Ec 143 - Econometrics: Advanced Methods and Applications

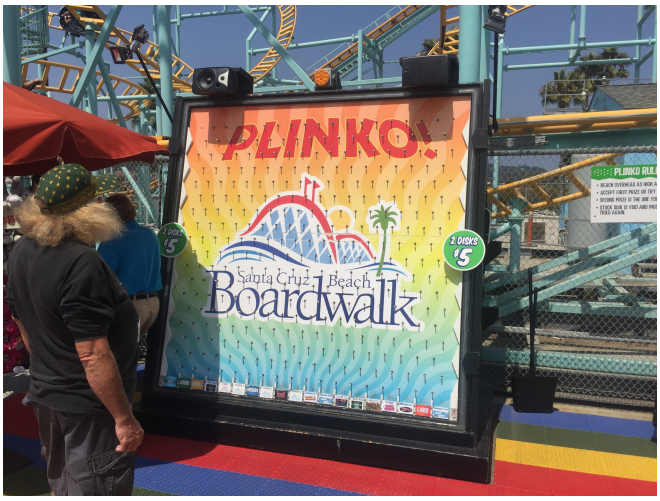
Department of Economics, UC - Berkeley

Spring 2022

Course Description

Catalog: This course introduces selected advanced data analysis and inference methods appropriate for economic data. Methods are taught in tandem with real world applications as encountered in policy analysis, industry and consulting work. Equal weight is given to theoretical development, computation and application. Exact topics and applications may vary across offerings. ECON C142 and 143 may be taken independently or together in any order.

Spring 2022 postscript: You will develop skills in empirical model building; loosely, the development of bespoke methods of data analysis for answering specific economic questions. In this semesters' offering we will (i) develop instrumental variables methods suitable for systems of simultaneous equations and apply them to the analysis of excise taxes (e.g., gas or soda taxes) and city growth in spatial equilibrium, (ii) discrete choice methods and their application to non-market evaluation of environmental services, (iii) methods of model selection and their use in nonparametric curve estimation with an application to calorie demand/hunger measurement, (iv) multi-step estimation problems with applications to understanding market power in oligopolistic settings, production function estimation, and explaining productivity differences across firms.



The central limit theorem at the beach!

https://en.wikipedia.org/wiki/Bean_machine

Course Logistics

Instructor: Bryan Graham, Department of Economics, University of California – Berkeley

Email: bgraham@econ.berkeley.edu

Time & Location: Tu/Th 11AM to 12:30PM, TBD

Office Hours: Thursdays 2PM to 3:40PM

Graduate Student Instructor: TBD, e-mail: bgraham@econ.berkeley.edu

Prerequisites: (i) A first course in econometrics, intermediate statistics or intermediate data science (Ec 140, Ec 141, ENVECON C118, DS 100 *or* STAT 135); (ii) Linear algebra (Math 54, Stat 89A *or* EE 16A). Exposure to economic theory at an intermediate level (e.g., Ec 100A, 101A etc.) is preferred, but not required. Prior exposure to scientific computing is also helpful, but not required.

Big Picture: I hope you will find this class interesting and challenging (i.e., difficult). At the end I hope you will feel a sense of accomplishment, as well as ownership over some new and valuable skills. I do *not* want you to find the class stressful. I am mindful that difficulty and stress often go hand-in-hand, but with some thoughtfulness on our part we can avoid this. While I will set and maintain high academic standards, I will also do my very best be supportive, encouraging and helpful. I also strongly urge you – the students – to try to be supportive, encouraging and helpful *to one another*. You'll have more fun (and learn more) if you work together. If a classmate reaches out for help, be generous. You will not regret it.

Grading: Grades for the class will equal a weighted average of those on homework (40%), the two mid-terms (30%) and your final project (30%). The in class midterm examinations will be held on March 17th, 2022 and the second on April 28th, 2022. Each midterm exam is graded on a scale from 0 to 100. I will only retain your highest midterm grade. Students that take both midterms (and get comparable/serious grades on both) will receive an additional bonus to their midterm grade component of 10 points. Consequently the highest available midterm aggregate is 110.

