

# AEDE 7140 Applied Econometrics II

Spring 2018

**Class:** Tuesdays & Thursdays, 2:20-3:40pm, 246 Agriculture Admin Bldg. (AA)

**Lab:** Mondays, 9:35-10:55, AA computer lab

**Professor:** Leah Bevis, Assistant Professor, AEDE (bevisl6@osu.edu; find me in 329 AA)

**TA:** Danbee Song (song.900@osu.edu)

**Professor hours:** Tuesdays 1pm-2pm, Thursdays 3:45-4:45pm in 329 AA

**TA office hours:** Wednesday 2pm-4pm in 344 AA, or by appointment

**Class Summary:** This class will include some extensions of basic econometrics models, primarily in module I on identification, endogeneity, fixed effects and panel data. It will then move to quasi-experimental approaches, including differences-in-differences, propensity score matching, regression discontinuity, spatial discontinuities and random shocks. We'll then examine a few sparsely related topics: variable subset and collapse in data rich environments, exogeneity tests and causal bounds around non-exogenous treatments. We'll end with an overview and brief examination of machine learning techniques.

**Required Textbooks:** [AP] Angrist, Joshua and Jorn-Steffen Pischke (2009). Mostly Harmless Econometrics: An Empiricists Companion. [W1] Wooldridge, Jeffrey M. Econometric Analysis of Cross Section and Panel Data, 2<sup>nd</sup> Edition.

**Supplementary Textbooks:** [W2] Wooldridge, Jeffrey M. Introductory Econometrics: A Modern Approach, 6<sup>th</sup> Edition. [G] Greene, William H. Econometric Analysis 7<sup>th</sup> Edition.

**Articles:** Required articles are listed by topic below, with full reference information below that. They are also posted on Carmen.

**Class Slides:** I will try to post weekly beamer presentations to the server every Monday evening. It's possible that sometimes this won't happen, but hopefully most Tuesday and Thursday mornings you can print and/or examine the beamer presentations prior to each class. I urge you to print them, if this will help you with note-taking. No point writing down a slew of equations that can be easily accessed via the slides.

**Class-Specific Reading:** Every class has required reading, and some have supplementary (optional) reading. I have tried to limit Wooldridge readings to only the subsections most central to the concept. Mostly Harmless readings are harder to shorten, but the context tends to be very useful. For most of you, reading these textbook sections (or sometimes assigned articles) *before* lecture will greatly improve your ability to engage with lecture material.

**Problem Sets:** Each problem set asks you to work through questions regarding a particular dataset, often related to an existing, published paper. Assignment data will be made available

on your server, and both Stata and R are available on the server for analysis. I don't mind if you collaborate as you work on the problem sets --- in fact, I encourage it. However, you must each turn in your own, individual results/answers and the Stata/R code that you used to procure those results. These answers may not be identical across students, though I realize that you'll likely give similar answers to the other people in your group.

All homework assignments will be handed in via Carmen, and grades/comments will also be returned via Carmen. If you are unable to complete a homework assignment on time due to a serious constraint, please let me know at least 3 days in advance, and we can discuss an extension. I am generally willing to grant extensions if you ask in advance.

**Replication:** Rather than a final, I require you to replicate a paper. To do this, you must choose a paper for which the data are available. You may NOT procure the code from the author. (If you do this and we catch you, you will fail the course and likely leave the program.) You will submit the paper and the data to me by April 2; if it's not too simple or too complex, I'll approve it. Ideally, you should choose a paper focused on causal identification, but I don't require this. Only the first three modules of this course have accompanying problem sets; this allows you to work on the replication for the final weeks of the class. It will be due on May 1.

**Server:** Because I require you to work with data for each problem set and the final replication, a server is available for each of you. This server will hold the problem set data and perhaps some sample code in Stata and/or R, from lecture or lab. You can access the server from any computer through Desktop Remote Connect or Microsoft Remote Desktop --- both freely available online. You do need to have an internet connection to do so, however.

