

Political Methodology Comprehensive Examination, May 2015
Department of Political Science, George Washington University

Instructions: *Read all questions before answering any of them. Answer 6 questions in total, with question 9, which will take longer, counting as two questions. Put differently, if you answer question 9, you only need to answer 5 questions in total. After the exam turn in an empirical paper demonstrating your ability to use statistical models OR schedule an oral exam. Good luck!*

1. Field experiments are increasingly used to improve causal inference in political science. What potential threats do field experiments confront that would compromise causal inferences?
2. Network analysis includes, among other things, a set of measurement tools. Give examples of two of these tools. Explain the theoretical basis for these tools, what advantages they have over more conventional measurement tools, and their limitations. For each tool, give an example of how one might use it in applied research, including an explanation of why it would be useful in this context and how it can improve our inferences.
3. In the statistical analysis of observational data, one might say that the key to designing a great empirical test is to make as tight of a connection between theory and method as possible. The goal should be to specify a statistical model that can generate evidence directly connected to one's underlying theory and hypotheses related to how X influences Y. Use this "theory-method congruence" principle to evaluate various methods commonly used in the analysis of "clustered data" (e.g., time-series cross-sectional data and multilevel data more generally).
4. Explain the concept of a "placebo test" as used in observational causal inference. In what ways are these tests useful for making causal claims? What is the relationship between these placebo tests and the traditional use of placebos (i.e., sugar pills) in experiments? How can certain types of placebo tests be used to supplement other observational causal inference techniques?
5. Assume you have time series (not cross-sectional time series) data at the annual level that you have used to model some aspect of the US political economy. A critic raises a concern that your time series regression suffers from a spurious regression problem. What is the spurious regression problem, and what evidence can you provide to alleviate your critic's concerns?
6. A major challenge when estimating non-linear models such as ordered probit, multinomial logit, weibull models, etc. is deciding what marginal effects to present to your audience. What principles should you apply when presenting and interpreting marginal effects of non-linear models? And are different types of marginal effects more appropriate given different research questions?
7. Earlier this week, the political science community buzzed with the news that Don Green of Columbia was publicly retracting an article that was published in Science, work that had received media coverage in the Washington Post, New York Times, This American Life, etc.. Three political scientists presented Green with evidence that the data collected by his co-author on the paper was too good to be true, and Green was convinced by the evidence. Throughout your career you will be required to evaluate empirical results, whether in your own research or others' research you are reviewing. Offer three specific ideas about how to explore data to look for evidence of data mistakes or fraud.
8. You have just presented a multiple regression model at a conference. The purpose of the regression model was to estimate a particular treatment of interest. An audience member objects to your results, claiming that your estimate is biased because you left out a particular variable of interest to her. Is she right? How could you respond in that situation, assuming you do not have the data on hand to estimate her preferred model? Put differently, how can you defend yourself?

9. Consider the OLS regression output below. The observations are white Americans in the 2004 National Election Study survey. The dependent variable is the feeling thermometer for Colin Powell, which ranges from 0 to 100. [At the time of the survey, Colin Powell was the United States Secretary of State.] All the independent variables have been scaled 0-1, so for continuous variables like Age, the minimum age is 0, the maximum age is 1, and the rest of the age are scaled accordingly. *Note: when answering the questions here, keep your responses as concise as possible.*

Source	SS	df	MS			
Model	5.47708	9	.608564445	Number of obs =	618	
Residual	21.8020069	608	.035858564	F(9, 608) =	16.97	
Total	27.2790869	617	.044212458	Prob > F =	0.0000	
				R-squared =	0.2008	
				Adj R-squared =	0.1889	
				Root MSE =	.18936	

