SYLLABUS FOR ARE212

ETHAN LIGON & SOFIA VILLAS-BOAS

ARE212 is the second class in the graduate econometrics sequence in the Department of Agricultural and Rsource Economics after ARE210, which covers probability and statistics. In 2020 the course will be jointly taught by Professors Sofia Villas-Boas and Ethan Ligon.

Some knowledge of probability and statistics is assumed. Many student will have encountered this material in ARE210, but ARE210 is not a prerequisite for ARE212. However, familiarity with both linear algebra, probability, and statistics is assumed. In particular, if you have not had ARE210 be sure you're comfortable and familiar with this material presented at the level of

Linear algebra: Appendix A of Bruce Hansen's (2019) Econometrics textbook

Probability & Statistics: Bruce Hansen's (2019) Probability & Statistics textbook

If this material is largely new to you then ARE212 is *not* the class for you. We will not cover this background material in lecture. Consider EEP118 as an alternative.

1. People

Sofia Villas-Boas: `sberto@berkeley.edu`
Office Hours: Tuesdays 11–12
Location: University Hall 223 or 275

Ethan Ligon: `ligon@berkeley.edu`
Office Hours: `https://are.berkeley.edu/~ligon/appointment.html`
Location: University Hall 783

Ben Krause (GSI): `benjaminkrause@berkeley.edu`
Office Hours: Wednesdays & Thursdays 9–10
Location: University Hall 234

. *Date*: April 1, 2020.

2. Structure of Course

The first half of the course (taught by Villas-Boas) follows closely notes developed by Max Auffhammer for an earlier incarnation of this class, and also borrows from lecture notes for classes in Berkeley's Economics department (e.g., Econ 240A, Econ 240B, and first parts of Econ 241A and Econ 241B). A textbook treatment of much of this material can be found in Greene (2018).

The second half of the course (taught by Ligon) will rely on a variety of sources; the most relevant textbook treatment will be Hansen's (2019) econometrics text. Ligon will pick up where Villas-Boas leaves off, and will also introduce certain topics of particular interest including models of discrete choice, economic models and the generalized method of moments, as well as others (as time permits).

Both halves of the class will require you to do some programming using `R` and `python`. If you do not have any prior experience with these languages we strongly recommend taking an introduction at the D-lab (`https://dlab.berkeley.edu/calendar-node-field-date/month?type=training`). However, programming novices can also expect to receive help in section, and some programming tools will be covered in lecture. In Part I of the class we will guide you in both lecture and sections in the use of `R` studio and also go over `R` coding using `Jupyter` notebooks in several lectures.

Sections will be held on Fridays from 10:10–11:00. You should review any section notes in advance, and come with a laptop prepared to delve into some programming & problem solving.

3. Course Outline

We're doing this for the first time together, and reserve the right to make changes! But our plans for when and what to cover are as follows:

3.1. **The Classical Linear Regression Model (CLRM).**

- Classical Assumptions (WG Ch. 2.1 - 2.4)
- The Algebra of Least Squares (WG Ch. 3.1 - 3.6 (3.4 on your own))
- Finite Sample Properties of OLS (WG Ch. 4.1 - 4.6)
- Hypothesis Testing (WG Ch. 4.7, 6.1 - 6.3)
- Data Problems - Multicollinearity and Influential Observations (WG Ch. 4.9)
- Generalized Least Squares (10.1 - 10.2.1)
- Maximum likelihood estimation (Notes and WG Ch. 17)
- Asymptotic distribution theory (Notes and WG Appendix D)

- Large Sample Properties of OLS (WG Ch. 5.1 and 5.2)
- Hypothesis Testing - Large Sample Results (WG Ch. 6.4)

3.2. **Non-spherical disturbances.**

- Non-spherical Disturbances (WG Ch. 10.2.2, 10.3, 10.5)
  - Heteroskedasticity (WG Ch. 11)
  - Serial Correlation (WG Ch. 12)

3.3. **Multiple Equation Models.**

3.3.1. *General Reading.*

- Hansen Ch. 11. General contemporary reference on systems of linear regressions.
- Trygve Haavelmo. 1944. The probability approach in econometrics. *Econometrica* 12 (Supplement): 1–118. Classic discussion of the "probability approach" to estimating economic models; one of the key documents distinguishing econometrics from statistics; one of the key
- Carl F Christ. 1994. The Cowles Commission's contributions to econometrics at Chicago, 1939-1955. *Journal of Economic Literature* 32 (1): 30–59. `https://www.jstor.org/stable/pdf/2728422.pdf`. Discussion of the birth of econometrics at the Cowles Commission; beyond its historical importance, worth reading to understand the source of much of the jargon we now use.
- Stock and Trebbi 2003; S. Wright 1921; P. G. Wright 1928 (Appendix B). On the origins of instrumental variables. The Wright papers are some of the earliest and clearest discussions of identification.

