Lecture 17: Cox Proportional Hazards Model

Lin Chen

Department of Public Health Sciences
The University of Chicago

The General Proportional Hazards (PH) Model

From last lecture:

- The general proportional hazards model is given by

$$h_i(t) = \exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi}) h_0(t) \tag{1}$$

  which assumes that covariates are multiplicatively related to the hazard.

  $\exp(\beta_1) \cdot \exp(\beta_2) \cdot \exp(\beta_3) \cdots$

- Two parts in the model:
  - Baseline hazard function: $h_0(t)$
  - PH "multiplier" – relative hazard function: $\exp(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_p x_{pi})$
- If $h_0(t)$ is specified according to a defined probability distribution (e.g. Weibull), then a *parametric* PH model is obtained.
- If the baseline hazard form $h_0(t)$ is unspecified, then a *semi-parametric* PH model is obtained.

Semi-parametric vs. parametric PH models

- In a semi-parametric PH model, the hazard function is not restricted to a specific functional form, thus the model has flexibility and widespread applicability.

- On the other hand, in a parametric PH model, if the assumption of a particular form of the hazard function is valid, inference will be more efficient (the estimates of parameters will tend to have smaller standard errors). Also, other quantities ($S(t)$, percentiles, etc) are easier to obtain.

Cox Proportional Hazards Model

- Sir David Cox (Cambridge) proposed that assuming the proportional hazards assumption holds, we can estimate the $\beta-$coefficients without any consideration of the form of the baseline hazard function.
  - Cox, DR (1972). “Regression Models and Life-Tables". *Journal of the Royal Statistical Society, Series B*.
  - The most cited paper in the whole history of JRSS, among the most cited statistics paper in medical literature, >53,000 citations.
- The semi-parametric proportional hazards model is referred to the Cox proportional hazards model or Cox model

Model and Interpretation of Parameters

- When X_j increases by 1 unit, given a PH model (parametric or semi-parametric),

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

$$\frac{h(t|(x_1, x_2, .., x_j+1, .., x_p))}{h(t|(x_1, x_2, .., x_j, .., x_p))} = \frac{\exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_j (x_j+1) + \cdots + \beta_p x_p) h_0(t)}{\exp(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_j x_j + \cdots + \beta_p x_p) h_0(t)} = e^{\beta_j}$$

Lecture 16 Slide 30

- The baseline hazard $h_0(t)$ cancels out when calculating hazard ratio. If the hazard ratio of a covariate X_j is of interest, $h_0(t)$ can be left unspecified.

- The coefficient e^{β_j} represents the hazard ratio for one-unit increase in X_j controlling for the other covariates.

Model and Interpretation of Parameters

- β_j represent the log hazard ratio for one unit increase in X_j controlling for the other covariates.
- $\beta_j = 0$ (HR=1.0) implies no association between the jth covariate value with the hazard of an event controlling for the other covariates.
- $\beta_j < 0$ (HR < 1.0) means larger values of X_j are associated with lower hazard at any time, and thus longer survival times controlling for the other covariates.
- $\beta_j > 0$ (HR > 1.0) means larger values of X_j are associated with higher hazard at any time, and thus shorter survival times controlling for the other covariates.

Hazard ratio of any two subjects at any time

- What is the relative hazard comparing an individual i with covariate values $(x_{i1}, ..., x_{ip})$ to an individual j with covariate values $(x_{j1}, ..., x_{jp})$?

$$\begin{aligned}\frac{h_i(t)}{h_j(t)} &= \frac{\exp(\beta_1 x_{i1} + \cdots + \beta_p x_{ip}) h_0(t)}{\exp(\beta_1 x_{j1} + \cdots + \beta_p x_{jp}) h_0(t)} \\ &= \exp(\beta_1 (x_{i1} - x_{j1}) + \beta_2 (x_{i2} - x_{j2}) + \cdots \beta_p (x_{ip} - x_{jp})) \qquad (2)\end{aligned}$$

- These calculations follow those for odds ratios from binary models, etc

Estimation

- Cox PH model and parametric PH are the same in terms of interpretation of parameters, however, they are different in terms of estimation.
  - Parameters in a parametric PH model are estimated by maximizing the likelihood.
  - Parameters in a Cox PH model are estimated by maximing a quantity identified (by Cox) as the *partial likelihood*.

Estimation in parametric PH model

- Once a parametric model is specified, the hazard function for subject i in the sample, $h_i(t)$, is a functions of $\boldsymbol{\beta}$ and parameters in $h_0(t)$. Given the relationship to $h_i(t)$, $f_i(t)$ and $S_i(t)$ are also functions of $\boldsymbol{\beta}$ and parameters in $h_0(t)$.
- Let (t_i, δ_i) be the pair of observations of survival time for the i^{th} subject, t_i is the observed survival time and δ_i is the indicator of event.
  - If $\delta_i = 1$, t_i is the actual survival time, the likelihood of observing t_i is given by $f_i(t_i)$
  - If $\delta_i = 0$, t_i is the censored survival time, under right censoring assumption, then we know that the actual survival time is at least t_i, the probability of which is given by $S_i(t_i)$
  - Thus, the likelihood function for subject i is given by $\{f_i(t_i)\}^{\delta_i}\{S_i(t_i)\}^{1-\delta_i}$
- The estimates of $\boldsymbol{\beta}$ and parameters in $h_0(t)$ are obtained by maximizing $\prod_{i=1}^{n}\{f_i(t_i)\}^{\delta_i}\{S_i(t_i)\}^{1-\delta_i}$

