Week 13

Hazard Models II

Rich Frank
University of New Orleans
November 15, 2012

Today

- Review non-parametric and parametric models
- Discuss semi-parametric models (e.g. Cox)
- Discrete data models
- How to handle time-varying covariates
- How to chose the "right" model
- Interpreting results
- Testing the proportional hazard assumption

- “Finding positive time dependence is like observing rust on a car: the older it gets, the more there is, and the likelier it is to fall apart.”

(Alt and King 1994: 193)

Hazard model review

- Like event counts, **duration must be greater than or equal to zero**.
- Survival at time *t* means that you have survived since *t-1*. This means that observations at time *t* are **conditional** on observations at time *t-1*.
- Some observations will survive the length of the study. These observations are considered **censored**.

- Why could we just not run a logit model where all observations are considered 0 until death, which is coded 1?

- Because of the **conditionality** of the observations.

- Therefore what we need is to figure out the conditional probability of t_i given the fact that a unit survived to t_i - 1.

- This can be done within the logit framework using conditional logit or other IVs that model time dependence.

Two Types of Hazard Models

- Continuous
  - Failure can happen and be captured at *any* time.
- Discrete
  - Observations are captured within certain regular measures of time (days, months, years).
  - The data we have are likely to be discrete time data.
  - However, as we will see there is a substantial overlap between the two approaches.

To review…

- $f(t)$ can be considered as the estimate of the instantaneous probability that the event (a dispute) occurs.

$$f(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t}$$

- Thus as Δt gets infinitesimally small you get an instantaneous estimate of the probability of failure at time t.

- The survivor function, $S(t)$—the probability of surviving at time t.

- As you might guess the chances of surviving past t are related to the chance of dying at t.

$$S(t) = 1 - F(t) = P(T \geq t)$$

- Therefore the failure and the survival rates are related to each other.

- This relation is given by the hazard rate.

$$h(t) = \frac{f(t)}{S(t)}$$

- In words, the hazard rate is the conditional failure rate—the rate that units fail by *t* given that they have survived until *t*.

- The (continuous) cumulative hazard function is given by:

$$H(t) = \int_0^t h(t)dt$$

Which can also be (and more often is) seen as:

$$H(t) = -\ln\{S(t)\}$$

- Last week we started with non-parametric models including the Kaplan-Meier estimator.
  - A non-parametric estimate of the survival function:

$$\hat{S}(t) = \prod_{j \mid t_j \leq t} \left(\frac{n_j - d_j}{n_j} \right)$$

Where n_j is the number of individuals at risk at time t_j, and d_j is number of failures at time t_j

- Last week, we also looked at parametric models like the exponential, the Weibull, and the Gompertz.

- Box-Steffensmeier and Jones (2004) also talk about several other parametric models.
  - Log-logistic
  - Log-normal

Why would we turn from parametric models?

- Because they require that we make assumptions about shape of the baseline hazard.
- What if we could use a method that would enable us to make no assumption about what the baseline hazard is—and let the data speak as to what the hazard rate is?
- That way we could focus on the effects of our main IVs without worrying about what would happen if we made the wrong assumption as to what the baseline hazard is.
- That would be cool!

Cox model

- In order to estimate the parameters, Cox (1972,1975) developed what he called partial likelihood.

- It is called partial because not all the information available in the data is being used.

- We are not estimating $(h_0(t))$.

- Let's remember what our data look like

Data sorted by ordered failure time

dyad	*Duration Time*	*Censored Case?*
2020	4	No
2091	16	No
345710	28	No
485623	30	Yes
301587	46	No
200356	51	Yes

Note: Data are sorted by duration time. The duration time for censored cases is the time in the last observation

Relevant data characteristics

- Events can be ordered.
- At t_0 all cases are at risk of failing.
- After the first failure, the risk set decreases by 1.
- As failures occur, the risk set keeps on getting smaller.
- For now we are assuming that there are no ties.
  - No dyads have disputes in the same time unit.

Partial Likelihood function

- The risk set ($R(t_i)$) represents the number of cases that are at risk of failing at time t_i.
- The probability that the *j*th case will fail at time T_i is given by:

$$\mathrm{P}(t_j = T_i \mid R(t_i)) = \frac{e^{\beta x_i}}{\sum_{j \in R(t_i)} e^{\beta x_j}}$$

- Which gives us the following log-likelihood function:

$$\ln(L_p) = \sum_{i=1}^{K} \delta_i \left[\beta x_i - ln \sum_{j \in R(t_i)} e^{\beta x_j}\right]$$

- Where δ_i is the censoring indicator (δ_i =1 if uncensored)

Cox assumptions

- The estimated partial maximum likelihood parameters are asymptotically normal, asymptotically efficient, consistent, and invariant (Box-Steffensmeier and Jones 2004: 52).
- However, note that there is no information included about the actual survival times, t_i , for both uncensored and censored cases.
- This is important because it means that the estimation takes no information from observations where there is no failure (censored cases) except in the risk set (the denominator not the numerator).

Risk set

- This partial likelihood is equivalent to

$$L_p = \frac{\Psi(7)}{\Psi(1)+\Psi(2)+\Psi(3)+\Psi(4)+\Psi(5)+\Psi(6)+\Psi(7)} * \frac{\Psi(4)}{\Psi(1)+\Psi(2)+\Psi(3)+\Psi(4)+\Psi(5)+\Psi(6)} * \frac{\Psi(5)}{\Psi(1)+\Psi(2)+\Psi(3)+\Psi(5)+\Psi(6)} * \frac{\Psi(3)}{\Psi(1)+\Psi(2)+\Psi(3)+\Psi(6)} * \frac{\Psi(1)}{\Psi(1)+\Psi(2)+\Psi(6)} * \frac{\Psi(2)}{\Psi(2)+\Psi(6)} * \frac{\Psi(6)}{\Psi(6)}.$$

Where $\Psi = e^{\beta x_i}$

- As you can see, this risk set cannot handle ties, because its focus is on ordering the failure times.

- If there are two failures in a particular time, it cannot tell which one failed first.

