Methodology Minor Field Exam

Fall 2016

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. You are free to use whatever word processing or typesetting software you like to write your answers. The questions must be answered in the allotted time.

Morning Session: Statistical Theory and Modeling Decisions

Answer one of the following two questions:

1. *Bayesian Statistics:* Western & Jackman (1994) argue that Bayesian inference is particularly appropriate for comparative research for two primary reasons: (1) the observed data constitute an entire population, and (2) the data are often weak due to collinearity and few observations. Whether studying comparative politics or not, how does encountering each of these two situations in any kind of research pose a problem for the applied researcher? How does Bayesian inference provide a solution to each problem?

   Drawing from any substantive area you wish, please provide a working example of a real or hypothetical dataset that would face these two problems. (Alternatively, you can provide separate examples of each problem and answer the following questions for each.) Explain what your outcome variable and predictors are, as well as the scope of the data. How is each of the two problems present in the data? As a Bayesian analyst, how would you specify your model of interest for these data? Be sure to explain how you would specify the structural form of the model, set the priors, estimate the model, and interpret the results. In what way does your model address the problems of modeling a full population of data and modeling weak data with few observations and/or collinearity? Why would a similar structural model estimated with least squares or maximum likelihood fail to address these issues as adequately?

2. *Panel Data:* When analyzing a limited dependent variable (such as a binary or count outcome) in panel data, there are two broad approaches: marginal models and generalized linear mixed effects models. What is the basic logic behind each approach? How is each estimated? What is the difference in interpretation between the two? How would you make the choice of which approach to use in a real study? (Feel free to use an example, if that helps illustrate, but that is not required.)

   Now imagine that you are studying air pollution policy in the United States. Suppose you have panel data for the 250 largest cities in the nation for the 20 years from 1991-2010. In every year, you have a binary measure for whether the city is a "nonattainment area": This variable is coded 1 if the city's air quality in a given year is worse than federally-prescribed national standards, and 0 if the air quality falls within national standards. Your predictors are how many Clean Air Act enforcement actions were taken by the federal EPA in that city in that year, how many actions were taken by state and local actors, the population density of the city, and the percentage of the city's jobs that are in the manufacturing sector.

   For this hypothetical study, how would you conduct this analysis? Who would be your target audience—federal policymakers, city officials, academic readers, or someone else? Would you estimate a marginal model or a generalized linear mixed effects model? Without seeing real data, how do you tentatively envision accounting for the error correlation structure in these data? Explain why you made these choices, and feel free to offer any other details you think would illustrate your choices.

Afternoon Session: Analyzing Data

Answer one of the following two questions:

3. *Poisson Regression:* Please analyze the data set *couart2.dta* using a poisson regression model. The data set contains information on the number of articles published by PhD students during the last 3 years of their education. The variables are as follows (you must use them all):

   art Outcome variable–number of articles published in last 3 years of PhD.

   fem Dummy for gender. 1 = female.

   mar Dummy for marital status. 1 = married.

   kid5 Number of children

   phd Prestige score of PhD granting institution (higher = more prestigious)

   ment Number of articles published by student's mentor in last 3 years

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

   Also, please present graphs of predicted counts against all covariates.

   Finally, discuss whether or not you think Poisson regression is the appropriate technique for these data and justify your answer. If no, then discuss other options and why they may be more appropriate.

4. *Ordered Logistic Regression:* Please analyze the survey data set *ordered.logit.dta* using an ordered logistic regression model.

   Conventional wisdom among scholars of interest groups in American Politics states that a primary goal of groups is to develop and maintain access to policy makers. While much of this work has focused on groups' ties to members of Congress, sometimes equally important is the extent to which groups cultivate connections within executive and regulatory agencies.

   Here, you will examine the casues of group access to federal agencies. The data are from a 1985 survey of interest groups and associations listed in the CQ's Washington Information Directory ($N = 892$). The groups were asked, "For the federal agency with which this association communicates, conslults, or interacts the most, does this association interact with it frequently, occasionally, seldom, or almost nevrer?"

   The outcome of interest is the variable **interact**—coded 1 for "almost never", 2 for "seldom", 3 for "occasionally", and 4 for " frequently". The explanatory variables are:

age —age of the group in years.

taxexmpt Dummy—whether the group is tax exempt or not.

indmembs Dummy–whether the group's members consist of individual persons.

orgmembs Dummy—coded 1 for associations where members are themselves associations.

Present the results of this model in a table including the coefficients, the standard errors, 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 model fit?

Please also provide predictied probabilty graphs displaying the effects of these explanatory variables on the probability of increasing the degree of interactions.