Estimation in Cox PH model

- Without assuming a particular form of $h_0(t)$, it turns out that the $\boldsymbol{\beta}-$coefficients can be estimated using the partial likelihood idea:
  - Suppose that there are r ordered survival times, let $t_{(j)}$ be the j^{th} ordered survival time and let $R(t_{(j)}) = \{i \mid t_i \geq t_{(j)}\}$ be the subjects at risk at time $t_{(j)}$, called the *risk set*.
  - Consider the simple scenario where there is only 1 death at time $t_{(j)}$ and that is the subject with covariate $\boldsymbol{x_{(j)}}$.
  - Then the probability that it is the subject with covariate $\boldsymbol{x_{(j)}}$ who died at time $t_{(j)}$ given that there is a death at time $t_{(j)}$ is given by

$$\begin{aligned} L_{(j)} &= Pr(\textit{subject with } x_{(j)} \textit{ died at time } t_{(j)} \mid \textit{ one subj. died at } t_{(j)}) \\ &= \frac{h_{(j)}(t_{(j)})}{\sum_{i \in R(t_{(j)})} h_i(t_{(j)})} = \frac{\exp(\boldsymbol{\beta}^T \boldsymbol{x_{(j)}})}{\sum_{i \in R(t_{(j)})} \exp(\boldsymbol{\beta}^T \boldsymbol{x_i})} \end{aligned} \tag{3}$$

- The estimates of $\boldsymbol{\beta}$ are obtained by maximizing the product over all unique times $\prod_{j=1}^{r} L_{(j)}$.

Inference in PH model: hypothesis tests and confidence interval

- In a PH model, the estimated coefficients have an approximate normal distribution when there are adequate numbers of events.
  - Thus the test statistic $Z = \hat{\beta}_j / se(\hat{\beta}_j)$ is used to test the null hypothesis $H_0 : \beta_j = 0$ in the presence of the other parameters $(\beta_1, ..., \beta_{j-1}, \beta_{j+1}, ..., \beta_p)$ in the model.
  - A $100 \times (1 - \alpha)\%$ confidence interval for β_j can be found using $\hat{\beta}_j \pm z_{\alpha/2} se(\hat{\beta}_j)$.

Inference in PH model: comparing two nested models

- Likelihood ratio test (difference in deviance) can be used to compare two nested models:
  - Model (1): $h(t) = \exp(\beta_1 x_1 + \cdots + \beta_p x_p) h_0(t)$
  - Model (2): $h(t) = \exp(\beta_1 x_1 + \cdots + \beta_p x_p + \beta_{p+1} x_{p+1} + \cdots \beta_{p+k} x_{p+k}) h_0(t)$
- We wish to test:

$$H_0 : \beta_{p+1} = \beta_{p+2} = \cdots = \beta_{p+k} = 0 \textit{ (i.e. Model (1) is correct)}$$

$$H_1 : \beta_j \neq 0 \textit{ for at least one } j = p+1, p+2, ..., k$$

- Let $\hat{L}_1, \hat{L}_2$ be the maximized likelihoods under model (1) and model (2), respectively. Under the H_0, the likelihood ratio test statistic $-2(\log\hat{L}_1 - \log\hat{L}_2) \sim \chi^2_k$ approximately.

Example 1: *Prognosis for women with breast cancer*

Table 1: Survival times (in months) of women with tumors that were negatively or positively stained with HPA. Censored survival times are labeled with an asterisk.

Negative staining	Positive staining	
23	5	68
47	8	71
69	10	76*
70*	13	105*
71*	18	107*
100*	24	109*
101*	26	113
148	26	116*
181	31	118
198*	35	143
208*	40	154*
212*	41	162*
224*	48	188*
	50	212*
	59	217*
	61	225*

Prognosis for women with breast cancer: Cox PH model

The Stata function `stcox` followed by covariates will estimate the Cox model

```
. use "prognosis_breast_cancer.dta", clear
. stset time status
  ......
. stcox i.stain, nolog

        failure _d:  status
  analysis time _t:  time

Cox regression -- Breslow method for ties

No. of subjects =           45                 Number of obs    =           45
No. of failures =           26
Time at risk    =         4331
                                               LR chi2(1)       =         3.87
Log likelihood  =   -85.047944                 Prob > chi2      =       0.0491

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     2.stain |   2.479398   1.241987     1.81   0.070      .9288808    6.618086
------------------------------------------------------------------------------
```

Prognosis for women with breast cancer by cell staining

- Under the Weibull PH model (from last lecture), the estimated hazard ratio of a women in a positively stained group compared to a women in a negatively stained group is $e^{\hat{\beta}} = 2.55$ with $se(e^{\hat{\beta}}) = 1.27$, and a 95% CI $(0.97, 6.78)$

- Under the Cox PH model, the estimated hazard ratio of a women in a positively stained group compared to a women in a negatively stained group is $e^{\hat{\beta}} = 2.48$ with $se(e^{\hat{\beta}}) = 1.24$, and a 95% CI $(0.93, 6.62)$

- The hazard ratios are almost the same for the two models. This is because the Weibull is good choice for the underlying hazard (as discussed in lecture 16). Note that both models are PH models, and PH is valid for the data.

