Political Methodology Comprehensive Examination, January 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 testing a variable X and find that it has no correlation with your dependent variable Y. However, you discover that larger values of X are associated with larger magnitudes of the error term. Can you think of an empirical example in which this holds? Can failing to control for X in a regression lead to biased coefficients for any other independent variables? If not, is there any reason you may still want to incorporate X into your analysis somehow?

2. In the linear regression model (or any other model, for that matter), we often talk about regression assumptions and appropriate model specification. What's the difference? Discuss why and how each facet matters for making inferences about political phenomena.

3. How do we know whether an effect we have uncovered is a "causal effect?" What are the best ways to ensure proper estimation of causal effects?

4. Instrumental variables have long been popular in economics, but have only recently become widely used in political science. Why are political scientists increasingly using instrumental variables? What problems can they solve? And what is the role of theory in the application of instrumental variables?

5. Assume you estimate the model: $y_i = \alpha + \beta x_i + \epsilon_i$ and the model has an R^2 of .64. After running the regression, you obtain the fitted values ($\hat{y}_i$) and the residuals ($\hat{\epsilon}_i$).

(5a) What is the correlation between y_i and $\hat{y}_i$?

(5b) What is the correlation between y_i and $\hat{\epsilon}_i$?

(5c) What is the correlation between $\hat{y}_i$ and $\hat{\epsilon}_i$?

(5d) Do any of the correlations in (a)-(c) provide evidence regarding the causal relationship between x_i and y_i? Why or why not?

Part II: Answer ONE of these two questions

6a. For roughly the past 15 years, various scholars have advocated the use of Bayesian statistics for the analysis of political phenomena. In practice, however, Bayesian techniques often produce results very similar to non-Bayesian techniques. Assuming that to be true, what are the advantages, if any, of using Bayesian statistics in political science?

6b. For various types of clustered data (hierarchical or panel/TSCS), we are incorporating information from two (or more) levels of analysis. First, lay out a rationale for why we would ever want to do that in the first place. Second, what are the obstacles associated with such a strategy? Third, what opportunities arise from this practice?