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

Instructions: Answer 3 of 4 questions in part I, 3 of 5 questions 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 3 of 4 questions; read all questions before answering any of them.

1. Suppose a researcher makes the following statement: "Our goal is to explain as much of political phenomena as we can. Therefore, when choosing a model specification, you should add in whatever explanatory variables maximize the R^2." What, if anything, is wrong with this claim? What are some of the better approaches to choosing a model specification?

2. 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?

3. Political scientists often confront the following situation. The outcome of interest to be explained by a statistical model is ordinal, with a small number of values (3, 4, or 5, say). A standard model to estimate in such situations is either the ordered probit or ordered logit model. Why do analysts argue that such models are preferable to OLS? Are there any situations where OLS would be preferable? What is the proper way to interpret the coefficient estimates in an ordered probit or ordered logit model?

4. Say you have 15 data pairs $\{x_i, y_i\}$ and have computed the OLS line of best fit: $\hat{y}_i = \hat{\alpha} + \hat{\beta} \cdot x_i$. What happens to the correlation r_{x_i,y_i} when you add a 16th data point at the point of averages $\{\bar{x}, \bar{y}\}$? In contrast, what happens to the correlation r_{x_i,y_i} when you add a 16th data point at the point $\{\bar{x} + s_x, \hat{\alpha} + \hat{\beta} \cdot (\bar{x} + s_x)\}$? What aspect of regression and linear correlation does this illustrate?

Part II: Answer three of these five questions. Question 5 will take longer and is worth 2 questions. That is, if you choose to answer question 5, only answer one other question in Part II.

5. Consider a three-player game with an Autocrat, an elite Rival, and a representative Citizen. They have a pot of size R to split amongst themselves. The order of moves is as follows:

1. The Autocrat offers a split of R between the three actors.
2. The Rival can choose to coup, with probability of success p. If unsuccessful, the Rival gets nothing. If successful, the Autocrat gets nothing and the Rival offers a split between himself and the Citizen.
3. The Citizen can choose to revolt with probability of success q. If unsuccessful, the Citizen gets nothing. If successful, the Citizen wins R.

Of course, if neither a coup nor revolt is attempted, the original offer is implemented. Assume that if an actor is indifferent between two actions, they choose the option leading to peace. Find the unique subgame-perfect Nash Equilibrium to this game. What does the autocrat keep in equilibrium?

6. The Empirical Implications of Theoretical Models (EITM) movement in political science has been around for about a decade now, but different advocates of the approach propose alternative empirical strategies. One such debate appeared in *Political Analysis* in 2007. On one side, Carruba, Yuen, and Zorn argued on behalf of deriving comparative statics from formal models, then testing them with simple models. In contrast, Signorino argued on behalf of deriving a stochastic formal model, which is then mapped into a non-linear model and estimated using maximum likelihood or an alternative estimation technique. What are the costs and benefits of each approach? Are there types of situations where one approach is more likely to be successful and other types of situations where the other approach is more likely to be successful?

7. In 2002, the prominent political methodologist Chris Achen wrote the following:

> Even at the most quantitative end of the profession, much contemporary empirical work has little long-term scientific value. "Theoretical models" are too often long lists of independent variables from social psychology, sociology, or just casual empiricism, tossed helter-skelter into canned linear regression packages. Among better empiricists, these "garbage-can regressions" have become a little less common, but they have too frequently been replaced by garbage-can maximum-likelihood estimates (MLEs). Beginning graduate students sometimes say, "Well, I don't really understand how these variables relate to each other and the data are bad, but I did use the newest estimator, downloaded from the Internet, and I do report heteroskedasticity-consistent standard errors."

Twelve years later, does Achen's claim still hold up? Do you see any countervailing trends or developments that would give us reason for optimism?

8. Regression discontinuity analysis and instrumental variables have become more widely used in political science over the last decade. Why is that the case? Some critics of these techniques argue that they encourage atheoretical applications, with an emphasis on cleverness in finding good instruments or discontinuities, then telling a story to wrap the analysis. Does this critique have merit? Why or why not?

9. Standard practice in political science in recent years has been to account for heteroskedasticity by the use of so-called "robust" standard errors. Are robust (Huber/White/sandwich) standard errors appropriate for all situations? If not, under what conditions would alternative remedies be preferable?