Powell thermometer	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
Racism	.096209	.0370106	2.60	0.010	.023525	.1688931
Republican	.0839401	.0328671	2.55	0.011	.0193933	.1484868
Democrat	-.0442625	.0334775	-1.32	0.187	-.1100081	.0214831
Political Awareness	.0901019	.0339432	2.65	0.008	.0234418	.156762
Age	.0458706	.0255014	1.80	0.073	-.004211	.0959521
Education	-.0091021	.0332541	-0.27	0.784	-.074409	.0562048
Female	.007228	.0156651	0.46	0.645	-.0235363	.0379922
Southerner	.0150523	.0168659	0.89	0.372	-.0180702	.0481748
Ideology	.0119771	.007057	1.70	0.090	-.0018819	.0258361
_constant	.505332	.0527084	9.59	0.000	.4018193	.6088446

- Now suppose that not all of the assumptions of the classic linear regression model (CLRM) hold. In particular, the data are characterized by heteroskedasticity, and it is a function of the Xs. Nevertheless, you estimated your model with OLS. What are the implications of this for your estimated coefficients, the standard errors, and your fit statistics?
- Suppose you have reason to believe (a theory!) that the heteroskedasticity you worried about in part (a) was a function of education (which is measured in seven ascending categories). Explain how you would conduct and interpret a formal statistical test of this theory. What would you do if you found support for your theory?
- Using this regression model, how would you test the hypothesis that “party doesn’t matter to Americans’ evaluations of Colin Powell”? If you can test the hypothesis from this output alone, do so (set-up/report/interpret). If you cannot, explain why not and what else you’d need to know. [Go back to assuming CLRM holds for this question.]
- All else equal, what is the expected difference in Powell ratings between women from the South and men from the North?
- You present the results from this regression at a conference and an audience member asks whether you “considered the argument that political information or awareness should have caused a divide between partisans? That Democrats and Republicans really would have reacted differently to more information about Powell’s role in the war effort?” Did this regression do that? If so, report and discuss the relevant results. If

not, write down an amended model that would consider the argument the audience member raised. Discuss what information from that model (including any necessary tests not immediately reported in Stata regression output) you would use to answer the audience member's question.

10. Consider a model of democratic politics with three actors: an incumbent, a rival, and a representative citizen. In the status quo, the incumbent holds a fraction $\alpha > 1/2$ of power and the rival $1 - \alpha$. The order of the game is as follows:

1. The incumbent decides to **break** the law or not. Only the rival sees whether this occurred.
2. The rival decides to **announce** whether the law was broken or not, at cost f . If the rival does not announce, the game ends. If the rival announces, the citizen gets a separate message (through the media) as to whether the law was actually broken. The message is correct with probability $q > 1/2$.
3. The citizen decides to **protest** or not, at cost c . He or she gets a payoff d for choosing correctly (i.e., for protesting if the law was broken or for not protesting if the law was not broken).

The outcomes for the politicians are as follows:

- If a protest occurs, the rival gets W and the incumbent gets 0.
- If the law is broken and no protest occurs (including if the rival doesn't announce), the incumbent is able to consolidate power. The incumbent gets W and the rival gets 0.
- If the law is unbroken and no protest occurs, the status quo remains. The incumbent gets αW and the rival gets $(1 - \alpha)W$.

We're going to solve for when a **democratic equilibrium** exists in which the incumbent does not break the law, the rival announces if and only if the law is broken, and the citizen protests if and only if they get a message indicating the law was broken. We proceed using backward induction.

- (a) Find the condition under which the citizen protests if and only if they get a law-breaking message. Note: A tricky part of this game is that the citizen could update his or her belief based on the rival's action. However, in a democratic equilibrium, the citizen only gets a message off the equilibrium path, so assume that if they get a law-breaking message, they believe the law was actually broken with probability q (and correspondingly for the opposite message).
- (b) Assuming (a), find the condition under which the rival announces if and only if the law was broken.
- (c) Assuming (a) and (b), when will the incumbent choose to break the law?
- (d) What do the set of conditions tell you about the role of informational quality in stabilizing democracy?