- Box-Steffensmeier and Jones (2004) discuss the four most popular ways of dealing with ties…

Breslow Method

- The default in Stata. Its popularity is attributed to its ease of computation.
- Say that both case 1 and case 2 fail at the same time.
- The Breslow method figures the likelihood of both:

$$L_1 = \frac{\Psi(1)}{\Psi(1) + \Psi(2) + \Psi(6)}$$

$$L_2 = \frac{\Psi(2)}{\Psi(1) + \Psi(2) + \Psi(6)}$$

Breslow likelihood function

$$L_{Breslow} = \prod_{i=1}^{k} \frac{e^{\beta si}}{\left[\sum_{j\in R(t_i)} e^{\beta x_j}\right]^{d_i}}$$

Where:

- s_i is the sum of the covariates for the tied cases.
- d_i is the number of tied cases at time t_i.

Efron Method

- This method differs from the Breslow in that it takes into account how the second risk set would differ depending on which case failed first.

- Breslow assumes the same risk set for both cases.

- However, the Efron method does assume that both orderings are equally likely.

$$L_{Efron} = \prod_{i=1}^{k} \frac{e^{\beta si}}{\prod_{r=1}^{d_i} \left[\sum_{j \in R(t_i)} e^{\beta x_i} - (r-1) d^{-1} \sum_{j \in D(t_i)} e^{\beta x_j} \right]}$$

Average Likelihood

- A more exact method for looking at all possible risk sets for ties.
- The likelihood function too complex for even Box-Steffensmeier and Jones to demonstrate.
- The important thing to remember is that it looks at all possible combinations of tie orderings.

Exact Discrete Method

- This method assumes that the cases actually do fail at the same time.

- In brief, the exact discrete method is identical to the conditional logit.

- Let’s see how these models work in the wild.

- To begin, let’s start with Alt and King’s (1994) data on cabinet failure.

```
. estout exponential weibull gompertz cox, cells(b(star fmt(3)) se(par)) ///
>    stats(N  ll  p) legend replace
----------------------------------------------------------------------------
              exponential        weibull       gompertz            cox
                     b/se           b/se           b/se           b/se
----------------------------------------------------------------------------
main
fract              -0.001         -0.001          0.001          0.001
                  (0.001)        (0.001)        (0.001)        (0.001)
polar              -0.016**       -0.015**        0.022***       0.019**
                  (0.006)        (0.005)        (0.006)        (0.006)
format             -0.091*        -0.087*         0.124**        0.107*
                  (0.046)        (0.035)        (0.044)        (0.046)
invest             -0.369**       -0.330**        0.346**        0.423**
                  (0.139)        (0.107)        (0.128)        (0.140)
numst2              0.515***       0.464***      -0.611***      -0.571***
                  (0.129)        (0.100)        (0.126)        (0.132)
eltime2             0.723***       0.664***      -0.891***      -0.818***
                  (0.135)        (0.104)        (0.135)        (0.141)
caretk2            -1.300***      -1.318***       1.704***       1.529***
                  (0.260)        (0.201)        (0.265)        (0.280)
_cons               3.725***       3.696***      -4.337***
                  (0.631)        (0.492)        (0.576)
----------------------------------------------------------------------------
ln_p
_cons                              0.261***
                                 (0.050)
----------------------------------------------------------------------------
gamma
_cons                                             0.051***
                                                (0.005)
----------------------------------------------------------------------------
N                 314.000        314.000        314.000        314.000
ll               -425.069       -412.932       -363.845      -1298.828
p                   0.000          0.000          0.000          0.000
----------------------------------------------------------------------------
```

- As you can see, the results are similar although there are differences in the estimated coefficients and standard errors.

- Let’s dig a bit deeper into the different Cox specifications

Cox with different tie methods

```
. estout Breslow Efron Averaged ExactD, cells(b(star fmt(3)) se(par)) ///
>    stats(N  ll  p) legend replace

----------------------------------------------------------------------------
                  Breslow          Efron         Averaged          ExactD
                     b/se           b/se             b/se            b/se
----------------------------------------------------------------------------
fract               0.001          0.001            0.001           0.001
                  (0.001)        (0.001)          (0.001)         (0.001)
polar               0.019**        0.020**          0.020**         0.021**
                  (0.006)        (0.006)          (0.006)         (0.006)
format              0.107*         0.111*           0.111*          0.116*
                  (0.046)        (0.046)          (0.046)         (0.048)
invest              0.423**        0.432**          0.433**         0.452**
                  (0.140)        (0.139)          (0.139)         (0.145)
numst2             -0.571***      -0.584***        -0.586***       -0.622***
                  (0.132)        (0.132)          (0.132)         (0.138)
eltime2            -0.818***      -0.846***        -0.847***       -0.872***
                  (0.141)        (0.141)          (0.141)         (0.146)
caretk2             1.529***       1.695***         1.719***        1.840***
                  (0.280)        (0.283)          (0.286)         (0.327)
----------------------------------------------------------------------------
N                 314.000        314.000          314.000         314.000
ll              -1298.828      -1286.678         -917.263        -917.399
p                   0.000          0.000            0.000           0.000
----------------------------------------------------------------------------
* p<0.05, ** p<0.01, *** p<0.001
```

Interpretation

- So we have run a few models.
- The Exact discrete method model has the largest estimated coefficients.
- How do we interpret what the output tells us?
- Let's start by looking at the values of our IVs to remind ourselves which ones are continuous and dichotomous.

```
. sum fract polar format invest numst2 eltime2 caretk2 if e(sample)

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       fract |       314    718.8121    87.98591        349        868
       polar |       314    15.28981    12.81406          0         43
      format |       314    1.904459    1.413235          1          8
      invest |       314    .4522293    .4985072          0          1
      numst2 |       314    .6305732      .48342          0          1
-------------+--------------------------------------------------------
     eltime2 |       314    .4872611    .5006355          0          1
     caretk2 |       314    .0541401    .2266552          0          1
```

