Empirical Project

For many students, this project becomes the kernel of their thesis (a good outcome) or a way to rule out a thesis topic that turns out *not* to be viable (also a good outcome!). Your project must use **methods developed in this class** (e.g., panel data methods, time series methods, limited dependent variables, instrumental variables, two-stage least squares).

You are strongly encouraged to meet with me to discuss your project.
Wooldridge has a great chapter (19) on carrying out an Empirical Project.

March 19 Submit table of summary statistics with source (name of datasets etc.)

April 30-May 1 15-minute presentations. Hard time limit.

If you are working with a professor on a thesis topic, you should (1) let me know and (2) not consult them on your presentation for this class.

The end product is a 15-minute presentation. You will submit your slides (including full tables) to me. Your slides will likely include:

I. Introduction:

- What is the main question of interest? This will likely be in the form of *"what is the effect of X on y?"*
- Is there a specific hypothesis you are testing?

  Note: Economic papers model the behavior of economic agents—individuals, state government agencies, political agents, firms, universities, etc. How does this agent make decisions? What is their objective? What are their constraints? Does your data represent the choices of only one agent or the interactions of several?

  You do not need to include a long theory section, but your regression equations should NOT represent "a list of variables I found that might be related to this topic" or even "a list of variables I found be statistically related."

- What is the motivation for this study? (Is there a disagreement in the literature? Is the basic correlation counterintuitive? Has this population not been examined before?)

II. Methodology:

- What are the key econometric/ identification issues? Will OLS results from a naïve regression identify a causal effect?

- Econometric model: Regression equation(s). Be precise with subscripts and notation. As appropriate:
  - Justification of variables used (often based on theory)
  - Justification of functional form (often based on theory)
  - Discussion of anticipated sign of coefficients (based on theory)
  - Treatment of standard errors (robust, clustered, adjusted for serial correlation, etc.)
- Potential empirical problems, your proposed solutions and justification for those solutions

III. Data:

- Describe the sources of your data.
- How does your data differ from data that would be "ideal" or from an "ideal experiment"?
- Summary statistics for key variables (formal table—not Stata output).
- Graphs of key relationship

IV. Results:

- Discuss and interpret basic OLS results
- Discuss and interpret alternative estimators (e.g., different functional forms, fixed effects, IV, 2SLS, Probit/Logit)
- Relevant hypothesis tests and interpret of their results
- Other robustness checks—e.g., other reasonable specifications, other ways to deal with outliers, etc.
- Conclusions about preferred model—of all the estimates which one is "best" and why?

All tables should be journal quality: Use full variable names, not abbreviations. Make the output "pretty"—no STATA output cut and pasted into the paper! Add notes to the bottom of the table so that tables can be interpreted standing alone.

V. Remaining problems/future work—at least 5 minutes of your presentation

I do *not* expect you to have results that will end up in your thesis—unless it is a relatively easy question, you will have many issues that still need to be resolved.

I DO expect you to be able to explain what problems remain.

- Are you lacking some key data? Sign the omitted variable bias.
- Are there econometric issues that you were not able to resolve?

- Do you have measurement error in some of your variables? Explain why and how your results are likely affected.
- Are the instruments weak/invalid? Discuss why that might be the case and sign the bias.

Econometrics II Final Empirical Project
Rubric

Project content (Equal weight on each component):	80%
1. Clearly stated main question of interest and the motivation for the study (how adds to literature/why is needed)	
2. Data and summary statistics well explained, using correlations, graphs, maps, tables as appropriate.	
3. Well-justified baseline econometric model. As relevant, discussion of variables in model, functional form, standard errors, etc. Precise subscripts and notation for equations.	
4. Thorough discussion of potential econometric/identification issues with baseline model. As appropriate, discussion of how OLS differs from "ideal experiment."	
5. Well-motivated alternative estimators and robustness checks. Discussion of how additional regressions add to understanding of question of interest.	
6. Accurate discussion and interpretation of results. Coefficients are contextualized to convey the economic relevance (e.g., relative to means, standard deviations, etc.)	
7. Appropriate hypothesis tests conducted. Precise interpretation of test results. Clear justification of preferred final model.	
8. Full discussion of remaining problems/directions for future analysis.	

Note:
If you read this rubric carefully, you will see that your grade does *not* depend on having results that have resolved all empirical issues.
Instead, you are expected to clearly lay out why your results are *unlikely* to be a final answer and what econometric/identification problems remain that future analysis would need to address to provide a defensible answer to your question of interest. The scores for (4), (6) and (8) will reflect this.

Presentation organization/style (Equal weight on each component):	20%
1. Organization: Content is clearly organized and easy to follow. Appropriate allocation of time for various sections. Pacing, delivery make the presentation engaging.	
2. Table and figures are easy to read and well documented. Each should be able to be interpreted as a "stand alone" item, with labels, units specified, no non-standard abbreviations, notes included at bottom.	