HW: Grouped Binary Duration Data

In the following, you will be expected to analyze real data (as opposed to the monte carlo data we've been using so far). Write your results as if you're writing a lab report. I do not expect a 20 or 30-page paper. However, clearly lead the reader through your model, results, and your interpretation/discussion. Additionally, while I do not expect you to be an expert in these substantive areas, do treat them seriously, as if you were actually conducting research on these topics.

Beck, Katz, & Tucker 1998 (BKT) study the onset of international disputes. In particular, they recommend using a binary data model that incorporated duration dependence, and they then reanalyze data used by Oneal and Russett in their 1997 *ISQ* article. The data (onealrusstt.dta) can be found on my web site. Use BKT as the codebook.

1. Replicate BKT's Table 1. For this, you will need to run the following regressions:

   (a) ordinary logit
   (b) logit with time dummies
   (c) logit with a natural cubic spline
   (d) cloglog with time dummies

   In addition to these, run the following regressions and append them to your table:

   (a) logit with t
   (b) logit with t, t^2, and t^3
   (c) cloglog with natural cubic spline
   (d) cloglog with t, t^2, t^3

2. Now remove continuation dispute years from the dataset and control for the number of prior disputes. Run the following regressions:

   (a) ordinary logit
   (b) logit with natural cubic splines
   (c) cloglog with natural cubic splines
   (d) logit with t, t^2, t^3
   (e) cloglog with natural cubic splines

3. Plot the hazards for the ordinary logit and the "best" model in 2. Demonstrate through the hazard plots whether the models are proportional hazard models.

Note that Table 2 in BKT removes continuing dispute years and Table 3 controls for prior disputes. You may want to replicate those to verify your code and data.

Note that the package `splines` contains a command `ns()` which takes a regressor and returns a length(x) x df basis matrix for the polynomial representation of the natural cubic spline for that regressor.