

Self-Study Material 3: Cox Proportional Hazards Model

Purpose:

1. Introduction to the Cox proportional hazards model
2. How to fit PHM using STATA

1. Introduction to the Cox proportional hazards model

Unlike parametric methods Cox's method does not require some particular probability distribution to represent survival times. That's why we call it semi-parametric model, which makes Cox's method more robust. Another advantage of using Cox's method is that it's relatively easy to incorporate time-dependent covariates. As we know, we use the method of maximum likelihood to estimate the regression parameter in parametric model, while we use the method of *maximum partial likelihood* to estimate the parameters in Cox's model. What's remarkable about partial likelihood is that you can estimate the coefficients without having to specify the baseline hazard function $\lambda_0(t)$.

The Cox's model is represented as:

$$\lambda_i(t) = \lambda_0(t) \exp\{\beta_1 x_{i1} + \dots + \beta_p x_{ip}\} \quad (1)$$

where $\lambda_i(t)$ is the hazard function for the i th subject, $\lambda_0(t)$ is the baseline hazard function, and β_1, \dots, β_p are parameters to be estimated.

2. How to fit PHM using Stata

```
. stcox treat
```

```
      failure _d: status == 1  
      analysis time _t: dur
```

```
Iteration 0:  log likelihood = -47.041906  
Iteration 1:  log likelihood = -46.382985  
Iteration 2:  log likelihood = -46.382377  
Refining estimates:  
Iteration 0:  log likelihood = -46.382377
```

```
Cox regression -- Breslow method for ties
```

```
No. of subjects =          25
```

```
Number of obs   =          25
```

```

No. of failures =          17
Time at risk   =         15334

LR chi2(1)      =          1.32
Log likelihood = -46.382377    Prob > chi2     =          0.2508

```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
treat	1.773237	.9036446	1.12	0.261	.6531198 4.814382

Statistics in Stata Output

- *Hazard ratios.*

The hazard is the probability of the event occurring in time t , given survival to time t . The `stcox` command generates the hazard ratio by default, along with its standard error, p value, and confidence intervals. A hazard ratio of 1.0 indicates the variables in the model have no effect on time to event for the status variable. The more the hazard ratio is below 1.0, the more the variables reduce the likelihood of the event. The more above 1.0, the more the variables increase the likelihood of the event. The hazard ratio of the covariates is the amount of change in the rate of the event occurring for each unit change in the covariate. For instance, a hazard ratio of 1.12 means that there is a 12% increase in the rate of the event for every 1 unit increase in the covariate, controlling for other covariates in the model. Where the covariate is a dichotomous grouping variable, such as treatment, the estimated hazard ratio for the treatment is 1.77, which means that the hazard of death for those who received the second treatment is about 1.77 of the hazard for those who receive the first treatment.

- *Likelihood ratio test of the model.*

The `stcox` command also generates a log likelihood value and its probability. If this value is significant, the model as a whole is significant.

- *Regression coefficients.*

After fitting a Cox regression model with `stcox`, it can be reinvoked to display regression coefficients rather than hazard ratios: `stcox, nohr`. Output will be very similar to the default, but with coefficients substituted for hazard ratios.

The first feature to note in such a table is the sign of the regression coefficients. A positive sign means that the hazard of risk is higher, and thus the prognosis worse, for subjects with higher values of that variable. For explanatory variables which are continuous, the regression coefficient refers to the increase in log

hazard for an increase of 1 in the value of the covariate.

```
. stcox treat, nohr
```

```
-----
```

_t	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
treat	.5728065	.5096018	1.12	0.261	-.4259946	1.571608

```
-----
```

Tests of Proportionality in STATA

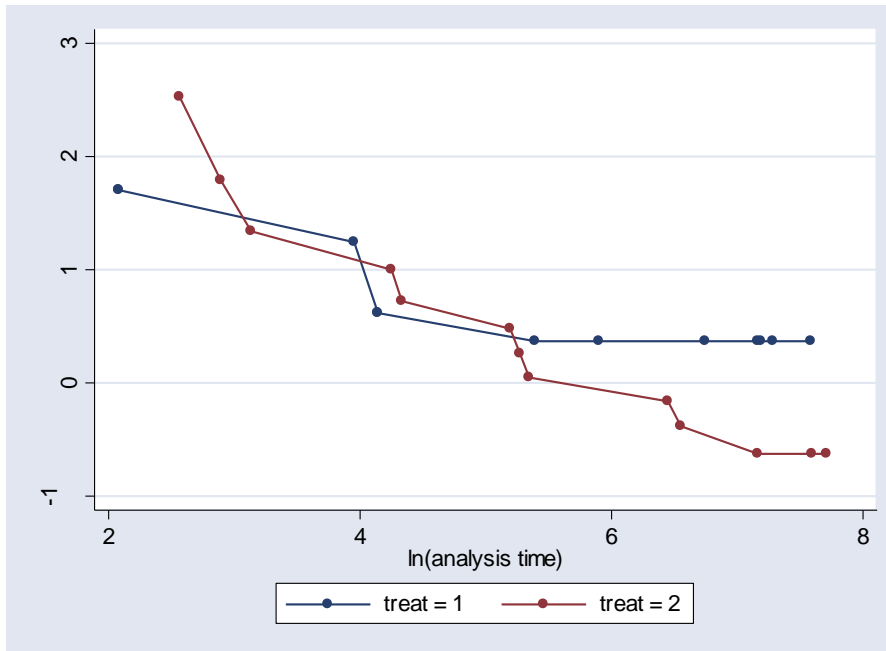
When modeling a Cox proportional hazard model a key assumption is proportional hazards. The goal of this lab is to illustrate how to test for proportionality in STATA using an example.