3.3.2. *Restricted Estimation.*

- Hansen Ch. 8

3.3.3. *Endogenous variables.*

- Hansen Ch. 12

3.3.4. *Measurement error.*

3.3.5. *Seemingly Unrelated Regression.*

- Hansen Ch. 11.7

3.3.6. *Panel Data.*

- Hansen Ch. 17
- Yair Mundlak. 1961. Empirical production function free of management bias. *Journal of Farm Economics* 43 (1): 44–56
- Gary Chamberlain. 1984. Panel data. In *Handbook of econometrics,* edited by Zvi Griliches and Michael D. Intriligator. North-Holland

3.4. **Correlation & Causality.**

- Ragnar Frisch. 2011. *A dynamic approach to economic theory: lectures by Ragnar Frisch at Yale University.* Edited by Olav Bjerkholt and Duo Qin. Lectures given by Frisch at Yale in 1930. Routledge. Highly insightful and only recently published lectures Frisch gave at Yale in 1930.
- Trygve Haavelmo. 1943. The statistical implications of a system of simultaneous equations. *Econometrica, Journal of the Econometric Society:* 1–12. A founding document of the structural (Cowles) approach to economics.
- Trygve Haavelmo. 1944. The probability approach in econometrics. *Econometrica* 12 (Supplement): 1–118. Classic statement regarding the "probability approach" to economics.
- Arthur S. Goldberger. 1972. Structural equation methods in the social sciences. *Econometrica* 40 (6): 979–1001. Insightful, historically informed reflections on identification of structural models in economics.
- Judea Pearl. 2015. Trygve Haavelmo and the emergence of causal calculus. *Econometric Theory* 31 (1): 152–179; James Heckman and Rodrigo Pinto. 2015. Causal analysis after Haavelmo. *Econometric Theory* 31 (1): 115–151; Judea Pearl. 2013. Reflections on Heckman and Pinto's "causal analysis after haavelmo". Unpublished working paper. `https://escholarship.org/content/qt5b27h1nm/qt5b27h1nm.pdf`. Back and forth between Pearl & Heckman/Pinto. They agree on one thing: Frisch & Haavelmo were right.
- Judea Pearl. 2009. *Causality.* Models, Reasoning, and Inference. Second. New York: Cambridge University Press. Pearl's textbook on causality.
- Milton Friedman. 1953. The methodology of positive economics. In *Essays on positive economics,* 3–43. Chicago: University of Chicago Press. Classic essay on the usefulness of "falsifying" economic models and the role of assumptions.

- Stanford Encyclopedia of Philosophy, "The Problem of Induction"
- Hume, An enquiry concerning human understanding
- Sewell Wright https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1200501/pdf/111.pdf
- Guido W Imbens and Donald B Rubin. 2015. *Causal inference in statistics, social, and biomedical sciences.* Cambridge University Press. Up-to-date statement and discussion of the Neyman-Rubin potential outcomes model, with focus on the assignment model.
- James J Heckman. 2010. Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic literature* 48 (2): 356–98. Contemporary survey comparing "reduced form" and "structural" approaches to causal inference.
- James J. Heckman. 1997. Instrumental variables: a study of implicit behavioral assumptions used in making program evaluations. *The Journal of Human Resources* 32 (3): 441–462. http://www.jstor.org/stable/146178. Critique of potential outcome approach in program evaluation.

3.5. **Instrumental Variables (GMM).**

- Hansen Chs 12, 13
- Joshua D Angrist and Alan B Krueger. 2001. Instrumental variables and the search for identification: from supply and demand to natural experiments. *Journal of Economic perspectives* 15 (4): 69–85
- Timothy G Conley, Christian B Hansen, and Peter E Rossi. 2012. Plausibly exogenous. *Review of Economics and Statistics* 94 (1): 260–272
- Victor Chernozhukov and Christian Hansen. 2008. The reduced form: a simple approach to inference with weak instruments. *Economics Letters* 100 (1): 68–71. https://faculty.chicagobooth.edu/christian.hansen/research/ch_weakiv_mar07.pdf

3.6. **Resampling & the Bootstrap.**

3.7. **Cross-Validation & Machine Learning.**

- Mullainathan, Sendhil, and Jann Spiess. 2017. "Machine Learning: An Applied Econometric Approach." Journal of Economic Perspectives 31(2): 87-106.
- Hansen Ch. 23

3.8. **Discrete Choice.**

- Philipp Eisenhauer, James J Heckman, and Edward Vytlacil. 2015. The generalized Roy model and the cost-benefit analysis of social programs. *Journal of Political Economy* 123 (2): 413–443
- Eric French and Christopher Taber. 2011. Identification of models of the labor market. In *Handbook of labor economics,* 4:537–617. Elsevier
- James J Heckman. 2010. Building bridges between structural and program evaluation approaches to evaluating policy. *Journal of Economic literature* 48 (2): 356–98
- Heckman-Taber (2010) (New Palgrave)
- Wooldridge (2012) Ch. 19

4. Discussion Sections

4.1. **Discussion Section Objectives.** To be achieved through participatory classroom engagement:

- Introduce you to data analysis in R and Rstudio for the purposes of econometrics
- Build the coding skills necessary through explanation and example to be successful in this class
- Provide a foundation and common language for you to be able to teach yourself the various tools for data analysis beyond the needs of this class

4.2. **What to expect in Discussion Section.** This particular Discussion Section has been developed iteratively over many years and as such rests on a uniquely well developed body of section notes. Each week I will provide the updated notes for the week by Monday so that you have time to review them. On Friday, I will start each class asking for questions or clarifications from your experience reviewing the notes ahead of time. Where appropriate I will jump first into those areas you highlight otherwise I will make sure we save time in class to review them before the end. I will then march through the examples of code provided. **I do not expect to ever complete all of the notes in a given Discussion Section**. Rather, my goal is to make sure you understand enough of the notes so that you can complete them on your own. In this way I ensure that you receive as much material as possible to successfully navigate this class (and your future data analysis needs).

4.3. **Assignments and Grading Policy.** Problem sets will be graded based on the following criteria:

Grade	Percent assigned	Criteria
Check-Plus	100	Assignment complete, all questions answered thoroughly
Check	70	Almost all questions attempted
Check-Minus	50	Most questions attempted
0	0	Less-than-half questions attempted or no submission

5. Grading

The class grade will depend on[1]

Problem sets in Part I: 20%
Midterm (March 11, in class): 30%
Discussion in Part II: 20%
Final: 30%

1. But see discussion of student tutors below.