**Students with Disabilities:** If you have a disability that requires special accommodations or modifications of some kind, please notify me within the first week of class. I'll do my best to accommodate you.

**Evaluation:** Grades will be based primarily on the problem sets (20% each, 60% total) and your final replication (30%), but also on class participation (10%). By class participation, I mean questions and discussion. I add in a grade for this because everyone learns econometrics better when questions, confusions and ideas are freely and openly offered. If you are a shy person, try to ask a question at least once a week. If you are a verbose person, try to ask your peers questions, or sometimes just listen --- that's also a form of participation.

## Topics and Assigned Readings

### Module 1: Extensions on Identification

- March 1: Endogeneity and Identification
  - Required: [AP] Ch 1, 2, 3.1, 3.2
  - *Topic: The potential outcome approach, causal identification*
- March 5 Lab: Stata Commands, Interfacing with Latex
- March 6: More on Instrumental Variables
  - Required: [AP] Ch 4 (IV)
  - Recommended: [WI] Ch 5.1.1 (IV motivation), 5.1.2 (2SLS), 5.2.3 (IV pitfalls), 5.3 (IV and omitted vars, measurement error)

- Required: Christian & Barrett (2018)
- *Topic: Review of IV, relevance & exclusion requirements, problems with time-trends and geographic pattern in, heterogeneous treatment effects*
- March 8: Propensity Scores; Panel Data and Fixed Effects
  - Required: [AP] Ch 5.1 (fixed effects), 8.3 (clustering in panels)
  - Recommended: [WI] Ch 10.1 (omitted variables problem), 10.5.1 (consistency of the FE estimator), 21.3.3 (propensity score methods)
  - **NOTE:** This class is at 11:10am, in Animal Sciences 210.
  - *Topic: propensity scores; Review of w/in vs. between variation, endogeneity threats under w/in vs. between, adding time/group dummies to FE, clustering under FE*
- Homework: *Endogeneity and Causality in Grandmothers and Granddaughters (Duflo)*
  - Relevant Reading: Grandmothers & Granddaughters (Duflo 2003)
  - Due: March 22, 8am

## Module 2: Parallel Universes

- March 19 Lab: TA-run lab to begin problem set 2
- March 20: TA-run discussion of assigned podcasts
  - Required: 2 EconTalk podcasts – “Susan Athey on machine learning, big data, and causation” & “Andrew Gelman on Social Science, Small Samples, and the Garden of the Forking Paths”
  - *Topic: Validity of p-values across multiple hypotheses, multiple analysis versions, and in small samples, causation in machine learning*
- March 22: Differences-in-Differences, Regression Discontinuity
  - Required: [AP] Ch 5.2 (DD), Ch 6 (RD)
  - Required: [WI] Ch 6.5 (pooled cross section and DD)
  - *Topic: DD, synthetic controls as a DD extension, RD, validity checks*
- March 26 Lab: Working with geospatial data in R
- March 27: Space: The Final Frontier
  - Required: Tarozzi & Deaton 2009 REStat; Antilla-Hughs & Hsiang 2013
  - *Topic: Spatial discontinuities for ID, random (climate) shocks, small area estimation and constraints on SAE validity; the role of spatial homogeneity for valid inference*
- Homework: *Natural Experiments in Deforestation (Alix-Garcia)*
  - Relevant Reading: The Ecological Footprint of Poverty Alleviation (Alix-Garcia et al. 2013)
  - Due: April 2 8am

## Module 3: Multi-dimensionality & Difficult Identification

- April 2 Lab: Running Lasso in R
- April 3: Data Rich Environments (Subsetting, Collapsing)
  - Required: Belloni, Chernozhukov & Hansen 2014 JEP; Bai & Ng 2009 JTSE
  - *Topic: PCA and SVA, Lasso Review, choosing “k”*
- April 5: Testing Exogeneity
  - Required: Oster 2017 JBES