There will be 5 homework assignments. Homeworks are due at 5PM on the assigned due date (the GSI may elect to make small modifications to all things homework related). Homeworks are graded on a ten point scale with one-half point off per day late. You are free, indeed encouraged, to work in groups but each student must submit an individual write-up and accompanying Jupyter Notebook (when required; see below). Your lowest homework grade will be dropped, with the average of the remaining scores counting toward your final grade. There will be no 'make-up' midterms. I will add 5 points to homework aggregates for students who make serious efforts to complete all five problem sets (concretely this means that students may amass up to 45 homework points). The due dates for the five problem sets are:

Problem Set	Due Date	Topics
1	February 7th	Instrumental Variables – Gasoline Demand
2	February 28th	Discrete Choice – WTP for Wildfire Risk Reduction
3	March 28th	Model Selection – The Demand for Calories
4	April 11th	Sequential Estimation – Price-Cost Margins in Semiconductors
5	April 25th	Panel Data – The Returns to R&D in Pharmaceuticals

Since there is no assigned textbook for the class, good note-taking is essential.

Your final project will be due on the day of the final exam for this class' final exampl group.

Overall numerical course grades will be calculated as follows:

$$\text{Grade} = 40 \times \frac{\text{Homework Points}}{45} + 30 \times \frac{\text{Midterm Points}}{110} + 30 \times \frac{\text{Final Project}}{100}.$$

Numerical grades will be mapped into letter grades. A default mapping is 100 - 97 A+, 93 to 96 A, 90 to 92 A-, 87 to 89 B+, 83 to 86 B, 80 to 82 B- and so on. In practice grades are sometimes curved (particularly depending on the difficulty of the two midterms).

Textbook: There is no required textbook for this class. Wasserman (2004) is a good, albeit challenging, reference. For a review of basic concepts in probability, the first few chapters of Mitzenmacher & Upfal (2005) are helpful. The recent book by Efron & Hastie (2016) is also a useful/fun reference, available online for free, and in hard copy form at a reasonable price. Your introductory statistics and Ec140/141 textbooks will also be useful references (indeed access to the Wooldridge or Stock and Watson textbook would be useful). While I will occasionally make lecture notes available to students via a course GitHub repository, students should plan on taking *detailed* notes on the material presented during lecture. If you miss class for any reason please be sure to get notes from a classmate. Good note-taking is essential for doing well in this course. Assigned readings are given in the course outline below. Any reading not available online (possibly via Oskicat) will be made available via the GSI.

Computation: All computational work should be completed in Python. Python is a widely used general purpose programming language with good functionality for scientific computing. There are lots of ways of accessing Python. We will use https://datahub.berkeley.edu for computation. More information will be provided in section on how to access and use this platform. For those wishing to manage a Python environment on their personal computer, the Anaconda distribution, which is available for download at

https://www.anaconda.com/products/individual

is a convenient way to get started. Some basic tutorials on installing and using Python, with a focus on economic applications, can be found online at `https://quantecon.org/`.

Good books for learning Python, with some coverage of statistical applications, are Guttag (2013), VanderPlas (2017), and McKinney (2017). These books are available online via http://oskicat.berkeley.edu/.

Any code I provide will execute properly in Python 3.6, which is (close to) the latest Python release. There are a large number of useful resources available for learning Python on campus (including classes at the D-Lab).

While issues of computation may arise from time to time during lecture, I will not teach Python programming. *This is something you will need to learn outside of class* (although help will be provided in section). I do not expect this to be easy. I ask that those students with strong backgrounds in technical computing to assist classmates with less experience. Problem sets will be more fun if you all work together and assist each other.

Extensions: Routine extensions for assignments will not be granted (i.e., only in exceptional circumstances). The penalty for lateness is relatively minor and I also drop the lowest homework grade. These features are designed to allow you some flexibility in workload management during busy times of the semester. Late work, in addition to being undesirable for the individual student, can delay your classmates getting feedback. Please do your best to start work well before the due date.

Accommodations: Any students requiring academic accommodations should request a 'Letter of Accommodation' from the Disabled Students Program at http://dsp.berkeley.edu/ *immediately.* I will make a good faith effort to accommodate any special needs conditional on certification.

Academic Integrity: Please read the Center for Student Conduct's statement on Academic Integrity at http://sa.berkeley.edu/conduct/integrity. We should all take issues of intellectual honest *very* seriously.

Additional notes: I prefer to avoid having substantive communications by e-mail. Please limit e-mail use to short (ideally yes/no) queries. I am unlikely to be able to respond to a long/complex e-mail. Don't be shy about approach me with questions immediately before/after class. For longer questions please make use of my office hours. This is time specifically allocated for your use; please come by (I will also be available via Zoom virtually during office hours)! I look forward to getting to know all of you.

