Political Methodology Minor Exam – Fall 2011

If you believe a question provides insufficient information to answer, explain why you think so, state some minimal additional assumptions necessary to answer, make those assumptions, and answer accordingly. There should be enough information, but this might earn partial credit. For questions that require calculation, please show your work.

The time for the exam is 8 hours. 8:00 a.m. to 5 p.m. on September 15th.

PART I.   Answer ALL of the following questions.

1. Consider the OLS regression output below. (The observations are white Americans in the 2004 National Election Study survey.)

```
      Source |       SS       df       MS                Number of obs =     618
-------------+------------------------------         #5 F(  9,   608) =   16.97
       Model |   5.47708        9  .608564445            Prob > F      =  0.0000
 #8  Residual |  21.8020069    608  .035858564           R-squared     =  0.2008
-------------+------------------------------         #6 Adj R-squared =  0.1889
       Total |  27.2790869    617  .044212458         #7 Root MSE      =  .18936
```

Powell thermometer	#1 Coef.	#2 Std. Err.	t	#3 P>|t|	#4 [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

a. For each emboldened item in the output (1-8), give a mathematical formula for how it is computed and briefly explain its meaning/interpretation, being sure to note what, if any, population parameter it is meant to estimate. (Note that the items are numbered—use that numbering for the order of your answers!)

b. Now suppose that not all of the assumptions of the CLRM hold. In particular, the data are characterized by pure heteroskedasticity, and it is a function of the Xs. Nevertheless, you estimated your model with OLS. For each of the eight emboldened items, explain the implications. Be sure to note whether or not you would expect a different value (as compared to the value you'd expect if the CLRM held) and if so, where that change in value would come from. Also be sure to mention statistical properties (bias, consistency, etc.), where applicable, when estimating via OLS under these data conditions.

c. Suppose you have reason to believe (a theory!) that the heteroskedasticity you worried about in part b was a function of education (which is measured in seven ascending categories). Explain how you would conduct and interpret both a Glejser's test and a Goldfeldt-Quandt test of this theory. What would you do if you found support for your theory?

d. 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.]

e. All else equal, what is the expected mean difference in Powell ratings between women from the South and men from the North?

f. You present the results from this regression at a conference and an audience member wants to dismiss them "because you forgot Iraq." You reply to the critic that you're not concerned about Iraq, because at the point at which the survey was conducted, support for the war in Iraq was high among all sorts of Americans. Explain the problem the audience member was claiming you had, and how your response was addressing it.

g. Another critic at that conference says they are concerned about the validity of your results because they have concerns about error in the Powell feeling thermometer measure. In particular, they worry that, in general, survey respondents aren't exactly sure where to place themselves on the 0-100 scale—that they have difficulty being so precise. And so the respondents are likely just picking some number out of a larger range of numbers that they would find as acceptable representations of their true opinions. Tell the critic whether or not you think their concerns are problematic for your estimates (why or why not).

2. You go to a job talk where the candidate is making the argument that individuals' issue positions and gender attitudes have differently structured their preferences for the Republican party over time because the time periods have offered contexts with greater or fewer women candidates on the ballot, which conditions the effects of attitudes and positions on partisan preferences. Regardless of context, however, the candidate argues, Republicans are more preferred if they are the incumbent party.

a. Use the following variables to write out a single regression equation that might capture the argument being made in the talk, adding any subscripts that might help clarify the equation:
   Republican Preference
   Gender Attitudes
   Issue Positions
   %Female Candidates
   Incumbent Party

b. Rather than run a single regression, the job candidate says that he estimated a multilevel model. The system of equations that specify his model are:

$$\text{Republican Preference}_{it} = \beta_{0t} + \beta_{1t}\text{Issue Positions}_{it} + \beta_{2t}\text{Gender Attitides}_{it} + \varepsilon_{it} \tag{1}$$

$$\beta_{0t} = \gamma_{00} + \gamma_{01}\text{Incumbent Party}_t + u_{0t} \tag{2}$$

$$\beta_{1t} = \gamma_{10} + \gamma_{11}\%\text{Candidates Female}_t + u_{1t} \tag{3}$$

$$\beta_{2t} = \gamma_{20} + \gamma_{21}\%\text{Candidates Female}_t + u_{2t} \tag{4}$$

Where ε_{it}, u_{0t}, u_{1t}, and u_{2t} are all stochastic.

