Workshop on Survival Analysis
Frederick J. Boehmke
Exercise #4: Competing Risks

In this exercise we will learn how to estimate a competing risks version of parametric and semi-parametric duration estimators in Stata. We will then plot the survivor function to see how a covariate can have a different effect on separate risks.

The data again represent information about position taking by members of Congress on NAFTA from Box-Steffensmeier, Arnold, and Zorn. The data indicate the number of days before the NAFTA vote that each member of Congress stated his or her preference on the bill started by President George H.W. Bush and supported by then-President Clinton. The first member announces 463 days before the vote, so that is set as time zero.

To add competing risks to the analysis, we will analyze the time until failure by announcing a vote in favor and the time until failure until announcing a vote in opposition. Once a position is announced we assume that member is right censored for the other position (i.e., no position switching).

Part I

Let's open the data and run the single failure model to set a baseline. Then we'll run the competing risks version.

1. `stset` the data with `timing` as the dependent variable.
2. Graph the Kaplan-Meier curve and the hazard with confidence intervals.
3. Re-run the single failure Cox model we used in the previous exercise:
```
.stcox corptpct labtpct mexbordr dleader rleader ncomact ideol
  pscenter hhcenter, nohr
```
4. Now run the competing risks version of the Cox estimator with this same specification. The two risks are failing by declaring support (vote=1) and failing by declaring opposition (vote=0). The adjustment is made in how you `stset` the data. You'll have to `stset` twice – once for each failure type.
5. Compare the three models. What differences do you see?

Part II

Now let's illustrate how to interpret the effect of a covariate on each of these two risks.

1. We'll illustrate this by exploring the effect of whether a Representative's district borders Mexico. So for each of the two risks we want to generate the predicted hazard with the Mexico border variable set to 0 and then to 1. Use the `at()` option to accomplish this. (For the purposes of this exercise we'll ignore any issues with nonproportional hazards.)

2. Generate and save the data from the predicted survivor function for both risks.

3. Now open and combine the two saved data sets, making sure to rename the variables as needed.

4. Create a graph with all four survivor functions. Think about how to use color and line pattern to accentuate the comparison.

5. How does the effect of the Mexican border variable change over the two risks?