The Cox with Efron Method

```
.    stcox fract polar format invest numst2 eltime2 caretk2, efron nohr

         failure _d:  ciep12
   analysis time _t:  durat

Iteration 0:   log likelihood =  -1369.664
Iteration 1:   log likelihood = -1307.7966
Iteration 2:   log likelihood = -1286.7898
Iteration 3:   log likelihood = -1286.6782
Iteration 4:   log likelihood = -1286.6782
Refining estimates:
Iteration 0:   log likelihood = -1286.6782

Cox regression -- Efron method for ties

No. of subjects =          314                     Number of obs   =       314
No. of failures =          271
Time at risk    =       5789.5
                                                   LR chi2(7)      =    165.97
Log likelihood  =   -1286.6782                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       fract |   .0013349   .0009231     1.45   0.148    -.0004743    .0031442
       polar |   .0198011   .0061485     3.22   0.001     .0077503     .031852
      format |   .1109851   .0459212     2.42   0.016     .0209813     .200989
      invest |   .4324272   .1393422     3.10   0.002     .1593215    .7055329
      numst2 |  -.5839958   .1323322    -4.41   0.000    -.8433622   -.3246295
     eltime2 |  -.8461374   .1407433    -6.01   0.000    -1.121989   -.5702857
     caretk2 |   1.694705   .2826007     6.00   0.000     1.140817    2.248592
------------------------------------------------------------------------------
```

- As we saw last week, the coefficients can be seen as hazard rates.

- To derive the hazard ratio all you have to do is exponentiate the coefficient.

- For the investiture variable, the estimated hazard ratio is:

```
.    display exp(_b[invest])
1.5409933
```

- This suggests that the risk of cabinet failure is about 1.54 (or about 1 ½ times) greater when there is an investiture requirement (1) than when there is not (0).

- What about whether it forms after an election?

```
.    display exp(_b[eltime2])
.42906907
```

- This suggests that if the government was formed immediately after an election, the risk is about 43% of those cases where there were longer, and more protracted negotiations necessary to form a government.

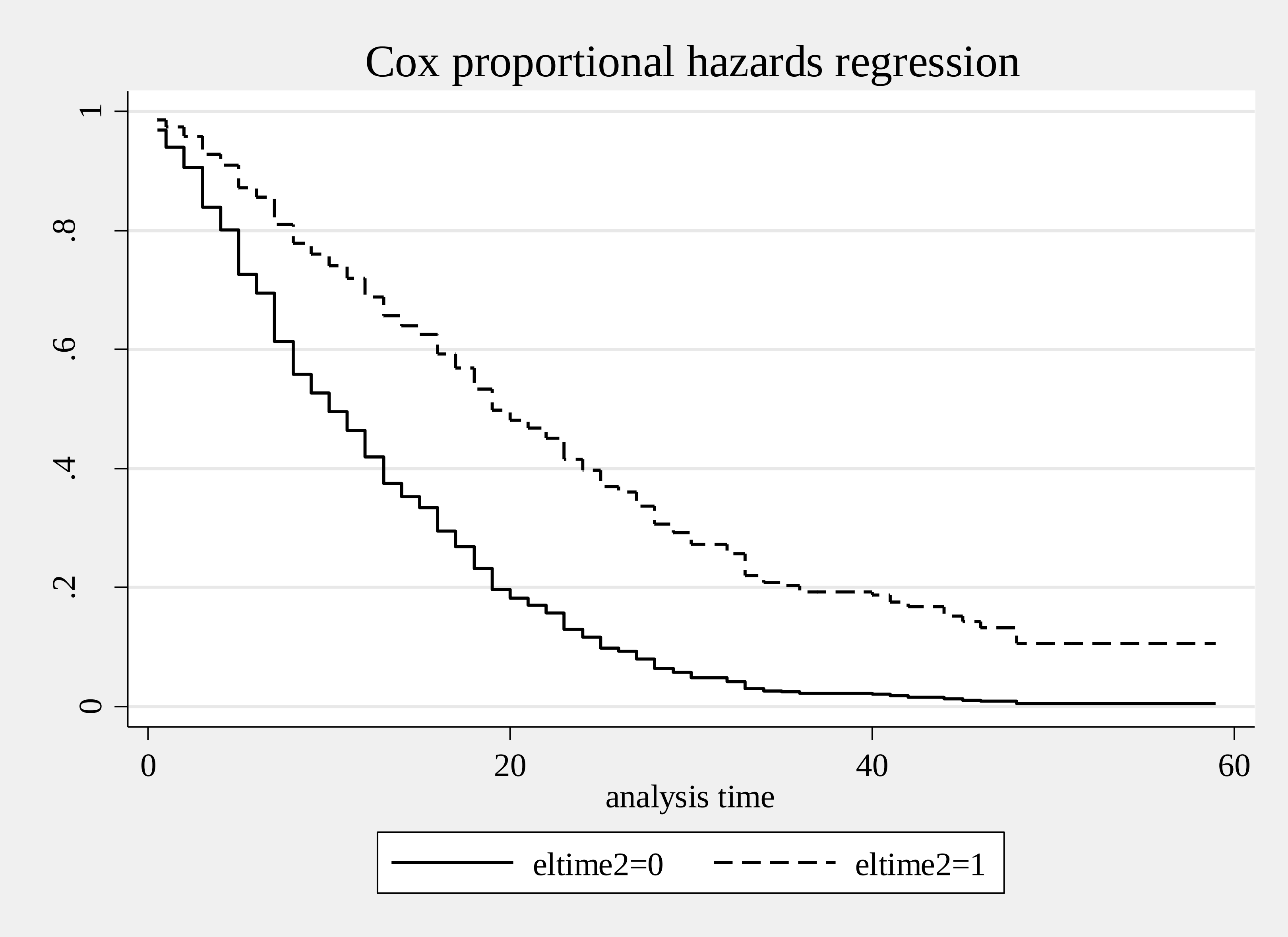
- We can also get the risk when eltime2=0