1. Kaplan-Meier Curves

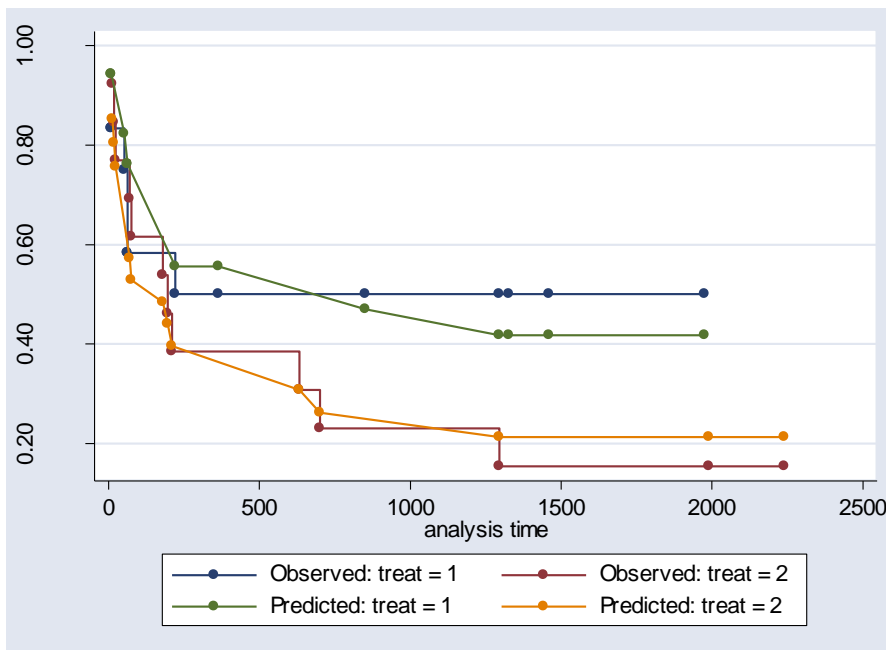
Works best for time fixed covariates with few levels. If the predictor satisfy the proportional hazard assumption then the graph of the survival function versus the survival time should results in a graph with parallel curves, similarly the graph of the $\log(-\log(\text{survival}))$ versus \log of survival time graph should result in parallel lines if the predictor is proportional. This method does not work well for continuous predictor or categorical predictors that have many levels because the graph becomes to "cluttered". Furthermore, the curves are sparse when there are fewer time points and it may be difficult to gage how close to parallel is close enough. Due to space limitations we will only show the graphs for the predictor treat.

sts graph, by(treat)

stphplot plots $-\ln(-\ln(\text{survival}))$ curves for each category of a nominal or ordinal independent variable versus $\ln(\text{time})$. These are often referred to as "log-log" plots. Optionally, these estimates can be adjusted for covariates. The proportional hazards assumption is not violated when the curves are parallel.



stcoxkm plots Kaplan-Meier observed survival curves and compares them to the Cox predicted curves for the same variable. (Do not run stcox prior to this command; stcoxkm will execute stcox itself to fit the model and obtain predicted values.) The closer the observed values are to the predicted, the less likely the proportional hazards assumption has been violated.



2. Tests Based on the Schoenfeld Residuals

stphptest can be used after stcox to test the proportional hazard assumptions based on Schoenfeld residuals. Use of this command requires that you previously specified

stcox's schoenfeld() option (if the global test is desired) and/or stcox's scaledsch() option (if the detailed test is desired).

```
stcox treat, schoenfeld(sc*)
stphtest
```

Test of proportional hazards assumption

```
Time: Time
-----
          |                chi2      df      Prob>chi2
-----+-----
global test |                2.81       1       0.0939
-----
```

```
. stcox treat renal, schoenfeld(sc*) scaledsch(ssc*)
```

```
-----
      _t | Haz. Ratio  Std. Err.      z    P>|z|    [95% Conf. Interval]
-----+-----
      treat |   3.466265   2.077394    2.07  0.038    1.070833    11.22024
      renal |  60.67034  70.65261    3.53  0.000    6.190477   594.6052
-----
```

```
. stphtest, deta
```

Test of proportional hazards assumption

```
Time: Time
-----
          |      rho      chi2      df      Prob>chi2
-----+-----
treat      |   0.26505    1.51       1       0.2195
renal      |   0.08993    0.25       1       0.6171
-----+-----
global test |                1.52       2       0.4667
-----
```

If the value of Prob>chi2 less than 0.05, the proportional hazard assumption does Not hold in this example.

If a covariate fails the test of proportional hazards, this is evidence that it is time-dependent and one needs to abandon ordinary Cox regression in favor of Cox regression with time-dependent covariates. One may include a time-covariate interaction term in the model.

Another approach to non-proportionality is stratification, a method that is useful when the variable is both categorical and not of direct research interest. Note that there is no coefficient estimated for the strata variable. The partial likelihood for the stratified data is the product of the partial likelihood for each stratum.

```
. stcox treat, strata(renal)
```

```
-----
```

_t	Haz. Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
treat	4.323158	2.851764	2.22	0.026	1.186609	15.75051

```
-----
```

Stratified by renal