Syllabus for AEM 7100: Econometrics I
Spring 2020

Lecture Time: Thursdays 7:00PM - 10:00PM
Class Location: 150 Warren Hall
Professor: Shanjun Li
           405 Warren Hall; SL2448@cornell.edu; Phone: 607-255-1832
           Will use Canvas to distribute homework and materials
           Office Hours: Thursday 4:00-5:00 PM
TA: Deyu Rao <dr526@cornell.edu>
    Office hour: Tuesday from 2:00-3:00 PM, 363 Warren

Summary: This course provides an introduction to empirical methods that are commonly employed in applied microeconomic research. The course is applied in nature: our focus will be on learning the tools necessary for carrying out empirical work through hands-on data work and analysis; less efforts will be spent on proving theories and deriving properties of the estimators. My goal is to provide you with an understanding of some commonly used empirical methods that will allow you to make informed methodological decisions when conducting your own empirical work. ***The main leaning method of this course is learning by doing so I expect you to spend a lot of time on problem sets, programming, and doing research project. Coming to classes alone is very far from being enough.***

Textbooks: The required text for the course is *Econometric Analysis* by William Greene 7th or 8th Edition (4th, 5th or 6th should be ok too), Prentice Hall. Another required text is *Discrete Choice Methods with Simulations* by Kenneth Train. Download at http://elsa.berkeley.edu/books/choice2.html.

Other texts that might be useful are: *Econometrics of Cross-sectional and Panel Data* by Jeffrey Wooldridge and *Microeconometrics: Methods and Applications* by Colin Cameron and Pravin Trivedi.

Prerequisites: Matrix algebra and statistical methods courses at level of AEM 6940, ILRST 3110 or ECON 6190. Multivariate calculus, matrix algebra, probability and distribution theory, statistical inference, and an introduction to the multiple linear regression model. Appendices A and B in *Greene (editions 6 or 7)* are assumed.

Software: You should learn to use both Stata and Matlab. Stata is a powerful statistical program with a broad set of pre-programmed econometric and statistical tools. It is quite popular among economists, and is continuously being updated with new methods. It is an excellent package for most econometric analysis, but is limited when you want to use new or less-common econometric methods which have not yet been programed. Matlab is a high-level matrix programming language with a wide variety of built-in statistical functions. Many econometric methods have been programed in these languages and are available on the web. The advantage of using Matlab is that you are in complete control of your analysis, and it is easier to program new methods than in Statat. Some

disadvantages are that you have to do much of the programming yourself, programming complicated procedures takes significant time, and programming errors are hard to prevent and difficult to detect and eliminate.

In short, Stata could be very handy for data cleaning and routine analysis and Matlab is preferred for more sophisticated methods (not yet canned in Stata routines). I switch back and forth between the two depending on the specific project and in many cases, I use both. If you strongly prefer to stick with only one software, that is fine. But you will find that for more involved methods such as simulated MLE, Matlab will be a better tool than Stata. Moreover, Matlab has built in parallel processing in recent versions, which are very helpful for computationally intensive work.

For many exercises, I will ask you to code things up in Stata and Matlab. Data sets needed for the exercises will be distributed to the class via blackboard course website.

Evaluation: Grades for the course will be based on:
• Six or seven problem sets (total 30%)
• Midterm examination (20%), (time: the week after GMM chapter, the 2nd half of March)
• Group/individual empirical project (50%)

Group/individual Empirical Project: I expect each of you to work in groups of 2-3 students (if you strongly prefer work by yourself, that is fine as well). You should form your groups no later than March 1st and start to think about the project as soon as possible. If you could not find a group to join by March 1st, email my TA so that you can be helped. Ideally, you should find your own topic and data (this will be rewarded in the evaluation) and I would be happy to guide you through the process. But if you have trouble with this, talk to me no later than April 1st, and I can provide a set of topics and potential data sources. Another way to fulfill this requirement is to replicate an existing study. Many journals now require authors to post their data set and you can find an article to replicate from these journals (American Economic Review, Journal of Political Economy, American Economic Journals, Journal of Applied Econometrics etc.).

The paper should follow the standard format of a research paper in economics and should be 15-30 pages (no including references but all else) in font 12 and double spacing. **Due: 5PM on May 15th.** I need an electronic copy to TA (cc-ing me).