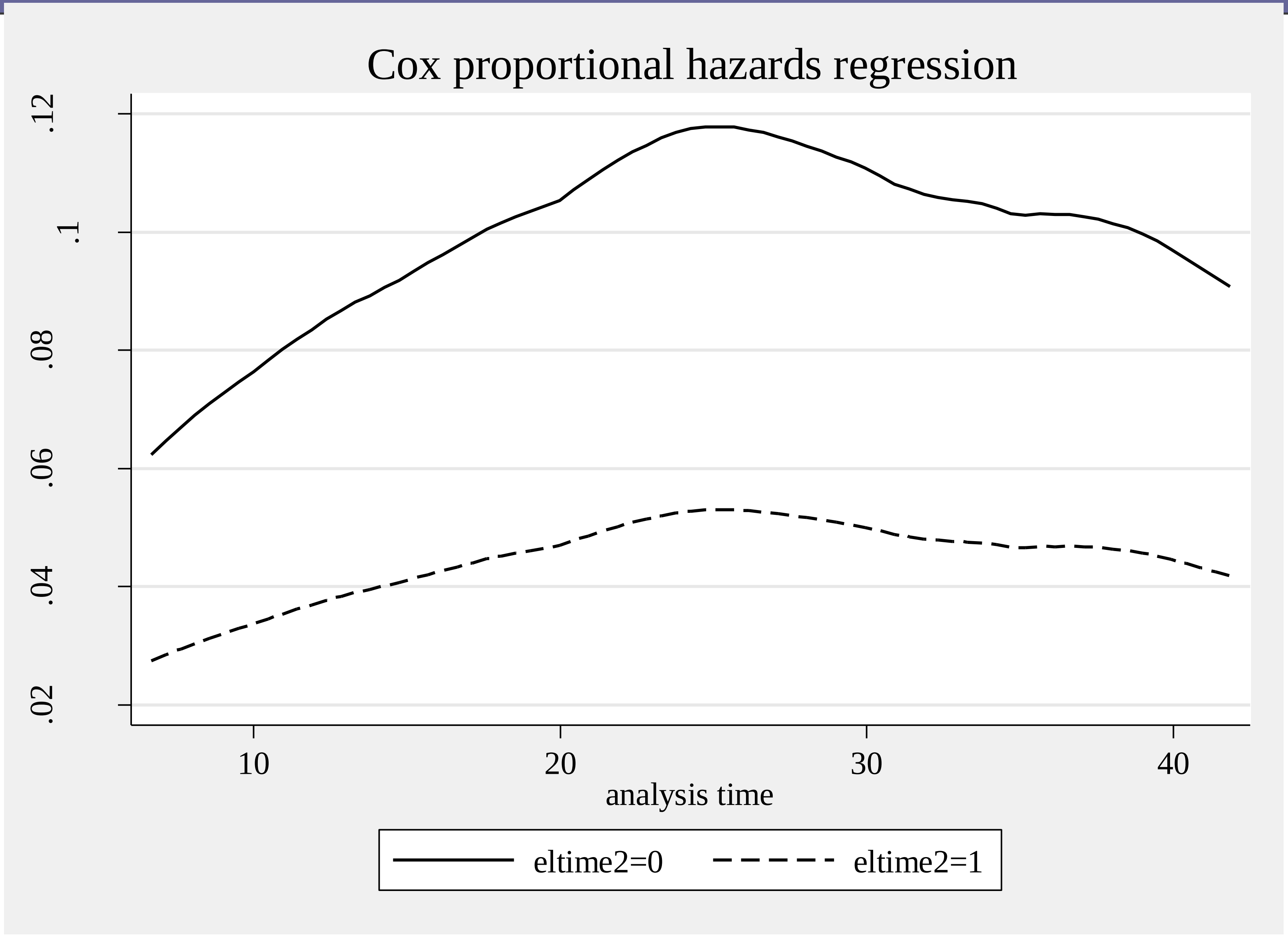
```
.    display 1/exp(_b[eltime2])
2.3306271
```

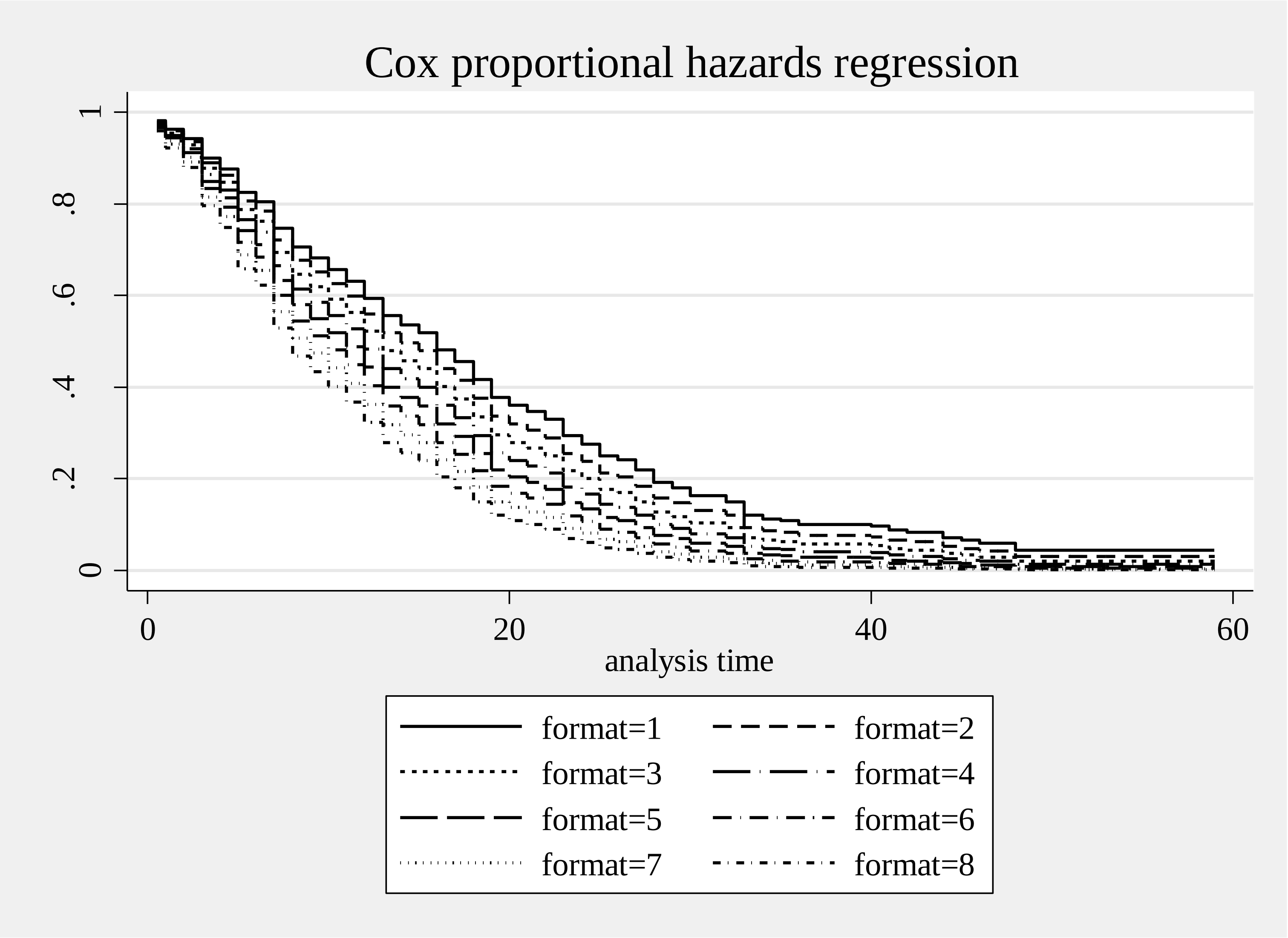
- This tells us that the risk of failure is 2.4 times greater than cases where eltime2=1

