihood Estimates

                                    Hazard      95% Hazard Ratio
                Variable             Ratio      Confidence Limits

                aftert               0.124       0.028       0.549
                pret                 2.950       0.636      13.692
```

iv. Interpretation: After 3.5 months, patients with percutaneously placed catheter do significantly better than patients given a surgically placed catheter (The conclusion in the book is wrong)

v. SAS program

```
options ls = 80;
libname fu '../../test1';
data work;
```

```
        set fu.kidd;
proc phreg;
        model survt*censor(0) = pret aftert/risklimits;
            if survt > 3.5 then aftert = cath; else aftert = 0;
            if survt <= 3.5 then pret = cath; else pret = 0;
run;
```

vi. Check PH assumption in the two time intervals: No violation.
SAS program for those checks:

```
options ls = 80;
libname fu '../../test1';
data work;
        set fu.kidd;
proc phreg;
    model survt*censor(0) = pret prett aftert aftertt/risklimits;
        if survt > 3.5 then aftert = cath; else aftert = 0;
        if survt <= 3.5 then pret = cath; else pret = 0;
        prett = pret*log(survt);
        aftertt = aftert*log(survt);
run;
```

Assignment eight: For recurrence of bladder cancer (Table B.2, page 501), use Cox model to investigate the effect of treatment (Placebo vs. thiotepa). In particular, by fitting a suitable time-dependent variable (or a varying coefficient model), test the assumption of proportional hazards with respect to all the covariates included in the model.

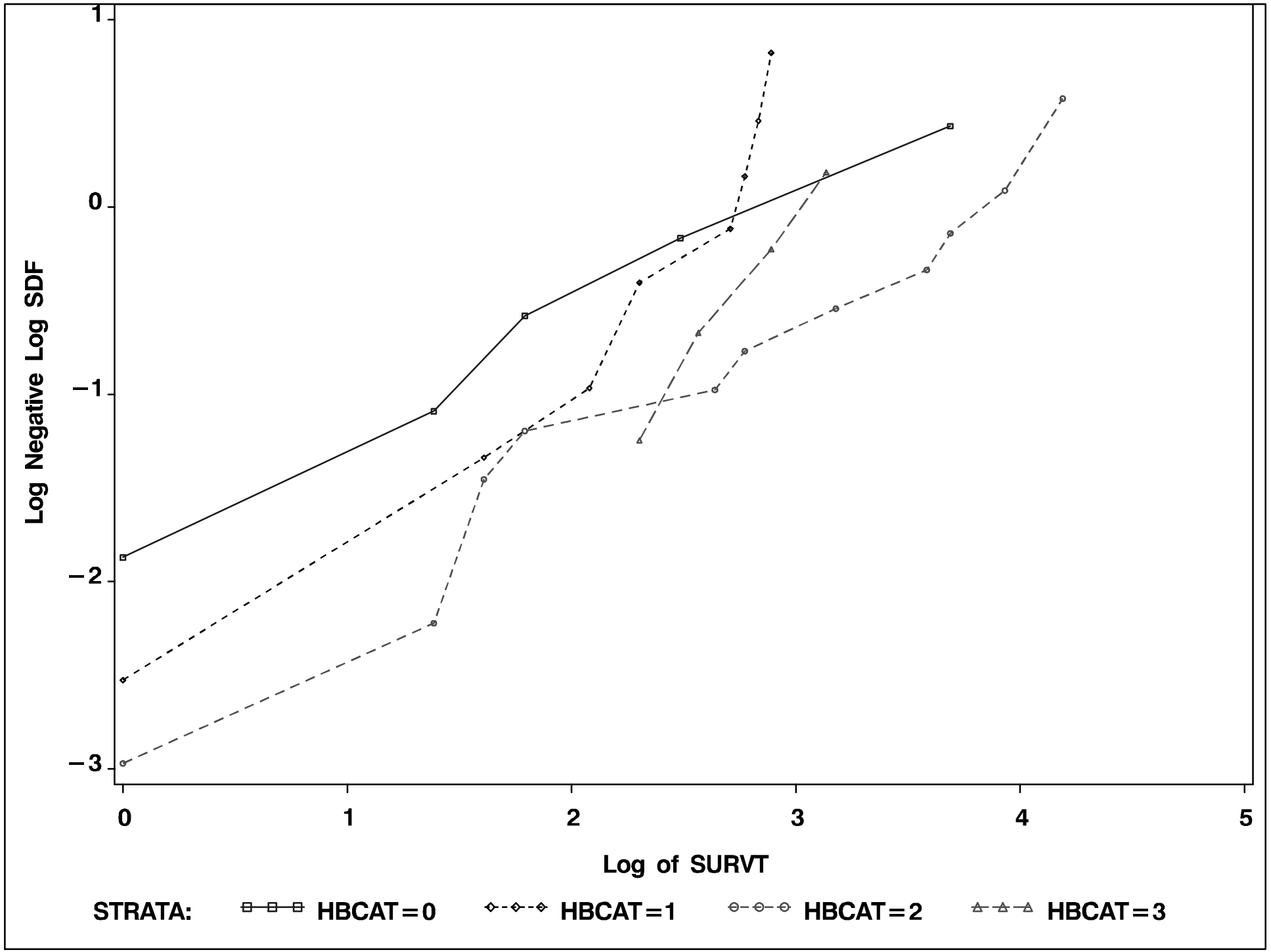
1
0
−1
−2
−3
Log Negative Log SDF
0
1
2
3
4
5
Log of SURVT
STRATA:
HBCAT=0
HBCAT=1
HBCAT=2
HBCAT=3

Figure 2: Example 4.11

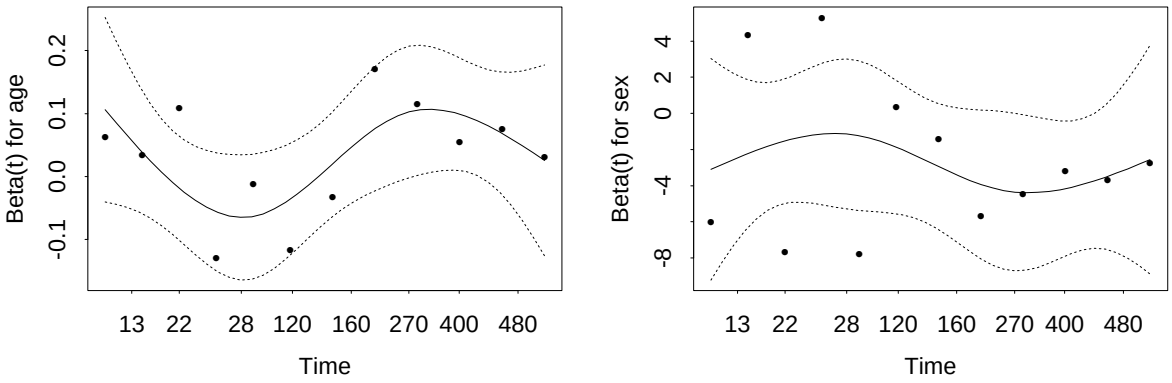

Figure 3: Example 9.2 in Klein and Moeschberger