

Political Methodology Comprehensive Examination, May 2011
Department of Political Science, George Washington University

Instructions: You have five hours to complete the exam. Good luck!

Part I

1. Former Illinois Governor Rod Blagojevich currently is facing a corruption trial for charges associated with his alleged attempt to sell Barack Obama's Senate seat. His attorney, who is considering a plea agreement, contacts you to calculate the probability that a jury will find him liable. You know that the prosecutor needs a unanimous vote of the jurors to win a criminal conviction. You also know that this is made more difficult if there is a person who is sympathetic to the defendant, like a fellow Democratic partisan or a contributor to the Obama campaign. Indeed, you know that 35% of voters are Democrats in Illinois and Obama received campaign contributions from about 2.3% of voters in 2008. Given these pieces of information, what is likelihood that a jury will find the Blagojevich guilty of the corruption charges given the chance that the jurors are a) Democrats or b) campaign contributors? [Note: Make two separate calculations here. Ignore the issue that most campaign contributors are likely to be Democrats.]

2. Members of the House of Representatives, many running against government spending in 2010, enacted earmarks for specific projects on average worth \$28,786,844 with a standard deviation of \$24,730,417. Assuming normality:
 - a. What is the probability that a member would sponsor \$500,000 in earmarked spending?
 - b. What is the probability that a member would sponsor \$75,000,000 in earmarks?
 - c. What is the probability that a member would sponsor between \$35,000,000 and \$60,000,000 in earmarks?

Do you have concerns about making the normality assumption for these data? If so, why?

3. You conduct a study of voting behavior in the 2008 election: why did people support Barack Obama? One hypothesis is that support for Obama was driven by negative assessment of the Bush administration (i.e., retrospective voting). Your model includes variables for ideology and views of how things went over the past 8 years of Republican administrations. A reviewer of your paper suggests that your model may evidence heteroskedasticity, suggesting that you include a variable capturing the respondents' level of information (i.e., how attentive they are to the news). Without examining the data, why might it be plausible to think that these data possess heteroskedasticity? How should you test for violations of this assumption? What solutions should you employ if you find evidence of heteroskedasticity? Is it appropriate to include an information variable to resolve any possible issue – why or why not?

4. Matching is an increasingly popular technique in political science. What problem is matching intended to solve? Under what conditions will matching work well? Under what conditions will matching break down or not work well? What alternative solutions are plausible when matching won't work?

5. Estimating nonlinear models like logit and probit and their ordered, multinomial, and heteroskedastic generalizations entails some costs in the hope of reaping additional benefits. What are the costs and benefits of estimating nonlinear models? How should we decide what models to estimate in specific applications?

Part II (Answer one of the two following questions)

6a. At a recent (fictional) talk, a political scientist gives a presentation where he is employing multilevel data. In the model specification and estimation portion of his presentation, he merely states that he "estimated a multilevel model, which offers an advantage over alternative models in the analysis of multilevel and hierarchical data." He then presents his model results, yet the audience is not sure what particular model he actually estimated. Offer an in-depth critique of this political scientist's presentation of his model specification and estimation. For example, what information is he omitting that the audience would surely want to know? Offer some advice on how he should characterize what he is doing that offers a sufficient explanation of his model specification and estimation procedures.

6b. From the work of Beck and Katz, one could argue that a consensus has emerged on how analysts should model time-series cross-sectional (TSCS) data: specify a fixed effects model with a lagged dependent variable and panel-corrected standard errors. What are the pros and cons of this approach? Compare and contrast it with alternative approaches for modeling TSCS data.

Part III

Either submit an empirical research paper along with the exam or schedule an oral exam after the written exam.