Prognosis for women with breast cancer: Cox PH model (log hazard ratio form)

The option `nohr` will give the results in the log hazard ratio form.

```
. stcox i.stain, nolog nohr

        failure _d:  status
  analysis time _t:  time

Cox regression -- Breslow method for ties

No. of subjects =           45                     Number of obs    =        45
No. of failures =           26
Time at risk    =         4331
                                                   LR chi2(1)       =      3.87
Log likelihood  =   -85.047944                     Prob > chi2      =    0.0491

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
    2.stain |    .9080157   .5009228     1.81   0.070    -.0737749    1.889806
------------------------------------------------------------------------------
```

- Under the Cox PH model, the estimated log hazard ratio of a women in a positively stained group compared to a women in a negatively stained group is $\hat{\beta} = 0.91$ with $se(\hat{\beta}) = 0.50$, and a 95% CI $(-0.07, 1.89)$

Estimate of the Survivor Function

- Recall in the PH Model

$$h_i(t) = h_0(t)\exp(\boldsymbol{\beta}^T \boldsymbol{x_i})$$

and

$$H_i(t) = H_0(t)\exp(\boldsymbol{\beta}^T \boldsymbol{x_i})$$

So

$$\begin{aligned} S_i(t) &= e^{-H_i(t)} \\ &= e^{-H_0(t)\exp(\boldsymbol{\beta}^T \boldsymbol{x_i})} \\ &= \{e^{-H_0(t)}\}^{\exp(\boldsymbol{\beta}^T \boldsymbol{x_i})} \\ &= \{S_0(t)\}^{\exp(\boldsymbol{\beta}^T \boldsymbol{x_i})}, \end{aligned}$$

where $h_0(t)$, $H_0(t)$, and $S_0(t)$ are the functions corresponding to $x_i = 0$. They are called the *baseline* hazard, cumulative hazard and survivor functions.

- Stata provides us the Kaplan-Meier based estimates of $H_0(t)$ and $S_0(t)$ using options `basechazard` and `basesurv`, respectively

Prognosis for women with breast cancer: estimate baseline survivor functions

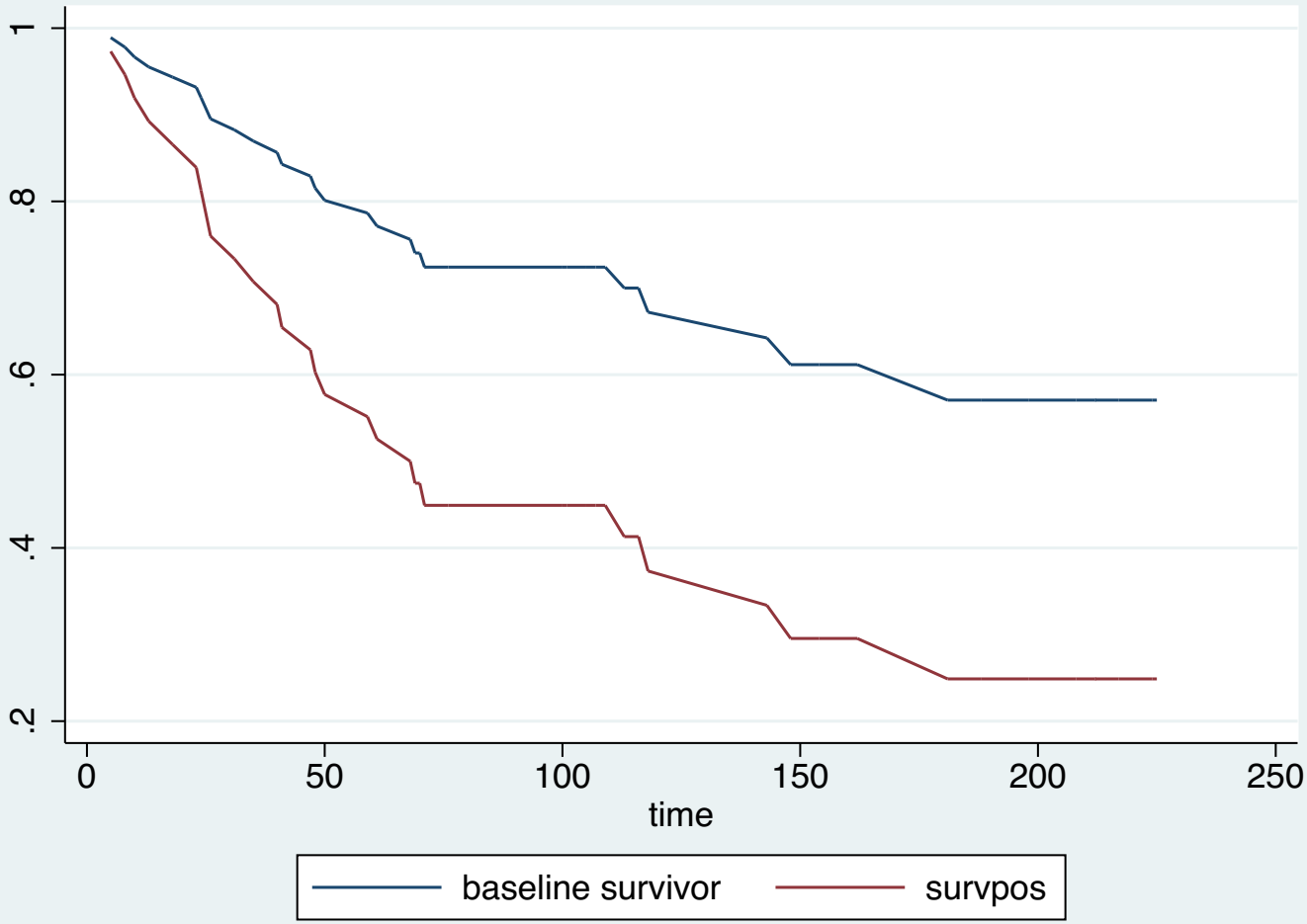
We first generate the baseline $S_0(t)$ function, then other $S_i(t|X)$ curves are a function of $S_0(t)$ times the quantity $\exp(\boldsymbol{\beta}^T \boldsymbol{x_i})$

```
. quietly stcox i.stain, basesurv(s0)

. sort time

. list time status s0 in 1/5

     +---------------------------+
     | time   status         s0  |
     |---------------------------|
  1. |    5        1   .98908253 |
  2. |    8        1   .97798381 |
  3. |   10        1   .96669555 |
  4. |   13        1   .95520883 |
  5. |   18        1   .94351404 |
     +---------------------------+
```