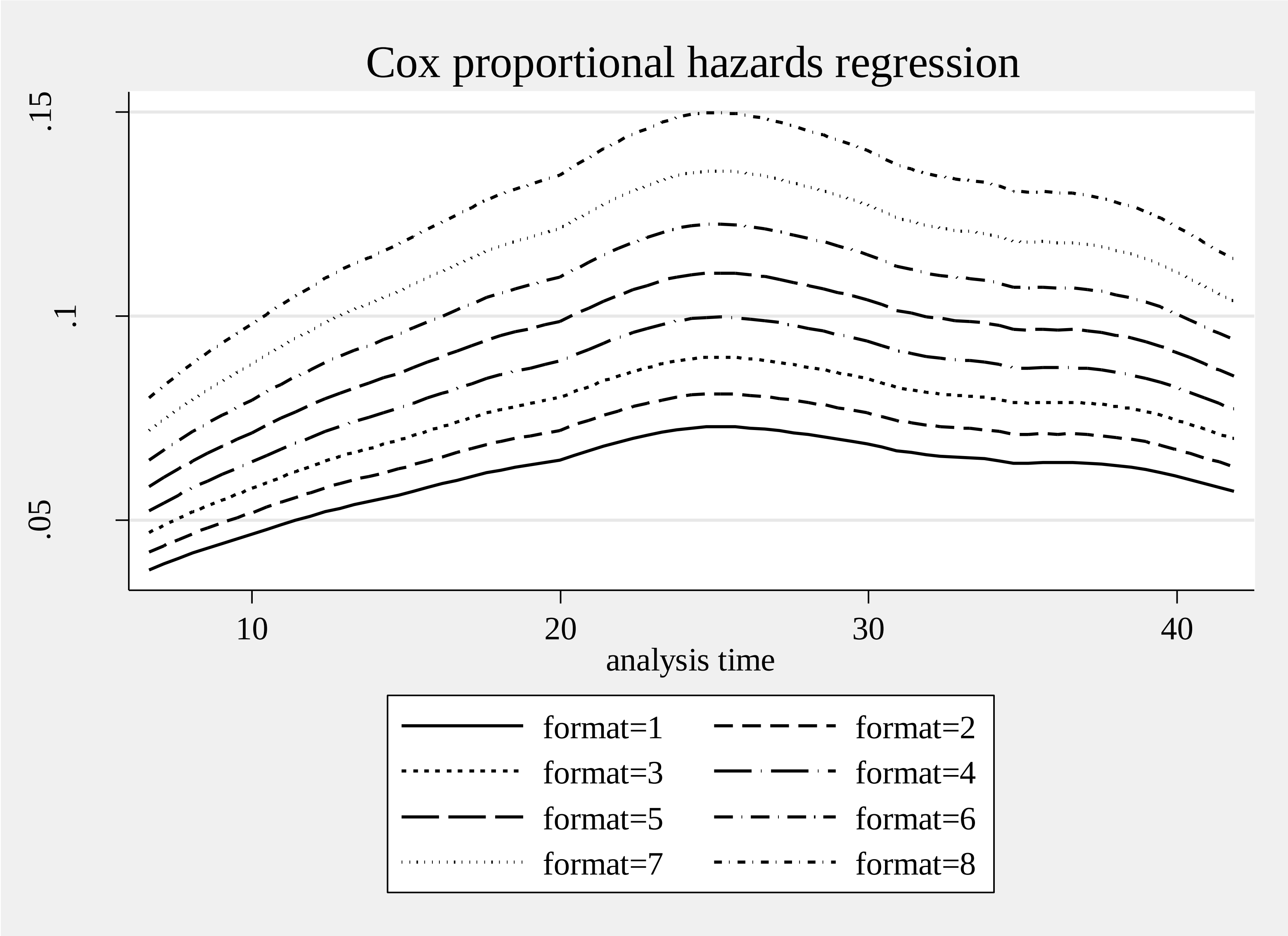
Or you can estimate the HR directly

```
. stcox fract polar format invest numst2 eltime2 caretk2, efron

         failure _d:  ciep12
   analysis time _t:  durat

Iteration 0:   log likelihood =  -1369.664
Iteration 1:   log likelihood = -1307.7966
Iteration 2:   log likelihood = -1286.7898
Iteration 3:   log likelihood = -1286.6782
Iteration 4:   log likelihood = -1286.6782
Refining estimates:
Iteration 0:   log likelihood = -1286.6782

Cox regression -- Efron method for ties

No. of subjects =          314                     Number of obs   =        314
No. of failures =          271
Time at risk    =       5789.5
                                                   LR chi2(7)      =     165.97
Log likelihood  =   -1286.6782                     Prob > chi2     =     0.0000

------------------------------------------------------------------------------
          _t | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       fract |   1.001336   .0009243     1.45   0.148     .9995258    1.003149
       polar |   1.019998   .0062715     3.22   0.001      1.00778    1.032365
      format |   1.117378   .0513113     2.42   0.016     1.021203    1.222611
      invest |   1.540993   .2147254     3.10   0.002     1.172715    2.024925
      numst2 |   .5576656   .0737971    -4.41   0.000     .4302614    .7227951
     eltime2 |   .4290691   .0603886    -6.01   0.000     .3256314    .5653639
     caretk2 |   5.445038   1.538772     6.00   0.000     3.129325    9.474386
------------------------------------------------------------------------------
```


