

Semi-Parametric Models

Use the data set called `coalum2.dta`. This is the coalition data set used by King et al. in their 1990 paper except that the variable names have been improved and labels have been provided. It should now be obvious what each of the variables are.

Exercise 1: Piecewise-Constant Exponential Model

1. The data set does not include an id variable. You should create one.
2. Now `STSET` the data using `CENSOR12` as the censoring variable and your new id variable for the ID option.
3. The maximum duration in the data set is 59 months. The duration data are single-spell data. Split each spell into six time intervals of length 10 months i.e. $[0, 10)$, $[10, 20)$... $[50, \infty)$.
4. Now generate six time dummies, `t1`, `t2`, ..., `t6`, for each of the six intervals of time.
5. Estimate a piecewise-constant exponential model where the independent variables are `MAJORITY_GOVERNMENT` `VOLATILITY` `POLARIZATION` `FRACTIONALIZATION` `CRISIS` `FORMATION_ATTEMPTS` `OPPOSITION_PARTY` `INVESTITURE`. Include the six time dummies and `noconstant`. Put the results in the first column of 1.
6. Interpret the coefficients on `MAJORITY_GOVERNMENT` and `INVESTITURE`. Now exponentiate the coefficients and interpret them.

Exercise 2: Cox Model

1. Estimate a Cox model where the independent variables are the same as before (not including the time dummies). Put the results in the second column of Table 1. Compare and contrast these results with those from the PCE model. Compare and contrast them to the results from the parametric models that we looked at last time as well.
2. Why is the log-likelihood from the Cox model so different from the log-likelihood from the parametric models? Why is there no constant?
3. Interpret the coefficients on `MAJORITY_GOVERNMENT` and `INVESTITURE`.
4. Now estimate the Cox model but get STATA to report the exponentiated coefficients. Interpret the exponentiated coefficients on `MAJORITY_GOVERNMENT` and `FORMATION_ATTEMPTS`. How much would the hazard rate increase if a government's fractionalization score increased by 20 units? How much would the hazard rate increase if the ideological polarization in the legislature increased by 10 units?
5. Experiment with the 4 different methods for dealing with ties (Breslow, Efron, and the 2 exacts). Put the results in the remaining columns of Table 1. Do the different ways of handling ties matter?

6. Use STATA to plot the baseline cumulative hazard function and the baseline survivor function.
7. Use STATA to plot the baseline hazard against duration. Unlike with the parametric models, the baseline hazard will be jerky. Why? In reality, the baseline hazard should not be jerky. As a result, you should use the LOWESS command to smooth the baseline hazard. The plot that you provide should show the jerky baseline hazard and the smoothed baseline hazard together.
8. Multiply the DURATION variable by two. Now rerun the basic Cox model from earlier (using Breslow). You will have to re-STSET your data. What do you notice about the results? Why does this happen?

Table 1: Models of Government Coalition Duration

Regressor	PCE	Cox Breslow	Cox Efron	Cox continuous	COX discrete
MAJORITY					
VOLATILITY					
POLARIZATION					
FRAGMENTATION					
CRISIS DURATION					
# FORMATION ATTEMPTS					
OPPOSITION CONCENTRATION					
INVESTITURE					
t1					
t2					
t3					
t4					
t5					
t6					
Observations					
Log-Likelihood					

* $p < 0.10$; ** $p < 0.05$; *** $p < 0.01$ (two-tailed); Standard errors are given in parentheses

Exercise 3: Diagnostic Tests – Government Duration Data

1. Estimate a Cox model where the independent variables are the same as before. Use the Efron method for dealing with ties. Plot the Schoenfeld residuals against time for each of the covariates in the model. Is there any evidence that the proportional hazards assumption is violated?
2. Instead of doing the graphical eyeball test, now conduct the formal test of the PH assumption.

Interpret the individual test and global test results. If you found that the proportional hazards assumption was violated, what would you do to solve it?

3. Now let's take a look at the Cox-Snell residuals from the Cox model to evaluate model fit. Graph the Cox-Snell residuals following the code in the lecture notes. Interpret what you find. Now repeat this process of looking at the Cox-Snell residuals for a parametric weibull model.
4. Now let's take a look at the functional form of the POLARIZATION covariate. Plot the martingale residuals against POLARIZATION. Interpret. (Use either of the approaches outlined in the lecture notes).