Prognosis for women with breast cancer: estimate survivor functions

Here the women with positive stain has $S_i(t|x_i = 1) = S_0(t)^{\exp(\hat{\beta})}$.

```
. gen survpos = s0^2.479398

. twoway line s0 survpos time
```



Example 2: *Treatment of hypernephroma*

- survival time is recorded in months for kidney cancer patients.
- nephrectomy codes: 0 = no nephrectomy, 1 = nephrectomy
- age codes: 1 for age<60, 2 for age 60-70, 3 for age >70

```
. use treatment_of_hypernephroma.dta
. list in 1/10
```

	nephre~y	age	time	status
1.	0	1	9	1
2.	0	1	6	1
3.	0	1	21	1
4.	0	2	15	1
5.	0	2	8	1
6.	0	2	17	1
7.	0	3	12	1
8.	1	1	104	0
9.	1	1	9	1
10.	1	1	56	1

Treatment of hypernephroma: Cox PH model

- Let $Age2$ be the indicator for age being in the range 60-70, and $Age3$ be the indicator for age being in the range >70. Then a proportional hazard model could be specified to be

$$h_i(t) = \exp(\beta_1 \cdot \textit{nephrectomy}_i + \beta_2 \cdot \textit{Age2}_i + \beta_3 \cdot \textit{Age3}_i) h_0(t) \qquad (4)$$

- $h_0(t)$ is the hazard for a subject that didn't receive a nephrectomy and of an age <60 at the time of diagnosis. It's unspecified in the Cox PH model.

Treatment of hypernephroma: Estimate the Cox model

```
. stset time status
. stcox nephrectomy i.age, nohr nolog basesurv(s0)

        failure _d:  status
  analysis time _t:  time

Cox regression -- Breslow method for ties

No. of subjects =           36                 Number of obs    =         36
No. of failures =           32
Time at risk    =         1340
                                               LR chi2(3)       =      12.16
Log likelihood  =    -82.75418                 Prob > chi2      =     0.0069

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 nephrectomy |  -1.411453   .515237    -2.74   0.006    -2.421299   -.4016071
             |
         age |
          2  |   .0125313   .4245943     0.03   0.976    -.8196582    .8447209
          3  |   1.341567   .5917646     2.27   0.023     .1817294    2.501404
------------------------------------------------------------------------------
```

Treatment of hypernephroma: Prediction

- The fitted model is given by

$$\hat{h}_i(t) = \exp\left(-1.411 \cdot \textit{nephrectomy}_i + 0.013 \cdot \textit{Age2}_i + 1.342 \cdot \textit{Age3}_i\right) \cdot \hat{h}_0(t)$$

- What is the estimated hazard ratio of a subject k of an age above 70 at the time of diagnosis without nephrectomy compare to a subject j of an age below 60 at the time of diagnosis with nephrectomy?

$$\frac{\hat{h}_k(t)}{\hat{h}_j(t)} = \frac{\exp(-1.411 \cdot 0 + 0.013 \cdot 0 + 1.342 \cdot 1)}{\exp(-1.411 \cdot 1 + 0.013 \cdot 0 + 1.342 \cdot 0)}$$

$$= \exp(-1.411 \cdot (-1) + 1.342 \cdot 1) = \exp(2.753)$$

```
. lincom nephrectomy*(-1)+3.age*1
 ( 1)  - nephrectomy + 3.age = 0
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         (1) |     2.75302   .7521929     3.66   0.000     1.278749    4.227291
------------------------------------------------------------------------------
```

The estimated hazard ratio is $e^{2.753}$ with a 95% CI $(e^{1.279}, e^{4.23})$

Treatment of hypernephroma: estimated survivor function

- Recall that under the PH model, the relationship between the survivor function for subject i and survivor function for a baseline subject is given by

$$S_i(t) = \{S_0(t)\}^{\exp(\beta^T x_i)} \tag{5}$$

- Given the estimated $S_0(t)$, the survivor function for a subject k of age above 70 at diagnosis without nephrectomy is estimated by

$$S_k(t) = \{S_0(t)\}^{\exp(-1.411\cdot 0+0.013\cdot 0+1.342\cdot 1)} = \{S_0(t)\}^{\exp(1.342)} \tag{6}$$