Cox proportional hazards regression
0
.2
.4
.6
.8
1
0
20
40
60
analysis time
eltime2=0
eltime2=1


Cox proportional hazards regression
.02
.04
.06
.08
.1
.12
10
20
30
40
analysis time
eltime2=0
eltime2=1


Cox proportional hazards regression
0
.2
.4
.6
.8
1
0
20
40
60
analysis time
format=1
format=2
format=3
format=4
format=5
format=6
format=7
format=8



Cox proportional hazards regression
.05
.1
.15
10
20
30
40
analysis time
format=1
format=2
format=3
format=4
format=5
format=6
format=7
format=8

- To look at how a particular variable affects the hazard rate:

$$\%\Delta h(t) = \left[\frac{e^{\beta(x_i = X_1)} - e^{\beta(x_i = X_2)}}{e^{\beta(x_i = X_2)}}\right] * 100$$

Where x_i is the covariate (e.g. post-election) and the X_1 and X_2 are two values of x_i.

- Let's look at how polarization can affect the hazard rate.

```
. codebook polar if e(sample)

-------------------------------------------------------------------------------------polar
Polarization Index
-------------------------------------------------------------------------------------
                  type:  numeric (float)

                 range:  [0,43]                       units:  1
         unique values:  34                       missing .:  0/314

                  mean:   15.2898
              std. dev:   12.8141

           percentiles:        10%       25%       50%       75%       90%
                                 0         3      14.5        25        36

.
. * Change from mean to 1sd above mean **
. display((exp(_b[polar]*28.1039 )-exp(_b[polar]*15.2898))/exp(_b[polar]*15.2898))*100
28.882878

. ** to 1sd below mean
. display((exp(_b[polar]*2.4757 )-exp(_b[polar]*15.2898))/exp(_b[polar]*15.2898))*100
-22.410175
```

- As you can see, there are a number of different ways of trying to look at the substantive effect of your independent variables on the hazard of failure.

- However, underscoring the Cox model (as well as the Weibull) is the assumption that the hazards are proportional.

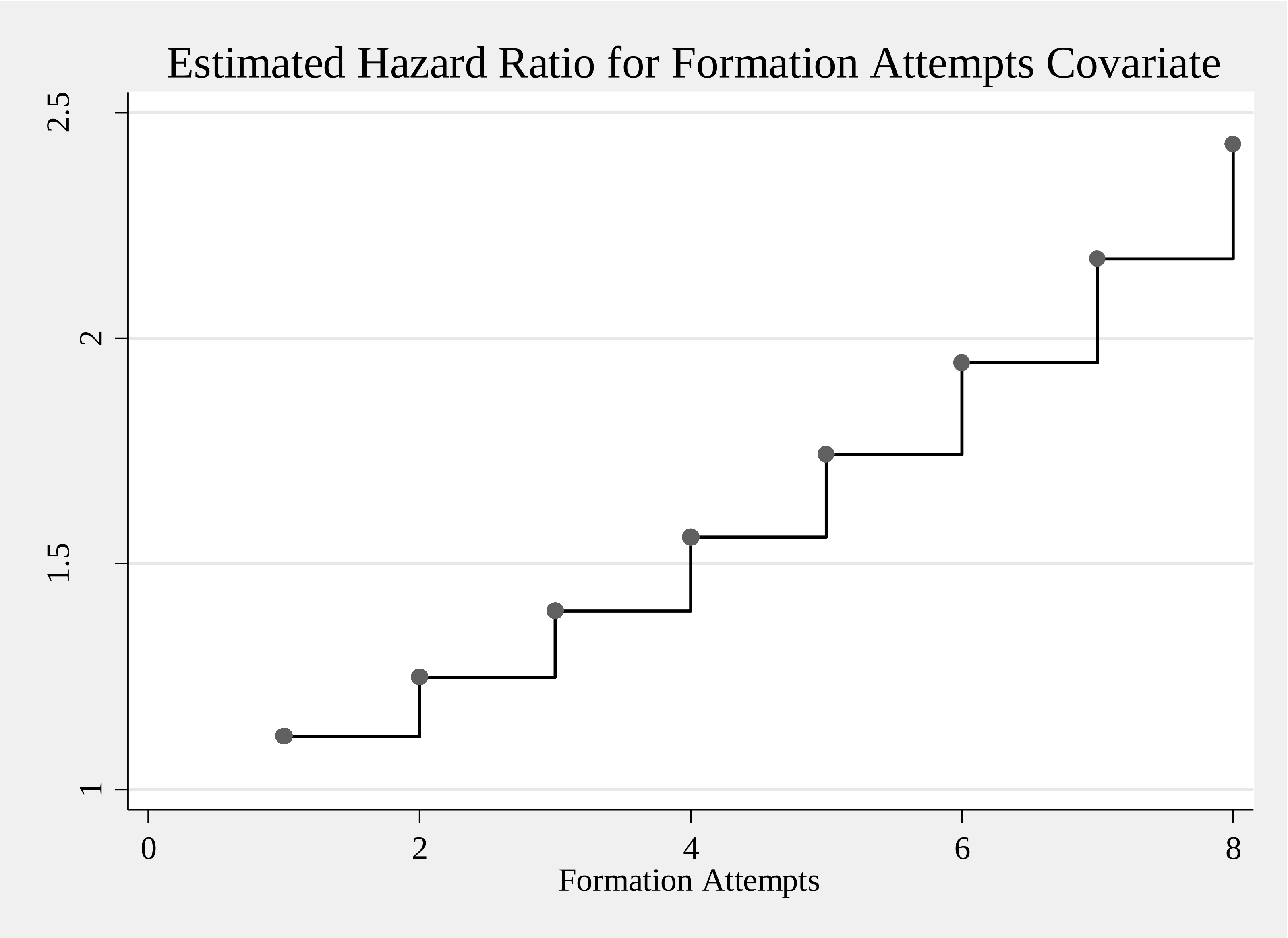
Let's look at the prop. hazards assumption

```
. tab format

  Number of |
  Formation |
 Attempts / |
  1-7 or 8+ |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |        179       57.01       57.01
          2 |         63       20.06       77.07
          3 |         36       11.46       88.54
          4 |         14        4.46       92.99
          5 |         12        3.82       96.82
          6 |          5        1.59       98.41
          7 |          1        0.32       98.73
          8 |          4        1.27      100.00
------------+-----------------------------------
      Total |        314      100.00
```

```
.    gen format_ratio=exp(_b[format]*format)

.    tab format_ratio

format_rati |
          o |      Freq.     Percent        Cum.
------------+-----------------------------------
   1.117378 |        179       57.01       57.01
   1.248534 |         63       20.06       77.07
   1.395085 |         36       11.46       88.54
   1.558838 |         14        4.46       92.99
   1.741812 |         12        3.82       96.82
   1.946262 |          5        1.59       98.41
   2.174711 |          1        0.32       98.73
   2.429975 |          4        1.27      100.00
------------+-----------------------------------
      Total |        314      100.00
```

