Methodology Minor Field Exam

Spring 2017

For the minor field exam, you must answer two questions, one in the morning session and one in the afternoon session. In the afternoon session, you may use the software of your choice and will have access to the internet–which you may use to help you analyze data but NOT to communicate with anyone. You are free to use whatever word processing software you like to write your answers. The questions must be answered in the alloted time.

Morning Session: Statistical Theory and Modeling Decisions

Answer one of the following three questions:

1. *Bayesian Statistics:* Even though Bayes developed his famous theorem in the 18th century, it has only been in the past few decades that Bayesian analysis has become popular in the social sciences. Why has it taken so long for Bayesian modeling to gain acceptance in the social sciences? What are the major criticisms of the Bayesian approach to statistical modeling? What are the strengths of Bayesian modeling, specifically when modeling social science data? How are the quantities of interest derived in the Bayesian context? (That is, describe at least one sampling algorithm used in MCMC.) How are model results from Bayesian estimation related to MLE or OLS estimates? Why would one choose to use Bayesian estimation, from both substantive and methodological perspectives? Finally, discuss how you see Bayesian modeling and Political Science progressing (or not) together in the future. Feel free to illustrate your answer using real or hypothetical examples.

2. *Panel Data:* Analyzing data that are repeated measures over the same units poses a unique set of estimation issues. Assuming a static model (as opposed to a dynamic model), discuss some of the major problems associated with estimating models with panel data while treating them like cross-sectional data (that is, what happens if we pretend each observation is completely independent, as opposed to repeated measures of the same individual). One of the most persistent problems is unit effects and the most commonly employed techniques to address these issues are fixed and random effects models. How do these 2 approaches differ in terms of estimation of the unit effects? How do these approaches differ in terms of substantive interpretation of the results. Finally, discuss the strengths and weaknesses of fixed vs. random effects models and provide examples of when each approach is superior to the other.

3. *Game Theory:* The Nash paradigm has been tremendously important for game theory. Nonetheless, the Nash equilibrium concept has been repeatedly refined. Why has the Nash equilibrium been refined? Alternatively, what do the refinements accomplish that the Nash equilibrium does not? How have extensive form games and games with incomplete information allowed us to incorporate beliefs into our analyses of strategic interactions? By considering actions and beliefs, how are the substantive applications of game theory enhanced?

Afternoon Session: Analyzing Data

Answer one of the following two questions:

(a) Please analyze the data set *incumbent.dta* using a logistic regression model.The data are available here:

`http://spia.uga.edu/faculty_pages/monogan/teaching/incumbent_logit.dta`

The data set contains information on members of the House of Representatives in 1990. The outcome of interest is the variable *returned*–whether the member was returned to the House in 1992, and the input variables are (you must use them all):

- age: The incumbent's age
- surplus: Surplus the incumbent could take home if they retired
- marginal: Dummy-whether the race was close last time
- resistm: Dummy-whether the incumbent's district was redistricted

Present the results of this model in a table including the coefficients, the standard errors, the proportional reduction in error, and any additional information you would like. What can you conclude from the z-ratios associated with each coefficient? What can you conclude from the proportional reduction in error?

Please test the conditional hypothesis that the effect of redistricting varies according to an incumbent's age. Please illustrate the nature of this conditioned relationship using predicted probabilities with confidence intervals. Please assess the substantive effect of the other input variables as well, reporting odds ratios and predicted probabilities. What are the tradeoffs of these two interpretation techniques?

Next, estimate the same model but *without including the interaction term* and use a likelihood ratio test to determine whether the more complex model offers a significantly better fit. Lastly, if you wanted to do some kind of residual analysis, how would you go about that in principle?

(b) Please analyze the data set *unrest.dta* using a count model. The data are available here:

`http://spia.uga.edu/faculty_pages/monogan/teaching/unrest_count.dta`

The outcome of interest is the variable *unrest*–a count of protest events in a given country, and the input variables (you must use them all) are:

- CL: Freedom House civil liberties index ($1-7$ scale, with higher values indicating lower levels of civil liberties).
- soviet: Dummy–whether a country is a former Soviet block country
- polity: an idex that ranges from -10 to 10 measuring level of democracy (higher values = more democratic)
- politysq: polity squared
- urbanpop: Percentage of a country's population that lives in an urban setting

Start by fitting a Poisson model and reporting these results. Please test for overdispersion in these data and describe what overdispersion is and why it is potentially a problem. What conclusions can you draw from these tests? What is the best choice of count model for these data and how did you make this choice?

For every set of results you report, present the results in a table (separate or combined, across models) including the coefficients, the standard errors, at least one fit statistic, and any additional information you would like. For the one model you determine to be best for these data, please tell us: What can you conclude from the z-ratios associated with each coefficient? For all models, what can you determine from the fit statistic?

Now test the hypothesis that the effect of Civil Liberties on unrest events is different in former Soviet countries than in the rest of the world. For the one model you determine to be the best for these data, Please illustrate the nature of this conditioned relationship using predicted counts with confidence intervals. For this one model, please assess the substantive effect of all the other input variables as well. When interpreting the effects of other predictors, you may choose among the methods of: partial changes in the conditional mean, factor change in the conditional mean, discrete change in the conditional mean (e.g., predicted counts), or predicted probabilities of counts.