Political Methodology, Fall 2007
Minor Exam Questions

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.

PART I – ANSWER 3 OF THE FOLLOWING 4 QUESTIONS

1. *Regression Theory*: State the Gauss-Markov theorem. Using this theorem, prove that the OLS estimator is BLUE. What is the significance of the Cramér-Rao Lower Bound?

2. *Applied Regression I*: Consider the OLS regression printout below, estimated on data for which the CNLRM holds:

   The dependent variable below is *Inflation* and the sample size is 82 observations.

Variable	**1** *Coefficient*	**2** *Std. Error*	t-statistic	**3** *Prob. (2-Sided)*
Constant	13.18958	1.687733	7.814967	0.0000
Central Bank Indep. (CBI)	-12.15446	3.555119	-3.418863	0.0010
Coord Wage Barg (CWB)	-6.853600	2.874575	-2.384213	0.0195
CBI*CWB	9.740278	5.640083	1.726974	0.0881

4 R^2	*0.227081*	Mean Dependent Variable	6.575396
Adjusted R^2	0.197353	S.D. Dependent Variable	4.064031
5 *S.E. of Regression*	*3.640988*	Akaike Info Criterion	2.632061
Sum Squared Residuals	1034.030	Schwartz criterion	2.749462
Log likelihood	-220.2674	**6** *F-statistic*	*7.638717*
7 *Durbin-Watson*	*1.952286*	Prob(F-statistic)	0.000153

   (a) For each italicized item above, explain how it is computed (give a mathematical formula) and give a brief description of its interpretation.

   (b) Now suppose that the CNLRM does not apply; rather the data exhibit pure heteroskedasticity that is not a function of the x's, of the x^2's, or of the cross-products of the x's. Nonetheless you employed OLS estimation. Tell how each of the labeled items is affected. In particular, be sure to mention whether their values change and how, if applicable and possible; be sure to mention whether and what changes occur in their statistical properties as well (e.g., are they biased, inconsistent, etc. under this kind of heteroskedasticity?), again as applicable and possible.

   (c) Again suppose that the CNLRM does not apply; again the data exhibit pure heteroskedasticity but this time it is a function of the x's, of the x^2's or of the cross-products of the x's. Nonetheless you used OLS estimation. Tell how each of the labeled items is affected, again being sure to mention whether their values change and how if applicable and possible, and whether and what changes occur in their statistical properties as applicable and possible.

3. *Applied Regression II*:You estimate the following regression:

$$\mathbf{y} = \mathbf{b}_0 + \mathbf{b}_1 \mathbf{x} + \mathbf{b}_2 \mathbf{z} + \mathbf{b}_3 \mathbf{x} \cdot \mathbf{z} + \mathbf{e} \tag{1}$$

STATA estimates that $\mathbf{b}_0 = 3.5$, $\mathbf{b}_1 = 2$, $\mathbf{b}_2 = 1$, $\mathbf{b}_3 = -3$. Assuming that $\mathbf{x}$ and $\mathbf{z}$ both vary from 0-1, graph the following in a spreadsheet (label appropriately and print):

(a) $\delta y/\delta x$ on the y-axis against z on the x-axis

(b) $\delta y/\delta z$ on the y-axis against x on the x-axis

(c) Interpret the two lines you've just drawn

(d) Generically, what is the variance of $\delta y/\delta x$?

(e) Generically, what is the variance of $\delta y/\delta z$?

(f) STATA also tells you that $\mathbf{V}[(\mathbf{b}_1, \mathbf{b}_2, \mathbf{b}_3)']$ is:

$$\mathbf{V} \begin{pmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{pmatrix} = \begin{pmatrix} +0.256 & -0.027 & -0.309 \\ -0.027 & +1.296 & -1.624 \\ -0.309 & -1.624 & +2.652 \end{pmatrix} \tag{2}$$

(g) Graph $\delta y/\delta x \pm 1$ std. dev.

(h) Graph $\delta y/\delta z \pm 1$ std. dev.

(i) Why are $\delta y/\delta x \pm 1$ std. dev. and $\delta y/\delta x \pm 1$ std. dev. shaped like they are?