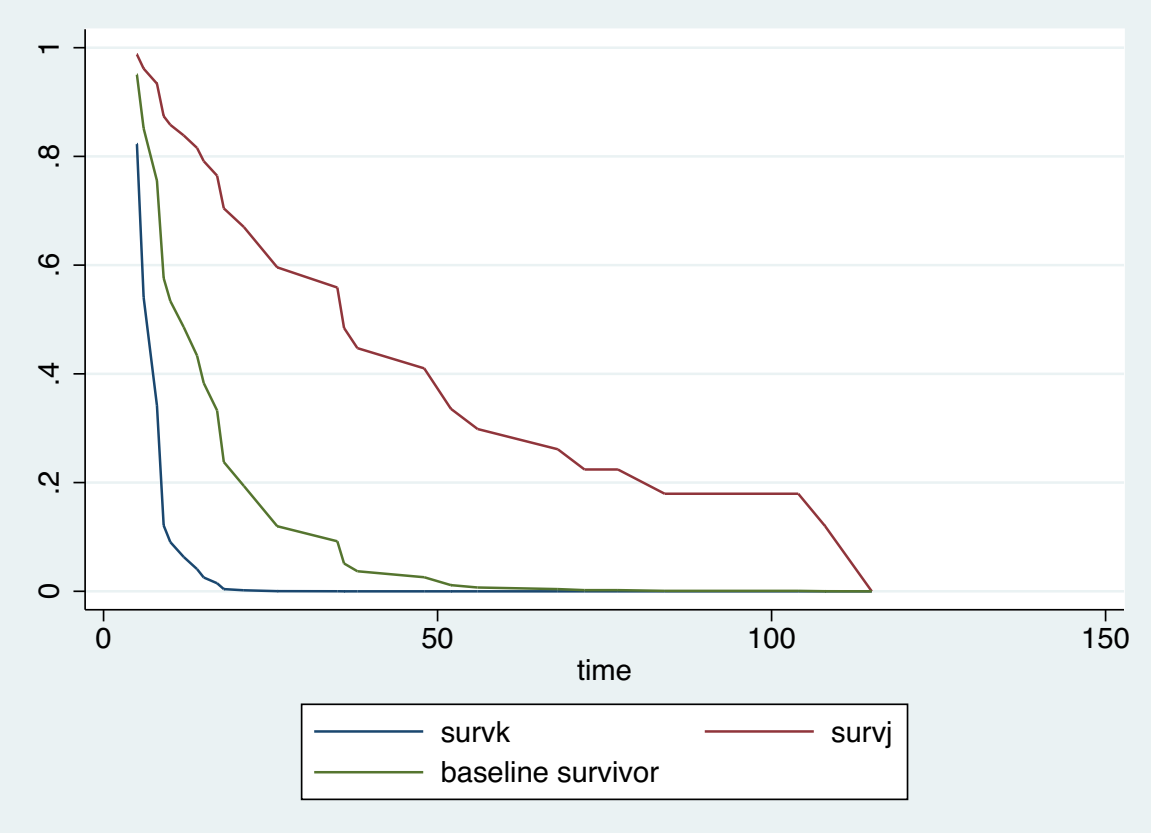
- Given the estimated $S_0(t)$, the survivor function for a subject j of age below 60 at diagnosis with nephrectomy is estimated by

$$S_j(t) = \{S_0(t)\}^{\exp(-1.411\cdot 1+0.013\cdot 0+1.342\cdot 0)} = \{S_0(t)\}^{\exp(-1.411)} \tag{7}$$

Treatment of hypernephroma: Survivor function

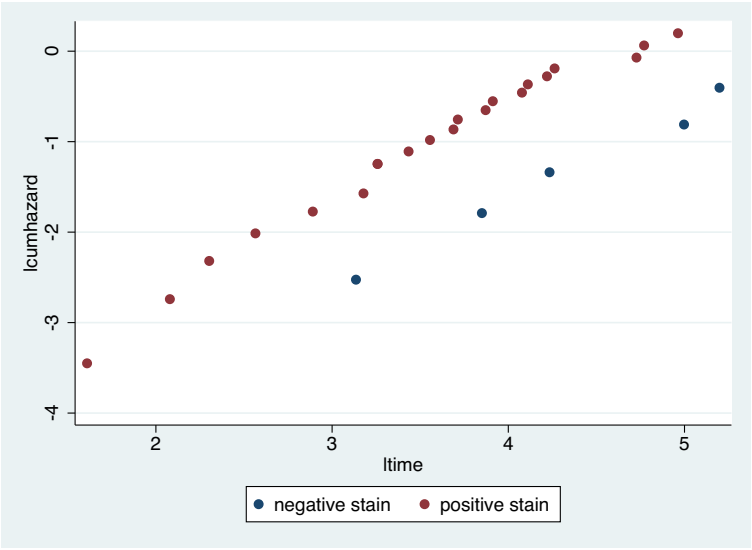
Estimate the survivor function for each group with different x_i

```
. gen survk = s0^(exp(1.342))

. gen survj = s0^(exp(-1.411))

. sort time

. twoway line survk survj s0 time
```



Evaluation of the Proportional Hazards Assumption

- There are several ways to evaluate the PH assumption, which is (perhaps more) important in the Cox model given it is the only assumption on which the model is based, and the method is used for widely
- We already looked at one simple approach, plotting $H(t)$ against time in groups, checking for parallelism.
  This is the breast cancer (HPA) data:



Evaluation of the Proportional Hazards Assumption

Another test - facilitated in Stata

- Like other models, survival models have residuals (observed - fitted value) quantities that can be used to diagnose aspects of the model fit, identify outliers, etc. These take on a bit of an odd form in survival data, but can serve certain purposes.
- One such use is a formal proportionality test for each covariate, based on *Schoenfeld residuals* -
- Schoenfeld test and related tests are framed as H_0 : PH holds vs. H_a : Deviation from PH

Evaluation of the Proportional Hazards Assumption

In Stata, we can use `estat phtest` (PH-Test) following a Cox regression to check PH assumption.

```
. use "prognosis_breast_cancer.dta", clear
. stset time status
. stcox i.stain, nolog
  ......
. estat phtest

        Test of proportional-hazards assumption

        Time: Time
        ----------------------------------------------------------------
                    |                 chi2       df       Prob>chi2
        ------------+---------------------------------------------------
        global test |                 2.08        1           0.1492
        ----------------------------------------------------------------
```

An insignificant p-value means the PH assumption is not violated. As expected, the PH assumption holds in this example.