Topics:
1. OLS mechanics
2. Finite sample and large sample properties of OLS
3. Hypothesis testing
4. Regression extensions
5. Instrumental variable
6. Panel Data
7. GMM
8. System of Equations
9. MLE
10. Discrete choice model

Class Schedule:

Discuss Syllabus and Introduction (lecture 1)
1. OLS mechanics
   1.0: notations   (lecture 1)
   1.1: least squares derivations (lecture 1)
   1.2: projection matrix (lecture 2)
   1.3: regression components and partial regression (lecture 2)
   1.4: goodness of fit (lecture 2)
   Show Matlab and Stata sample files on OLS mechanics (lecture 2)

2. Finite sample and large sample properties of OLS
   2.0: distribution of random variables (lecture 3)
   2.1: finite sample properties of OLS (lecture 3)
   2.2: large sample properties
       2.2.1: primary tools of asymptotic theory (lecture 3 half way through)
       2.2.2: consistency, relationship between bias and consistency (lecture 4)
       2.2.3: asymptotic normality (lecture 4)
       2.2.4: estimation of asymptotic covariance matrix (lecture 5)
       2.2.5: asymptotic efficiency (lecture 5)
       2.2.6: function of parameters (lecture 5)
       2.2.7: t-statistic (lecture 5)
       2.2.8: confidence interval (lecture 5)

3. Hypothesis testing
   3.1.   Definition   (lecture 6)
   3.2.   Hypothesis testing decisions (lecture 6)
   3.3.   Method of single hypothesis testing (lecture 7)
   3.4.   Joint Hypothesis Testing (lecture 7)
   3.4.1.   Wald Test (lecture 7)
   3.4.2.   F-Test (lecture 7)

4. Regression extensions
4.0: Review of Assumptions and OLS properties (lecture 8)
4.1: Multicolinearity (lecture 8)
4.2: GLS and FGLS
4.2.1: Robust Var-cov (lecture 8)
4.2.2: GLS (lecture 8)
4.2.3: FGLS (lecture 8)
4.2.4: Test of Heteroskedasticity (lecture 9)
4.3 Functional Form – Nonlinearity (lecture 9)
4.3.1: Nonlinearity in variables (lecture 9)
4.3.2: Nonlinear regression Models (lecture 9 and lecture 10)

5. Instrumental variable
5.1 Properties of OLS under endogeneity (lecture 11)
5.2 Causes of Endogeneity (lecture 11)
5.3 Definition of Valid IV (lecture 11)
5.4 Exactly Identified Case
5.5 Overidentified case
5.6 Specification Tests
5.7 Notes on nonlinear model

6. Panel Data
7. GMM
8. System of Equations
9. MLE
10. Discrete choice model

Some Notes on Empirical Project for AEM 7100

- Please review the section in the course syllabus devoted to the Empirical Project
- Recapitulation of a few points from that document with additional thoughts:
  - Ideally you should find your own topic and data. One good approach is to use a dataset that is ready to go (perhaps from someone else's paper) to ask a question different from the one for which the data were created.
  - If you choose to replicate a study, you should still attempt to add your own insights through extensions using different methods or asking different empirical questions. If you are replicating a study, you should aim to replicate the core results and robustness checks (and any extensions in the paper that are interesting), but it is probably not a good use of time (read: the ratio of time invested to additional credit you will earn will be low) to replicate every single result in the paper and appendices for papers with very lengthy robustness checks (many empirical papers published in top journals now include 70-80 pages of robustness checks).
  - If you are still having trouble with your topic, you should arrange a meeting with Professor Li or the TA as soon as possible.
  - The paper should follow the standard format of a research paper in economics and should be 15-30 pages (no including references but all else) in font 12 and double spacing. The standard format for a research paper in Economics is:
    1. Introduction
    2. Literature Review
    3. Theoretical Model or Motivation (optional)
    4. Description of Data Used
    5. **Empirical Methodology**
    6. **Results**
       A. **Main Results**
       B. **Robustness Checks**
       C. **Extensions**
    7. Discussion and Economic Significance of Results
    8. Conclusion
- For the purposes of this paper, you should think of following this framework, but where sections 5 and 6 (in **bold**) may be disproportionally larger relative to the other sections compared to a published paper (depending on the paper). It may make sense to collapse sections 1 and 2 into a single section, and the same may be the case for sections 7 and 8.
- Depending on the questioned asked, your empirical methodology may invoke a theoretical model (e.g., a simultaneity bias problem may begin with an explanation of the supply and demand equations underlying the structural relationship before proceeding into an identification strategy to disentangle the effects),[1] or it may not[2] (especially where there is no clear economic model underlying the question asked). If you do invoke a theoretical model, you should only devote enough of your write-up to explain how the theoretical model motivates the empirical

methodology used.

