Methodology Minor Field Exam

Spring 2020

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.

For the morning session, internet usage is prohibited. For the afternoon session, you may use the internet to download software packages or look up reference information as you complete the data analysis. Your work must be fully your own. Enjoy this opportunity to showcase your skills.

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 recently 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 model results from Bayesian estimation related to MLE or OLS estimates? Compare the interpretation of estimates from a Bayesian to estimates from an OLS or MLE model. 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.

2. *Linear Model Theory:*

   When estimating a linear model via OLS, one assumes the Gauss-Markov assumptions to be true in order for the estimates to be BLUE. What are these Gauss-Markov assumptions? What does it mean for an estimator to be BLUE? What are the most common violations of Gauss-Markov assumptions and what types of data are most likely to lead to violations of these assumptions? What are the implications to violating Gauss-Markov, specifically in terms of interpreting model results? Finally, which assumption do you feel is the most important to not violate and why?

3. *Rational Choice Theory:*

   Nash equilibria have 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:

4. *Ordered Response Model:* Please analyze the dataset *torture.csv* using a regression model appropriate for a discrete, ordered response. The dataset contains information on 199 countries for the years 1982–2011. The dependent/response variable, **ciri_tort**, indicates whether evidence suggests state agents did not practice torture at all (category 1), practiced torture occasionally (category 2), or practiced torture frequently (category 3). The input variables (you must use them all) are:

   conflict_25 Dummy for civil conflict (armed conflict resulting in $\geq$ 25 battle-related deaths)

   leg_con Index of legislative constraints on the executive (0 to 1, higher = stronger constraints)

   log_gdp Gross domestic product in US dollars (log transformed)

   log_pop Population size in thousands (log transformed)

   Present the results of this model in a table or figure that shows the coefficient estimates, their standard errors/confidence intervals, and any additional information you would like. What can you conclude from the hypothesis tests for each coefficient?

   Present graphs of predicted probabilities for each category of the dependent variable against two of the covariates.

   Finally, assess and discuss whether or not an ordered response model is appropriate for these data. If the model is not appropriate, discuss other options and why they may be more appropriate.

5. *Linear Regression:* Please analyze the survey data set *engagement.dta* using a linear regression model. The data set contains information on overall civic engagement during the 2012 presidential election. The variables are as follows (you must use them all): The input variables are as follows (you must use them all):

   engscale Civic engagement scale (**dependent variable**; 0 = doesn't participate in any activity, 5 = participates in all activities).

   education Educational attainment(1=didn't complete high school, 6=post-graduate degree).

   income Household income (1 = less than $10,000, 18 = $250,000 or more).

   age Respondent's age.

   stghpid A measure of strength of partisanship (0 = Independent, 3 = strong party ID).

   ideodist Ideological distance to President Obama (0 = same position as President Obama, 6 = at the opposite extreme relative to President Obama).

   Present the results of this model in a table including the coefficients, the standard errors, the R^2, 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?

   Please test the conditional hypothesis that educational attainment fosters civic engagement, and that the positive influence of education is especially intense for individuals with strong partisan attachments. Estimate a new model to test this hypothesis and discuss the results. Illustrate the nature of this conditioned relationship by graphing predicted values and confidence intervals. Provide a detailed interpretation of the conditional relationship and whether or not you think it matters.

   Then compare the fit of the two models and discuss the implications of including the conditional relationship described above relative to not including this. Which model do you feel is a better fit to the data and why?

   Finally, discuss whether or not you think OLS is the appropriate estimator for these data. If so, justify your response. If not, what model do you think would be a better estimator and why?