4. *Logistic Regression* You have data from 250 murder trials in which the dependent variable ($\mathbf{D}$) is the jury decision (1 if acquit, 0 if convict). Your independent variables are minority status of the accused ($\mathbf{MA} = 1$ if minority, 0 if not), the proportion of the jury that are minorities ($\mathbf{MJ} \in [0\ldots1]$), and the true facts of the case in the form of Θ, the *true* probability the accused is actually guilty given the available evidence. (You just happen to know the truth for this problem; so much for realism.) You decide to estimate $\Pr(\mathbf{D} = 1) = \mathbf{f}(\mathbf{MA}, \mathbf{MJ}, \Theta)$ using logit; specifically you estimate by maximum likelihood:

$$\Pr(\mathbf{D} = 1) = \frac{e^{\beta_0 + \beta_1 MA + \beta_2 MJ + \beta_3 \Theta}}{1 + e^{\beta_0 + \beta_1 MA + \beta_2 MJ + \beta_3 \Theta}}$$

(a) Generically (*i.e.*, formulaically), what is the marginal effect in this model of the accused's being a minority on the probability of her conviction?

(b) Suppose you estimate that $\beta_0 = 0$, $\beta_1 = -0.5$, $\beta_2 = +0.25$, and $\beta_3 = -1.5$. Holding $\mathbf{MJ}$ and Θ both fixed at 0.5, what is the estimated effect of the accused's being a minority as opposed to non-minority on the probability of conviction?

(c) If there were no prejudice (racial bias) in the system, what would the coefficients in this model be? How would you test that hypothesis?

(d) If you estimated that β_1 was negative and β_2 was positive, what would this mean substantively? What if β_1 was negative and β_2 was zero?

PART II – ANSWER 1 OF THE FOLLOWING 2 QUESTIONS

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 much 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. Qualitative and Interpretive Methods

   In recent years there has been considerable interest among scholars working within a positivist frame of reference in advancing qualitative methods as a tool for the study of world politics. By virtue of its emphasis on case studies and its focus on institutional and historical context the qualitative methods "movement" might seem - depending on your point of view - to bridge, co-opt, or marginalize the preferred methods of scholars who identify themselves as interpretivists.

   Compare and contrast qualitative and interpretive methods. To what extent are the goals, techniques, and logics of inquiry of the two traditions similar, and how are they different? Are qualitative IR scholars interpretivists in all but name, or is positivism now so broad that there now no need for interpretivism? And what, if anything, might each side learn from the other?

PART III – ANSWER 4 OF THE FOLLOWING QUESTIONS

1. Calculate the total long run effect for the following error correction model:

$$\Delta y = -.45 y_{t-1} + .35 \Delta x_t + .25 x_{t-1} \tag{3}$$

   What component of that effect would be considered short-term? Describe the logic behind the error correction model.

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. Explain the logic behind Heckman selection models. Provide a formal outline of why one must correct for selection. What is the exclusion restriction and how would one make sure your model satisfies the exclusion restriction? Why can't you trust the naive estimates of the standard errors from a two-step Heckman selection model? Finally discuss the advantages and disadvantage of the Heckman selection model.

4. 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?

5. Explain Krosnick's application of the satisficing notion to the evaluation of survey questions, and also explain Tourangeau's four-component theory of survey response. Would you consider these two approaches to be opposing one another or complementing one another, and why? How can these two perspectives be used to help guide question writing and questionnaire construction?

6. Internet surveys have become common. What advantages, if any, do Internet surveys have over other survey modes? Summarize any research you know about how similar the results of Internet surveys are to surveys taken through other modes. Discuss the limitations of Internet surveys from the Total Survey Error perspective. (In answering this question, be sure to indicate what type of Internet surveys you are critiquing.)

7. Explain the considerations involved in choosing a prior for Bayesian analysis. How much difference does the choice of a prior usually make? In what circumstances are diffuse priors used rather than informed priors? If possible, give an example of a situation in which an informed prior gives a different result than a diffuse prior would. How would Bayesians answer critics who argue that using an informed prior is an instance of "cooking" the results to obtain the results the researcher wants?

8. 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 design? In your answer, compare experimental designs to "pseudoexperiments" and cite specific examples from the experimental literature whenever possible.

   item Most of the empirical evidence on public opinion was obtained through survey research. Recently, many psychologists and survey researchers have begun to examine in-depth the psychological processes involved in the answering of survey questions. What have we learned about these processes? How might we adjust our research techniques to take advantage of the latest thinking about the psychology of the survey response? Can these studies tell us anything useful about political thinking outside the survey setting?

9. What are the advantages and disadvantages of discrete versus continuous time event history models? What are the 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 appropriate corrections? What complications are added by multiple events? How can they be addressed?

10. Consider the following simultaneous move game as represented in its normal form.

(see figure, attached next page)

a. Draw an extensive form representation of this game. b. What are the Nash equilibria if x = 0? (Consider both pure and mixed strategies.) c. What are the pure-strategy Nash equilibria if x = 2? d. What are the Nash equilibria if x = -2? (Consider both pure and mixed strategies.) e. Characterize the pure-strategy ash equilibria for all possible values of x between negative infinity and infinity.