- Required: Altonji, Elder & Taber 2005 JPE
- *Topic: Various balance tests, AET, Oster's bounds, placebo tests*
- April 9 Lab: Supervised Replication Work
- April 10: Non-Exogenous Instruments
  - Required: Conley, Hansen & Rossi 2012 ReStat
  - *Topic: Conditional IV, bounding the bias, IV for measurement error*
- Homework: *Using PCA and Lasso to Lower First Stage Dimensionality*
  - Due: April 16 8am

#### Module 4: Difficult Data and Dependent Variables

- April 12: Binary and Multivariate Outcomes
  - Required: [AP] Ch 3.4.2 (limited dep vars & marginal effects)
  - Required: [WI] Ch 15.8.3 (unobs effects of logit models), 16.2.1, 16.2.2 (multinomial response), 16.3.1 (ordered response), 18.1 (count) 18.2.1 (poisson)
  - *Topic: Conditional logit, multinomial logit/probit, ordered logit/probit, poisson*
- April 16 Lab: Supervised Replication Work
- April 17: Differentiating Censoring, Corner Solutions, Selection
  - Required: [WI] Ch 17.1 (motivation), 17.2, 17.3 (type I tobit) 17.6 (2-part models and type 2 tobit)
  - Recommended: Ricker-Gilbert, Jayne & Chirwa 2011 AJEA
  - *Topic: Censoring and corner solutions and tobit, differentiating from selection, double and triple hurdle models*

#### Module 5: A Bit on Machine Learning

- April 19: Machine Learning
  - Required: Mullainathan & Speiss 2017 JEP, Athey & Imbens 2017 JEP
  - *Topic: Supervised vs. unsupervised, and overview of ML groups; heirarchical and non-heirarchical clustering, random forests, endogenous determination of groups / mixture models; choosing k (w/in all topics).*

#### Replication

- Paper proposal due April 2 (can be submitted earlier, if you wish)
- Final replication due May 1 (no exceptions)

### Assigned Articles

Alix-Garcia, J., McIntosh, C., Sims, K.R. and Welch, J.R., 2013. The ecological footprint of poverty alleviation: evidence from Mexico's Oportunidades program. *Review of Economics and Statistics*, 95(2), pp.417-435.

Altonji, J.G., Elder, T.E. and Taber, C.R., 2005. Selection on observed and unobserved variables: Assessing the effectiveness of Catholic schools. *Journal of Political Economy*, 113(1), pp.151-184.

Anttila-Hughes, J. and Hsiang, S., 2013. Destruction, disinvestment, and death: Economic and human losses following environmental disaster. Working Paper.

Athey, S. and Imbens, G.W., 2017. The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2), pp.3-32.

Belloni, A., Chernozhukov, V. and Hansen, C., 2014. High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives*, 28(2), pp.29-50.

Christian, P.J. and Barrett, C.B., 2017. Revisiting the effect of food aid on conflict: A methodological caution. World Bank Policy Research Working Paper 8171.

Conley, T.G., Hansen, C.B. and Rossi, P.E., 2012. Plausibly exogenous. *Review of Economics and Statistics*, 94(1), pp.260-272.

Duflo, E., 2003. Grandmothers and granddaughters: old-age pensions and intrahousehold allocation in South Africa. *The World Bank Economic Review*, 17(1), pp.1-25.

Mullainathan, S. and Spiess, J., 2017. Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2), pp.87-106.

Ng, S. and Bai, J., 2009. Selecting instrumental variables in a data rich environment. *Journal of Time Series Econometrics*, 1(1).

Oster, E., 2017. Unobservable selection and coefficient stability: Theory and evidence. *Journal of Business & Economic Statistics*, pp.1-18.

Ricker-Gilbert, J., Jayne, T.S. and Chirwa, E., 2011. Subsidies and crowding out: A double-hurdle model of fertilizer demand in Malawi. *American Journal of Agricultural Economics*, 93(1), pp.26-42.