

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 fine to use to estimate the coefficients of a structural equation in a simultaneous equation model.
 - b) It is always better to use GLS than to use OLS when you suspect nonspherical errors.
 - c) First-differencing your data always adjusts fully for autoregression of order one.
 - d) You cannot include dummy variables in random effects models, as their coefficients will not be estimated.
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) inclusion of an extraneous variable in the regression (i.e., a variable that is not in the true model)
 - b) a lagged value of the independent variable used as one of the regressors
 - c) multicollinearity
 - d) AR(1) in the error structure
3. Wesleyan hires you for the summer to study the factors that determine which students decide to major in economics, and why. Explain what data you would want to use (be realistic about what data Wesleyan would have on students) 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 you have selected 900 individuals to participate in a negative income tax experiment. You have also selected 900 similar people to serve as a control group. Given observations on all these people's annual hours of work during the year before the experiment and during the year of the experiment, explain how to use a regression to estimate the difference in differences estimate of the impact of the negative income tax on hours worked.

5. You start your new job as a research assistant at an economic consulting firm. Your boss calls you in the first morning and tells you to estimate the average effect for the US of violent crime on residential housing prices — by tomorrow.
- What kind of data would you use to do this?
 - What equation would you estimate, and using what econometric method?
 - When you show your results to your boss, what complaint(s) might he have about your work?
 - If you were given unlimited time and money to do this project, what would you do instead?

6. Three of your friends come to you for advice on their Econ 385 projects. In each case, explain what you might tell them, stating why they should indeed be worried, what might happen to the coefficient if the problem were fixed, and what they might do to address the problem.
- Friend 1 runs a probit model where the Y variable is whether or not a woman gets a mammogram, and the X variable of interest is whether or not the person had health insurance coverage. The coefficient is positive and statistically significant. Your friend is worried about endogeneity.
 - Friend 2 runs an OLS regression where the Y variable is hourly earnings and the X variable of interest is whether or not the person is a queer male (=1 if queer). The coefficient is positive and statistically significant. Your friend is worried about omitted variable bias.
 - Friend 3 runs a fixed effects model where the Y variable is course grade and the X variable of interest is whether or not the person is a member of a fraternity or sorority (=1 if Greek). The coefficient is negative and statistically significant. Your friend is worried about sample selection bias.

7. A student who has 3-period panel data for a set of Tanzanian households runs a regression of household food expenditures on the occurrence of welfare shocks (and other variables) and gets the following coefficients on a shock to household welfare under three different estimation strategies:

	pooled OLS	random effects	fixed effects	
β :	.092	.121	-.37	
	(.053)	(.048)	(.057)	(robust standard errors)

- What is the economic theory behind running an equation like this? Is it a structural equation?
 - Is the pattern of the relative sizes of the three standard errors as you would expect? Explain.
 - 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?
8. A student who has time-series data for the US runs a regression of the Case-Shiller National Composite Home Price Index on the Homeownership Rate (and other variables) and gets the following coefficients on home ownership under three different estimation strategies:

	OLS	FGLS	ARIMA(4,1,0)	
β :	.668	1.530*	-.067	(* .01 < p < .05)

- What is the economic theory behind running an equation like this? Is it a structural equation?
- A Dickey-Fuller test where the Index is run on four lags and the time trend yields a test statistic of -4.2. What null hypothesis is being tested here, and what is the result?
- 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 coherence and completeness of your answers and on your demonstrating knowledge of the material covered in this course. 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) Discuss how sample selection can affect both estimation and interpretation of results, including how sample selection can be explicitly accounted for during estimation. Describe an example of a case in economics where sample selection occurs, and how you would deal with it in your case