Final Exam: Problems (2 hours)

Each problem is weighted equally. In order to get full credit, you must show any calculations used to arrive at your answers and completely answer the questions.

1. Answer true, false, or uncertain for each statement below. Explain your choice.

    a) OLS is a special case of 2SLS.

    b) It is always better to use IV than to use OLS when you suspect endogeneity.

    c) A good way to deal with multicollinearity is to use principal components of variables instead of the variables themselves.

    d) Use of top-coded data for the y-variable without adjusting for the top-coding leads the OLS slope estimates to be biased downwards.

2. Indicate for each of the following issues whether or not it causes bias for any coefficients in the estimated equation and if so, what (if anything) can be done about it:

    a) heteroscedasticity

    b) using the level of an x-variable when the true model has the natural log of the x-variable

    c) measurement error in the y-variable

    d) MA(2) in the error structure

3. Wesleyan hires you for the summer to study the factors that determine which alumni donate to Wesleyan and how much they donate. Explain what data you would want to use (be realistic about what data Wesleyan would have on alumni) and what you would do with the data. Anticipate complaints that econometrics professors might have when they review your study at the end of the summer, and explain how you will answer them.

4. Suppose $Y_1 = \beta_0 + \beta_1 Y_2 + \beta_2 X + \varepsilon_1$ and $Y_2 = \delta_0 + \delta_1 Y_1 + \varepsilon_2$
    and the reduced form estimates are $Y_1 = 2 + 4X$ and $Y_2 = 1 + 8X$

    a) Estimate the identified structural coefficients. Show your work.

    b) Assume that $\beta_1 = 0$ and estimate the identified structural coefficients. Show your work.

    c) Alternatively, assume that $\beta_0 = 0$ and estimate the identified structural coefficients. Show your work.

5. Suppose you want to test whether girls who attend a girls' high school do better in math than girls who attend coed schools. You have a random sample of senior high school girls from a state in the United States, and *score* is the score on a standardized math test. Let *girlhs* be a dummy variable indicating whether or not a student attends a girls' high school.

   a) What other factors would you control for in the equation?
      (Note: you need to be able to reasonably collect data on these factors.)

   b) Write an equation relating *score* to *girlhs* and the other factors you listed in (a).

   c) Suppose that parental support and motivation are unmeasured factors in the error term for (b). Are these likely to be correlated with *girlhs*? Explain.

   d) Suppose you have another variable, *numhs*, which is the number of girls' high schools within a 20-mile radius of a girl's home. What is necessary for this to be a good instrument for *girlhs*?

   e) Suppose you estimate the reduced form for *girlhs* and find that the coefficient on *numhs* is negative and statistically significant. Are you comfortable proceeding with IV estimation where *numhs* is used as an IV for *girlhs*? Explain.

6. Suppose you have quarterly data on new housing starts, interest rates, and real per capita income.

   a) Specify a model for new housing starts that includes accounting for possible trend and seasonality.

   b) In your model, why might you expect there to be autocorrelation even after accounting for trend and seasonality?

   c) What might you do to test for autocorrelation and, if you find it, what might you do to adjust your model to take it into account?

7. a) A friend has regressed kilograms of Brazilian coffee purchased on the real price of Brazilian coffee, the real price of tea, and real disposable income. She found the wrong sign on the real price of Brazilian coffee with a *t*-value of 0.5. She reestimated without the real price of Brazilian coffee and found little change in the other coefficient estimates, so she adopted the latter specification and concluded in her writeup that the demand for Brazilian coffee is price inelastic. Before handing her project in, she asks for your advice. What advice would you offer, and why?

   b) A friend has regressed output of an agricultural product on number of acres devoted to its production, amount of seed used, inches of rainfall during its growing season, and hours of sunlight during its growing season. He cannot get any reasonable results in terms of statistically significant coefficients with the expected signs. Before handing his project in, he asks for your advice. What advice would you offer, and why?

8. A student who has panel data for a set of countries on GDP growth and social spending runs a regression of GDP growth on social spending (and other variables) and gets the following coefficients for social spending under four different estimation strategies:

	OLS	FE	FE+Prais-Winsten	First-Differences	
β:	-.017	.070*	.089***	.507**	(***p<.01; **p<.05; *p<.10)

   a) What is the economic theory behind running an equation like this? Is it a structural equation?

   b) Why would the coefficient become positive when fixed effects are included (as in what does that imply about the relationship between different types of countries and social spending)?

   c) No additional information is available, but you need to choose an estimate to use for policy purposes. Which estimate do you choose to use, and why?

Final Exam: Essays (1 hour)

Your grade on this part of the exam will depend on the <u>coherence</u> and <u>completeness</u> of your answers and on your <u>demonstrating knowledge of the material covered in this course</u>. Write both essays in the same bluebook, but in a different bluebook than the answers to the problems.

1)   Explain the relationships between OLS and the following alternative estimators: IV, GLS, ML, GMM.

2)   Cases where the estimated model is not the true model; give several examples where this happens and discuss the problem(s) that occur in each case.