- Like with other PH models, the ratio of each value to the adjacent one is identical…

```
.    display  1.395085/1.248534
1.1173785

.    display  1.741812/1.558838
1.1173785

.    display 2.429975/2.174711
1.1173784
```


Estimated Hazard Ratio for Formation Attempts Covariate
2.5
2
1.5
1
0
2
4
6
8
Formation Attempts

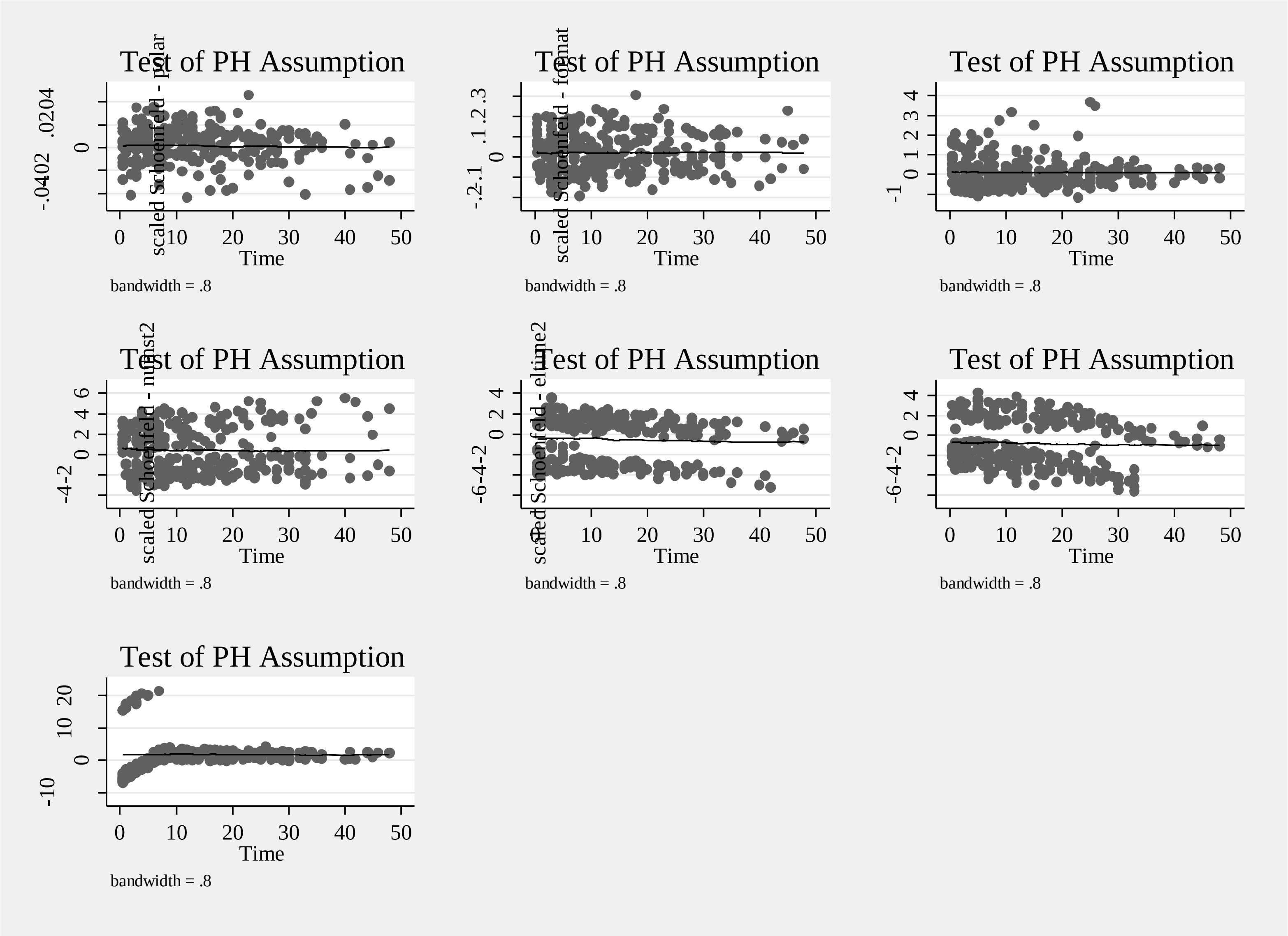
Proportional hazards property

- Same as the Weibull.

