

# Advanced Maximum Likelihood Estimation

## ICPSR 2017

### Exercise Three

#### Introduction and Data

Box-Steffensmeier et al. (1997) studied the timing of U.S. House members' announcements of their positions on the upcoming North American Free Trade Agreement (NAFTA) enabling legislation. Their theory suggested a set of covariates would be predictive of the timing of members' announcements; in the interest of brevity, I'll simply list those covariates and the direction of the expected marginal covariation with the hazard of announcement (see their Table 1 for more details):

- **pscenter**: The (mean-centered) percentage of the member's district that were members of a private sector labor union. (+)
- **ideol**: A binary (AFL-CIO COPE-based) indicator, coded 1 if the member is ideologically liberal, and 0 if they are conservative. (+)
- **inter1**: the multiplicative interaction of **pscenter** and **ideol**. (−)
- **pecenter**: The (mean-centered) percentage of the vote received by presidential candidate H. Ross Perot in that member's district in the 1992 general election. (−)
- **perotsq**: **pecenter** squared. (+)
- **mexbodr**: An indicator variable, coded 1 if the member's district shared a geographic border with Mexico, and 0 otherwise. (+)
- **hhcenter**: The (mean-centered) median annual household income of individuals in that member's district. (+)
- **inter2**: The multiplicative interaction of **hhcenter** and **ideol**. (−)
- **corptpct**: The percentage of all campaign contributions received by that member in the previous electoral cycle that came from corporate interests. (+)
- **labtpct**: The percentage of all campaign contributions received by that member in the previous electoral cycle that came from labor unions. (+)
- **ncomact**: An indicator variable coded 1 if a member served on a committee to which the NAFTA legislation had been referred, and 0 otherwise. (+)
- **dleader**: An indicator variable coded 1 if the member in question was a member of the Democratic House leadership, and 0 otherwise. (+/−)
- **rleader**: An indicator variable coded 1 if the member in question was a member of the Republican House leadership, and 0 otherwise. (+)

## Assignment

### Part I

In their study, Box-Steffensmeier et al. treat all House members as uncensored (that is, eventually declaring their position, by doing so on the day of the actual vote). They then model the direction of the `vote` itself (`yeas=1`, `nays=0`) separately. An alternative view might consider the possibility that individuals voting “yes” were motivated differently than those voting “no,” with the result that covariates might be differentially related to the two outcomes. This suggests a competing risks framework for analysis.

This part of the assignment is relatively straightforward: Specify, fit, and interpret one or more competing risks models of House members’ NAFTA decision. Data for the analysis are available in the course github repo, in [Stata](#) and [CSV](#) formats. If you are so inclined, you should feel free – but not obligated – to think about how the change to a competing risks framework might alter the specification of the model’s right-hand side (interactions, nonlinearity, etc.). You *are* obligated to use everything you have learned to this point in the completion of the exercise; that means (a) discussion of the statistical and substantive conclusions one might draw from your results, (b) explanation of the choices you made regarding the statistical model(s) you chose, etc., and (c) any appropriate diagnostics.

### Part II

In their article, Box-Steffensmeier et al. discuss the idea that, with respect to NAFTA, districts that bordered Mexico (i.e., those with `mexbodr = 1`) are different in important ways from those that did not. In Part II of this exercise, consider empirically (and assess) the hypothesis that districts bordering Mexico had had differently shaped (but not systematically higher or lower) hazards than those lacking such a border.\* In doing so, fit models to the original (non-competing risks) data (that is, treat all observations as uncensored). Be sure to set out and discuss your findings on this point in substantive terms, using whatever method(s) you feel are appropriate.

This assignment is due on Wednesday, August 16 at 11:59 p.m. EDT. Exercises should be submitted electronically, *in PDF format*, in the usual way. In addition to your responses to the items above, please include all code used to fit models, conduct diagnostics, generate plots, and so forth.

---

\*Hint: This is best done in a parametric formulation, like a Weibull, relying on the discussion from August 14.