- [ ] When you write a paper, you should always keep your readers in mind: you are writing for them: not yourself. What point are you trying to make for each sentence you write? Would what you write make sense to others who may not know as much as you for the specific question? What kind of questions would a reader have and then you need to address in writing?
- [ ] To that end, the tables and graphs in the paper should be self-explanatory. That is, a reader does not need to go to your paper in order to get a good idea on what these tables and graphs convey. You can add a note below the table and graph if you need to explain some of the details. But the title of the tables (and graphs) and variables/labels should be self-explanatory and convey as much information as possible. For example, from a regression result table, the read should know: what is the dependent variable? What kind of regressions are we looking at (OLS, IV, MLE…)? the unit of analysis (e.g., household-year, country-month…) and the number of observations.

In general, to decide on a topic of your research (for example your thesis or dissertation), you need to ask these three questions:

1. Why is it interesting/important (why should we care)? You can approach this by thinking about these questions: how does it inform the broad economic literature? What is the real-world examples (or the magnitude of the issue)? What are the policy implications? Is this a big (general-interest) question that people from different field would be interested in or a question that may only appeal to those in you field?
2. What is my contribution?
3. What are the (empirical) challenges and how should I address them?

Read some papers from top 5 economics journals or top field journals and see how good papers look like. Here are four great papers that you should look at and emulate in thinking about your research. They use different strategies including quasi-experiment (three first three), experiment (number 4), and structural method (number 5).

- [ ] [The Impact of Air Pollution on Infant Mortality: Evidence from Geographic Variation in Pollution Shocks Induced by a Recession](#) Kenneth Chay and Michael Greenstone, *Quarterly Journal of Economics*, 2003, 118(3): 1121-67.
- [ ] The Effect of Driving Restrictions on Air Quality in Mexico City Lucas Davis, *Journal of Political Economy*, 2008, 116(1), 38-81. http://faculty.haas.berkeley.edu/ldavis/Davis%20JPE%202008.pdf
- [ ] [Do Consumers Respond to Marginal or Average Price? Evidence from Nonlinear Electricity Pricing](#), Koichiro Ito. *American Economic Review*, 104(2): 537-63, 2014. Abstract | Slides | NBER WP | EI WP
- [ ] "The Short-Run and Long-Run Effects of Behavioral Interventions: Experimental Evidence from Energy Conservation." Allcott, Hunt, and Todd Rogers (2014). *American Economic Review*, Vol. 104, No. 10 (October), pages 3003-3037.

Download PDF. Download replication code.

☐ "Market-Based Emissions Regulation and Industry Dynamics" Meredith Fowlie, Mar Reguant and Stephen Ryan. April 2014. Forthcoming, *Journal of Political Economy*. [PDF] [Appendix]

Useful references:

On how to avoid plagiarism:
http://www.ox.ac.uk/students/academic/guidance/skills/plagiarism

On dealing with graduate school stress:
http://www.uhs.berkeley.edu/bewell/grad%20student%20stress.pdf
http://thegradstudentway.com/blog/?p=1727

Coding and data practices:
https://web.stanford.edu/~gentzkow/research/CodeAndData.pdf

Check out Greg Mankiw's resources for economics research here (read each of the articles)
http://gregmankiw.blogspot.com/2006/05/advice-for-grad-students.html

John Cochrane's Writings Tips for PhD Students:
http://faculty.chicagobooth.edu/john.cochrane/research/papers/phd_paper_writing.pdf

Plamen Nikolov's article on writing tips that combine multiple sources including Cochrane's:
http://www.people.fas.harvard.edu/~pnikolov/resources/writingtips.pdf

---

[1] See, for example Black, et al. from February 2015 AER (link is ungated version) for an illustrative use of a theoretical model:
http://harris.uchicago.edu/sites/default/files/The%20Great%20Migration%20and%20Mortality%20201 4_05_05.pdf. Note the use of the phrase "to fix basic ideas and set the stage for empirical analysis to come," in the relevant section is exactly the approach that you should take if elaborating a theoretical model.
[2] See, for example, Currie & Tekin in February 2015 AEJ: EP (link is ungated version) for an empirical paper without a theoretical discussion:
http://www.princeton.edu/~jcurrie/publications/Foreclosure_Health_Jan_2014.pdf .