Substitute the values assigned to β_{0t}, β_{1t}, and β_{2t} by equations 2 through 4 into equation 1, multiply out, and collect the stochastic terms together. What problems would you have if you ran OLS to estimate this equation? (i.e., explain what assumptions would be violated and what the consequences for the OLS estimator would be.)

c. Comment on the similarities and differences between the equation you wrote in part b and the model you wrote out in part a. What does this tell you about why the job candidate made the modeling choice(s) he did?

3. Is it possible for the OLS residuals generated by some model to "pass" every statistical test we discussed for "white noise," and still contain substantively important information we could use in refining and improving our model? In other words, is it possible for residuals to look like white noise on all statistical grounds, and yet not be white noise? If so, describe how this could be true.

1. Causal Mechanisms and Lawlike Generalizations.
   Within positivist social science a debate has emerged between scientific realists who think the surest road to useful knowledge about politics is through the identification of causal mechanisms, which may or may not travel to other cases; and empiricists who think it is through the identification of lawlike generalizations, which are necessarily muh broader in scope. Perhaps ideally we would want both kinds of information, but since different methods and logics of inquiry follow from the two approaches individual social scientists have to choose between them in their research projects.

   Compare and contrast the arguments made by each side in this debate about why their approach should be favored. On balance, what is your view and why?

2. Suppose that you are analyzing voting data with measures at both the individual and county level. Write the separate level-1 and level-2 equations as well as the combined multi-level model. What assumptions are necessary to estimate this model? How would you interpret a random intercept and random coefficient model? Besides a multi-level model, what other models could you estimate? Why is a multi-level model superior?

3. What are fixed effects? Provide a formal derivation of the OLS estimators with fixed effects. What are random effects? Which should you use? What problem in TSCS/panel data are fixed and random effects designed to solve? What do fixed and random effects not solve for TSCS/panel data?

4. Experimental research is lauded for its superior "internal validity." What does this phrase mean, and how does internal validity differ from external validity? How is internal validity achieved? What is a true experiment anyway, and how does it differ from other varieties of experimental research (quasi-experiments, simulations)? What are some of the major threats to internal validity, and how can they be minimized through careful research design? In your answer, compare experimental designs to "pseudoexperiments" and cite specific examples from the experimental literature whenever possible.

5. What are the advantages and disadvantages of discrete versus continuous time event history models? What are advantages and disadvantages of parametric versus Cox models? What major assumption(s) should be tested in a Cox model? How do you conduct such testing and what are the appropriate corrections? What complications are added by multiple events? How can they be addressed?

6. There are several different statistical estimation approaches available, including least-squares, maximum-likelihood, and Bayesian. Take at least two approaches, and compare their assumptions and relative advantages and disadvantages. Under what circumstances would the approaches you choose give identical results? Which one is considered more general and why?

7. Find the solution of the difference equation for the given initial condition. Graph the solution sequence with time on the horizontal axis. Characterize the limiting behavior of the sequence.

$Y_{t+1} - 3Y_t = .10$ $Y_0 = 1$

$4Y_{t+1} - Y_t = 2$ $Y_0 = 4$

$Y_{t+1} + Y_t = 0$ $Y_0 = -10$

$3Y_{t+1} + Y_t = 1$ $Y_0 = 1$

$7Y_{t+1} + 7Y_t = 3$ $Y_0 = 5$

Why are nonlinear difference equation important in the study of time series analysis?

8. What does it mean for an estimator to be consistent? If you can give a formal definition, be sure to also give an explanation in words that could be understood by a colleague who has not taken a class in statistics. What must be proved for an estimator to be consistent?

9. What is the Independence of Irrelevant Alternatives assumption? Use conditional probabilities to sketch how it might be violated. What does it imply for models of unordered choices? What models might one use to model data when IIA is violated?

10. Under heteroskedasticity, the least squares regression model is inefficient. An alternative is the use of generalized least squares (GLS). Derive the GLS estimator. What is the chief drawback of GLS?