

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) Ridge regression has lower MSE than OLS and is thus preferable to it.
 - b) 2SLS is always preferable to picking just one instrument to use out of a set of possible instruments.
 - c) A good way to deal with heteroskedasticity is to take the natural logs of all of your variables (dependent and independent) and run the equation on the logs instead of on the levels.
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) autocorrelated errors
 - b) a lagged value of the dependent variable used as one of the regressors
 - c) measurement error in one of the regressors
 - d) omission of the age-squared term in a wage regression (so only include age, not age and age²)
 - e) inclusion of a time trend in a time-series regression
3. Wesleyan hires you for the summer to study the factors that determine which admitted students actually come. Explain what data you would want to use (be realistic about what data Wesleyan would have on admitted 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. Wikipedia describes an **abstract** as "a brief summary of a research article...[that] is often used to help the reader quickly ascertain the paper's purpose" [[http://en.wikipedia.org/wiki/Abstract_\(summary\)](http://en.wikipedia.org/wiki/Abstract_(summary))]

The article goes on to say that an abstract:

"typically outlines four elements germane to the completed work:

- The research focus (i.e. statement of the problem(s)/research issue(s) addressed);
- The research methods used (experimental research, case studies, questionnaires, etc.);
- The results/findings of the research; and
- The main conclusions and recommendations"

Write an abstract for your project. Limit it to one paragraph and no more than 180 words.

5. A student runs the following equation:

$$\log(wkearn) = \beta_0 + \beta_1 09 + \beta_2 male + \beta_3 09 * male + \beta_4 white + \beta_5 09 * white + \beta_6 age + \beta_7 age^2 + \beta_8 educ + \beta_9 married + u$$

It is run using a pooled cross-section of data from 2008 and 2009, where the dependent variable is the natural log of weekly earnings; *09* is a dummy variable indicating the observation is from 2009; and *male*, *white*, *age*, *educ*, and *married* are what they sound like.

a) What is the student setting up to test, and how would one test it?

The results are (assume all coefficients are statistically significant):

$$\log(wkearn) = 2.20 - .04 \cdot 09 + .39 \cdot male - .02 \cdot 09 * male + .03 \cdot white + .06 \cdot 09 * white + .11 \cdot age - .0012 \cdot age^2 + .11 \cdot educ + .09 \cdot married \quad ; \quad R^2 = .35$$

b) A friend (who knows some economics, but knows nothing about econometrics) looks at the results. Explain this equation to your friend (explain all of it, not just what you discussed in a).

6. A student wants to estimate the likelihood that a student will drop out of high school using a large set of variables, but discovers that many of the variables that he wants to use have missing values for many of the students. In particular, information is more likely to be missing if the student is a dropout than if the student stays in school. Discuss what he should do in this situation.

7. A student runs a regression of the property crime rate on the unemployment rate (and other variables) using a panel of state-level data and gets the following coefficients for unemployment under four different estimation strategies:

	OLS	FE	IV	IV-FE
β :	275	54	-1491	281
SE(β):	(6.09)	(2.34)	(-2.96)	(1.67)
R ² :	.14	.18	.06	.17

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

b) Why would one want to use fixed effects and/or instrumental variables for this equation?

c) What other information would you like to know in order to help decide which estimate to prefer?

d) No other 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 runs a time-series regression of the S&P 500 index on oil prices (and other variables) and gets the following coefficients for oil prices under four different estimation strategies:

	OLS	AR(1)	GARCH	ID(1)	
β :	-7.62***	-3.16***	-2.18**	-0.06	(***p<.01; **p<.05)

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

b) Why would one want to use either AR(1), GARCH, or ID(1) to estimate this equation?

c) The Dickey-Fuller test statistic was .7326. What does this indicate to you? What other information would you like to know in order to help decide which estimate to prefer?

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