Evaluation of the Proportional Hazards Assumption

Treatment of hypernephroma data: we can also test separately PH assumption for each of the multiple covariates, as well as globally testing all. The option ", `detail`" evaluates the PH assumption for each predictor and overall.

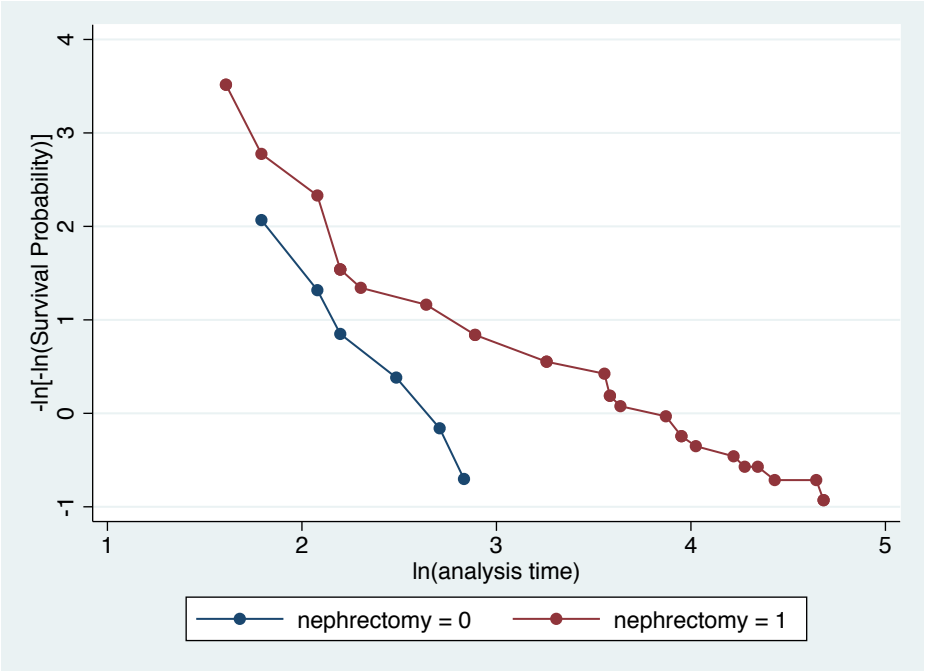
```
. use "treatment_of_hypernephroma.dta"

. stset time status

. stcox i.age nephrectomy , nolog

. estat phtest, detail

        Test of proportional-hazards assumption

        Time:  Time
        ----------------------------------------------------------------
                    |        rho       chi2       df       Prob>chi2
        ------------+---------------------------------------------------
        1b.age      |          .          .        1               .
        2.age       |    -0.26604       2.39        1          0.1221
        3.age       |    -0.05804       0.12        1          0.7281
        nephrectomy |    -0.06828       0.18        1          0.6715
        ------------+---------------------------------------------------
        global test |                   2.51        3          0.4736
        ----------------------------------------------------------------
```

The PH assumption holds for all covariates in this example.