- Changes to the baseline hazard rate in the exponential are in multiples of the baseline hazard.

$$\frac{h_i(t\,|x_1\!=\!1)}{h_i(t\,|x_1\!=\!0)} = e^{-\beta_1}$$

Testing the PH assumption: Schoenfeld Residuals



```
stcox fract polar format invest numst2 eltime2
caretk2, nohr efron schoenfeld(sc*)
scaledsch(ssc*)

stphtest, plot(fract  ) saving(sc1, replace )
stphtest, plot( polar ) saving(sc2, replace )
stphtest, plot(format  ) saving(sc3, replace )
stphtest, plot(invest  ) saving(sc4 , replace)
stphtest, plot(numst2  ) saving(sc5, replace )
stphtest, plot(eltime2  ) saving(sc6 , replace)
stphtest, plot(caretk2 ) saving(sc7, replace )
graph combine sc1.gph sc2.gph sc3.gph sc4.gph
sc5.gph sc6.gph sc7.gph
```

- The visual test of the Shoenfeld residuals relies largely on the eye of the beholder to decide of the PH assumption is met.

Testing the PH assumption

```
. stphtest, detail

      Test of proportional-hazards assumption

      Time:  Time
      ----------------------------------------------------------------
                  |       rho            chi2       df       Prob>chi2
      ------------+---------------------------------------------------
      fract       |     -0.10251         2.93        1         0.0868
      polar       |     -0.01391         0.05        1         0.8265
      format      |     -0.04127         0.45        1         0.5039
      invest      |     -0.01072         0.03        1         0.8615
      numst2      |     -0.07590         1.68        1         0.1947
      eltime2     |     -0.06889         1.35        1         0.2457
      caretk2     |      0.00063         0.00        1         0.9913
      ------------+---------------------------------------------------
      global test |                      8.83        7         0.2653
      ----------------------------------------------------------------
```

- What does a significant rho tell us?
- It suggests that the correlation between the scaled Schoenfeld residuals and the rank of the survival time is significantly different than 0.
- The null is that rho equals 0.
- What can we do if there is some time correlation?
- The most frequent method (used by Box-Steffensmeier, Reiter, and Zorn 2003) is to interact some function of time with the problematic variable.

- I tried the one variable that could be problematic, the fractionalization index, but the likelihood curve was flat, and Stata would not estimate it.

Which model is "best"

- The short answer: it depends…
- It is easier with parametric models to look at the estimated hazard function with the generalized gamma or the Weibull to see what direction the baseline hazard is going.
- With parametric and semi-parametric models one method is using the Akaike Information Criterion (AIC) for non-nested models.

AIC

$$AIC = -2(\log L) + 2(c + p + 1)$$

- Where c is the number of independent variables
- And p is the number of structural parameters

- The idea is to "reward parsimonious models" by adding to the LL for every parameter estimated

AIC

Model	Log-Likelihood	AIC
Exponential	-425.07	866.14
Gompertz	-363.84	745.69
Weibull	-412.93	843.86
Cox with Breslow	-1298.83	2611.66
Cox with Efron	-1286.68	2587.35
Cox with Average L	-917.26	1848.52
Cox with Exact D	-917.40	1848.80

Baseline hazard in Cox

- What if we do want to look at the baseline hazard function $\widehat{h_i}(t)$ after running a Cox model?
- Box-Steffensmeier and Jones (2004) discuss a method suggested by Kalbfleisch and Prentice (1973, 1980) for recovering the baseline hazard after the Cox.

$$\widehat{h_i}(t) = -\ \mathrm{ln}\hat{S}_o(t) = -\sum_{i=1}^{n} ln\widehat{\alpha}_i$$

Generating Baseline functions

- To obtain a natural 0 point in the IVs, we can mean center `polar` and `format`.

```
. egen meanpolar=mean(polar)
. gen polarmean=polar-meanpolar
. egen meanform=mean(format)
.  gen formmean=format-meanform
```

```
. stcox invest polarmean numst formmean eltime2 caretk2, nohr exactm
         failure _d:  ciep12
   analysis time _t:  durat

Iteration 0:   log likelihood = -1000.4512
Iteration 1:   log likelihood = -937.95835
Iteration 2:   log likelihood = -918.35198
Iteration 3:   log likelihood = -918.30728
Iteration 4:   log likelihood = -918.30727
Refining estimates:
Iteration 0:   log likelihood = -918.30727

Cox regression -- exact marginal likelihood

No. of subjects =          314                     Number of obs   =       314
No. of failures =          271
Time at risk    =       5789.5
                                                   LR chi2(6)      =    164.29
Log likelihood  =   -918.30727                     Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      invest |   .3881819   .1372435     2.83   0.005     .1191896    .6571743
   polarmean |   .0234117   .0056238     4.16   0.000     .0123893    .0344341
      numst2 |  -.5842937   .1323193    -4.42   0.000    -.8436348   -.3249526
    formmean |   .1303069      .0439     2.97   0.003     .0442643    .2163494
     eltime2 |  -.8623465   .1407457    -6.13   0.000    -1.138203     -.58649
     caretk2 |   1.735869   .2868057     6.05   0.000      1.17374    2.297997
------------------------------------------------------------------------------
```

- Actually estimating the baseline hazard in Stata requires continuing on with a series of steps that are outside the scope of this class.
- They can be done, and Brad Jones has kindly put the do-file to do so on his website.
- An example can be seen on pages 64-66 of Box-Steffensmeier and Jones (2004).

- There are a number of extensions to the basic Cox model and conditional logit to better model the data generating process:
  - Frailty models (e.g. Aydin 2010)
  - Repeated events
  - Competing risks
  - Non-proportional hazards (Box-Steffensmeier, Reiter, and Jones 2004)

- Let’s look at several of these articles in more detail…