Political Methodology Comprehensive Examination, May 2014
Department of Political Science, George Washington University

Instructions: Answer all questions in part I, one question in part II, and after the exam turn in an empirical paper demonstrating your ability to use statistical models OR schedule an oral exam.

Part I: Answer all 5 questions; read all questions before answering any of them.

1. Suppose you are interested in testing how X affects Y. Based on theory and the literature, there are a number of potential controls you can add to your regression. You don't have enough data to control for all of them at once, forcing you to choose some subset of controls. How might this choice affect statistical inference? How should this choice be made? Should it be based on theory or the data? What are some alternatives to making a single choice and instead using information from multiple models?

2. Inference involves using known facts and sample data to learn about unknown quantities of interest (typically, relationships between concepts), and, importantly, communicating (and quantifying) the degree of uncertainty we have about such inferences. While standard errors are the primary means by which we communicate this uncertainty, discuss multiple additional factors that influence the degree of uncertainty surrounding the inferences we make about unknown quantities of interest. Feel free to use examples to illustrate your arguments.

3. Suppose you have a theory about how X affects Y and wish to test it quantitatively. The problem is you have multiple plausible measures of X. For instance, if your theory is about the effect of political violence, you could potentially test multiple examples of violence: riots, guerrilla warfare, violent protests, etc. What are the advantages and disadvantages of testing all of these measures simultaneously in a regression? What are some of the alternatives to this and what are their advantages?

4. On occasion, people will wander into the methods faculty's offices with problems like the following. "I have 20 years of cross sectional time series data. Some of the data are missing. The dependent variable is ordinal, and I am worried about the endogeneity of my key independent variable of interest. What model should I estimate?" The usual advice provided to such queries is roughly "There is not a single model that will solve *all* your problems, so decide which features are most important to your analysis and argument, and work on solving those problems." Explicate how you would proceed if faced with a similar situation. That is, how would you decide which problems to solve? What are the advantages and disadvantages of the approach you take?

5. You have some survey data with 600 observations. Four variables are of interest. The variable "support" is coded 1 if the respondent supports a proposal, 0 if the respondent opposes the proposal. You also know the respondents' party, divided into three categories, Democratic, Republican, and Independent. Of the 260 Democrats, 80% support the proposal. Of the 240 Republicans, 15% support the proposal. Of the 100 Independents, 30% support the proposal.

(5a) If you cross-tabulated party and support, would you reject the null of independence at the .05 level?

(5b) Assume you recoded the party variable into three dummy variables, Democrat (0/1), Republican (0/1), and independent (0/1). What would the values of $\hat{\alpha}, \hat{\beta}_1, \& \hat{\beta}_2$ be if you then used the 600 observations to estimate the following equation? $support_i = \alpha + \beta_1 Democrat + \beta_2 Republican + \epsilon_i$

(5c) Repeat (3b), this time for $support_i = \alpha + \beta_1 Independent + \beta_2 Republican + \epsilon_i$

Part II: Answer ONE of these two questions

6a. A major methodological trend in political science over the past ten years has been the increasing use of experiments (field, lab, and survey). Why is this the case? What do the advocates of experiments argue to justify their use? Critics of experimental methods in political science argue that the use of experiments restricts the domain of topics that can be studied by political scientists. Is this a fair critique? Why or why not?

6b. You have been asked to give a 45-minute presentation on the topic of "clustered data analysis" (either TSCS/panel data analysis or multilevel modeling). The goal of the presentation is to spread awareness of the methods and models in this area and to clear up misconceptions analysts may have about this type of data and methods for analyzing it. You are asked to choose what you think are the most important issues to discuss. Provide an outline and discussion of the key topics on which you would present.