Evaluation of the Proportional Hazards Assumption

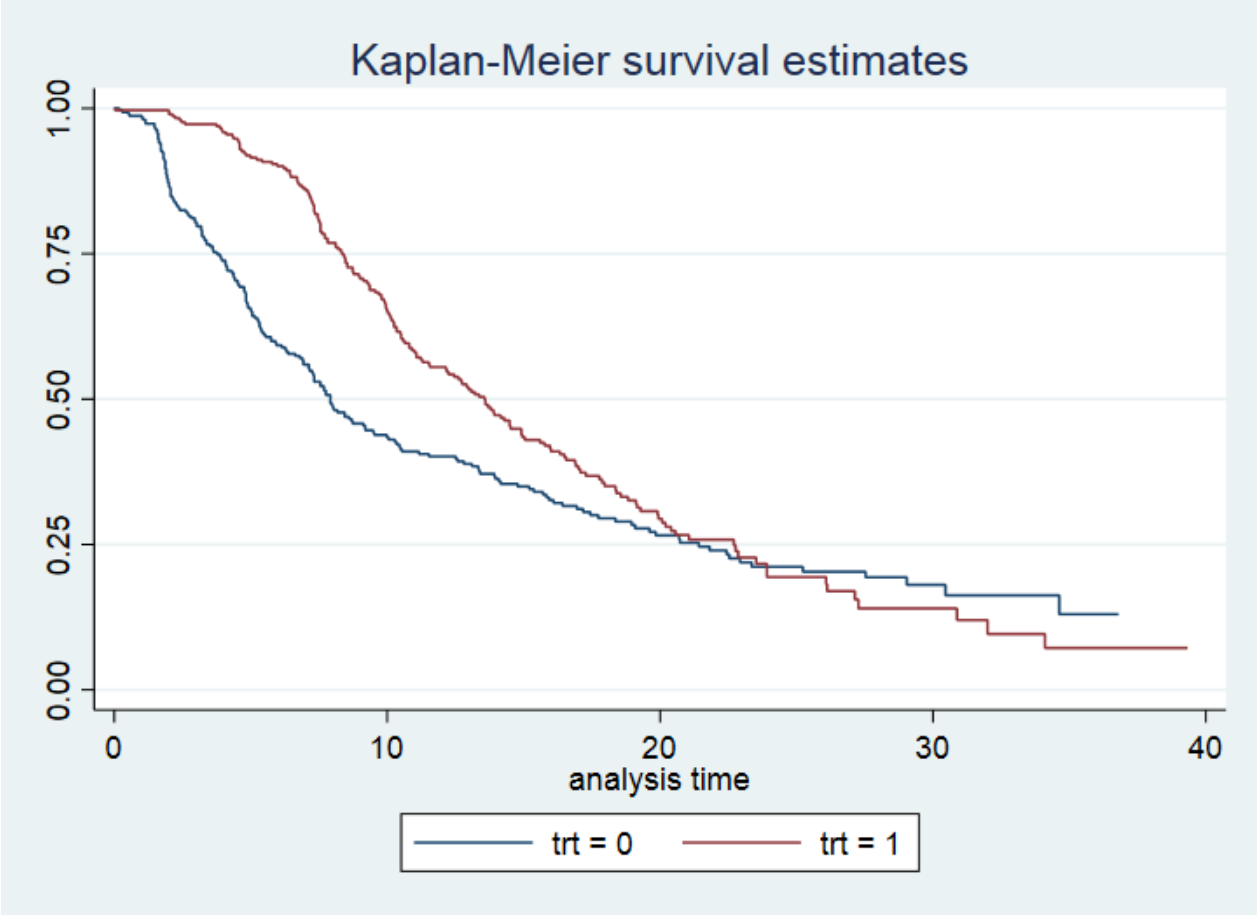
The Stata function `stphplot` can also be used to check PH assumption. It plots $-\log\{-\log(\text{survival})\}$ curves for each category of a categorical covariate against $\log(\text{analysis time})$. It can adjust for other covariates. The PH assumption holds if the curves are parallel.

```
. stphplot, strata(nephrectomy) adjust(2.age 3.age)

        failure _d:  status
  analysis time _t:  time
```



Another example when PH assumption does not hold

Randomized trial of bevacizumab (Avastin) added to chemo/RT for glioblastoma. Addition of bevacizumab showed early delay of progression, but the survival curves later converge.



Another example when PH assumption does not hold

Testing the PH assumption

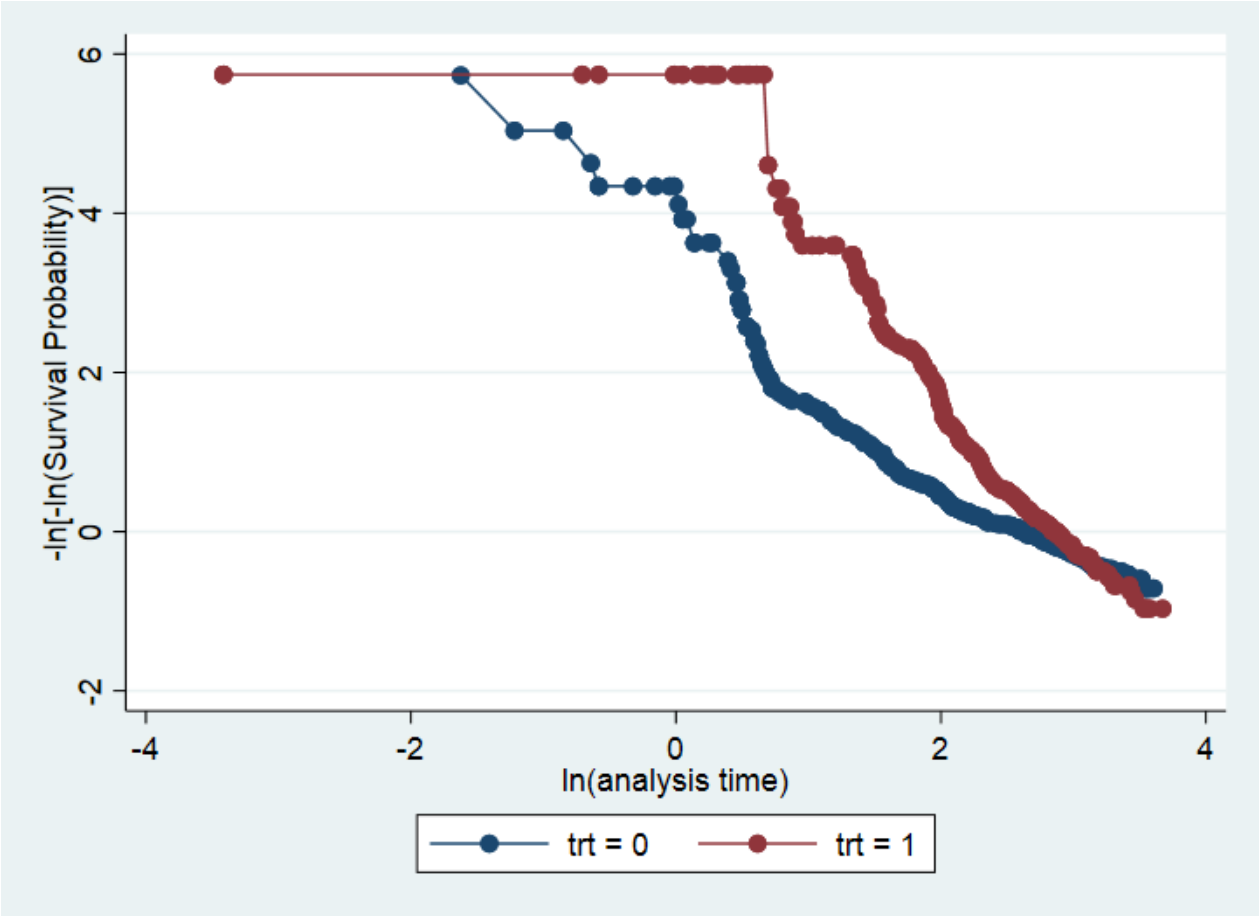
```
. use "GBM.dta"
. stset time event
. stcox trt , nolog

.....

Cox regression -- Breslow method for ties

No. of subjects =          621                     Number of obs    =          621
No. of failures =          396
Time at risk    =  6679.434945
                                                   LR chi2(1)       =        11.91
Log likelihood  =   -2222.8793                     Prob > chi2      =       0.0006

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         trt |   .7054104   .071377    -3.45   0.001     .5785128    .8601431
------------------------------------------------------------------------------

. estat phtest

        Test of proportional-hazards assumption

        Time:  Time
        ----------------------------------------------------------------
                    |                 chi2        df        Prob>chi2
        ------------+---------------------------------------------------
        global test |               49.96         1          0.0000
        ----------------------------------------------------------------
```

PH Fails

Proportional Hazards Assumption Gone Wrong

Diagnostic plot:

```
. stphplot, by(trt)
```



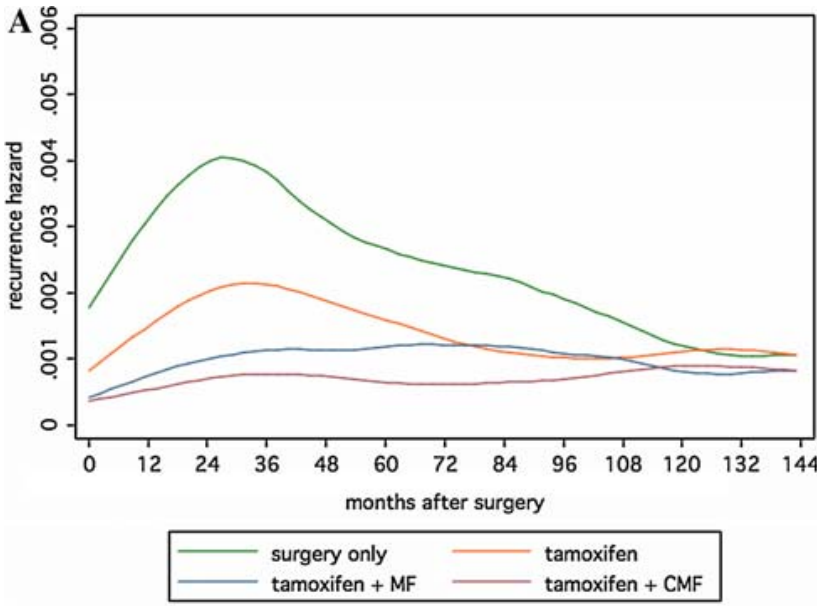
Extensions of the Cox PH Model

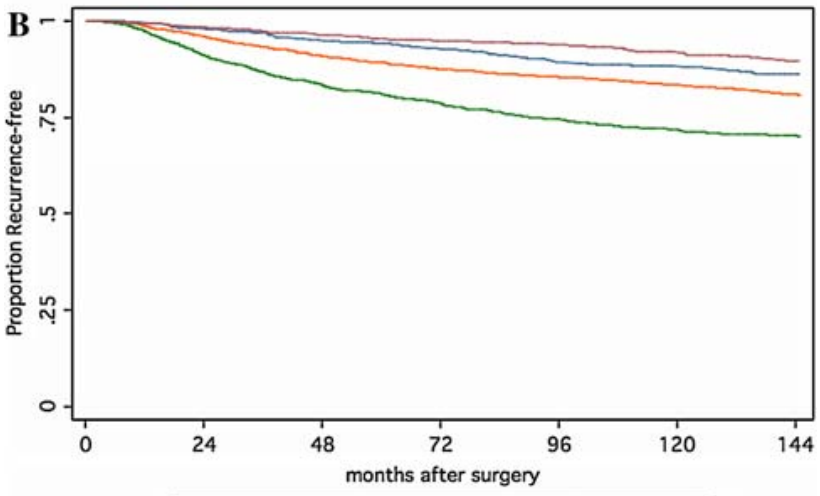
To address model limitations and extend to specific data structures, there are many extensions of the model

- Stratified PH model - can stratify on non-PH factors, or in cases where different 'baseline' hazard expected
- Time-varying covariates - Instead of covariate x fixed at start of follow-up, covariate $z(t)$ evolves over time
- Time-varying coefficients - covariate effect is a function of time (this is non-PH by definition)
- multiple events, multi-state models - some concepts from survival, such as the hazard, extend to recurrent events, etc

Hazards after Treatment in Early Stage Breast Cancer - Example of more flexible analysis

These estimates are actually from *stratified* (by treatment type) model - to permit inference on whether hazard shapes differ by treatment group. Other covariates within strata are assumed PH





Summary

The Cox Proportional Hazard Model

- A widely used semi-parametric method to relate covariates to hazard and survivor function
- Flexible, but does have key assumption PH that should be checked – extensions are available

Course Wrap-up