Course Outline

(Additional references may be added)

	Date	Topic	Readings
Week 1	Tu 1/18 Th 1/20	Introduction/ Computation	Shen (2014)
Week 2	Tu 1/27 Th 1/28	*Simultaneous* *Equations*	Graddy (1995) Angrist et al. (2000)
Week 3	Tu 2/1 Th 2/3	Excise Tax Incidence	Pirog (2011) Cawley & Frisvold (2017)
Week 4	Tu 2/8 Th 2/10	City Growth	Rauch (1993) Shapiro (2006)
Week 5	Tu 2/15 Th 2/17	*Binary Choice*	Amemiya (1981) Efron & Hastie (2016, Ch. 8)
Week 6	Tu 2/22 Th 2/24	Contingent Valuation	Carson (2012) Loomis & González-Cabán (2008)
Week 7	Tu 3/1 Th 3/3	*Model* *Selection*	Efron (2004), Abadie & Kasy (2019) Efron & Hastie (2016, Ch. 12)
Week 8	Tu 3/8 Th 3/10	Engel Curves, Calorie Demand & Hunger	Ye (1998),Efromovich (1999) Subramanian & Deaton (1996)
Week 9	Tu 3/15 **Th 3/17**	Review/Catch -Up **Midterm 1**	
Week 10	Tu 3/22 Th 3/24	*Spring Recess*	
	Tu 3/29 Th 3/31	*GMM &* *Sequential Estimation*	Newey (1984)
Week 11	Tu 4/5 Th 4/7	Measuring Market Power in Oligopolies	Bresnahan (1989),Puller (2009) Genesove & Mullin (1998),Baker & Bresnahan (2008)
Week 12	Tu 4/12 Th 4/14	*Linear* *Panel Data*	Griliches & Mairesse (1998) Blundell & Bond (2000)
Week 13	Tu 4/19 Th 4/20	Firm Productivity	Syverson (2011),Brynjolfsson & Hitt (2003) de Loecker (2013),Bloom et al. (2013)
Week 14	Tu 4/26 **Th 4/28**	Review/Catch -Up **Midterm 2**	

References

Abadie, A. & Kasy, M. (2019). Choosing among regularized estimators in empirical economics: the risk of machine learning. *Review of Economics and Statistics*, 101(5), 743 – 762.

Amemiya, T. (1981). Qualitative response models: A survey. *Journal of Economic Literature*, 19(4), 1483 – 1536.

Angrist, J. D., Graddy, K., & Imbens, G. W. (2000). The interpretation of instrumental variables estimators in simultaneous equations models with an application to the demand for fish. *Review of Economic Studies*, 67(3), 499 – 527.

Baker, J. B. & Bresnahan, T. F. (2008). *Handbook of Antitrust Economics*, chapter Economic evidence in antitrust: defining markets and measuring market power, (pp. 1 – 42). The MIT Press: Cambridge, MA.

Bloom, N., Schankerman, M., & van Reenen, J. (2013). Identifying technology spillovers and product market rivalry. *Econometrica*, 81(4), 1347 – 1393.

Blundell, R. & Bond, S. (2000). Gmm estimation with persistent panel data: an application to production functions. *Econometric Reviews*, 19(3), 321 – 340.

Bresnahan, T. F. (1989). *Handbook of Industrial Organization*, volume 2, chapter Empirical studies of industries with market power, (pp. 1011 – 1057). North-Holland: Amsterdam.

Brynjolfsson, E. & Hitt, L. M. (2003). Computing productivity: firm-level evidence. *Review of Economics and Statistics*, 85(4), 793 – 808.

Carson, R. T. (2012). Contingent valuation: a practical alternative when prices aren't available. *Journal of Economic Perspectives*, 26(4), 27 – 42.

Cawley, J. & Frisvold, D. E. (2017). The passthrough of taxes on sugarsweetened beverages to retail prices: the case of berkeley, california. *Journal of Policy Analysis and Management*, 36(2), 303 – 326.

de Loecker, J. (2013). Detecting learning by exporting. *American Economic Journal: Microeconomics*, 5(3), 1 – 21.

Efromovich, S. (1999). *Nonparametric Curve Estimation.* New York: Springer.

Efron, B. (2004). The estimation of prediction error. *Journal of the American Statistical Association*, 99(467), 619 – 632.

Efron, B. & Hastie, T. (2016). *Computer Age Statistical Inference.* Cambridge: Cambridge University Press.

Genesove, D. & Mullin, W. P. (1998). Testing static oligopoly models: conduct and cost in the sugar industry, 1890-1914. *Rand Journal of Economics,*, 29(2), 355 – 377.

Graddy, K. (1995). Testing for imperfect competition at the fulton fish market. *Rand Journal of Economics,*, 26(1), 75 – 92.

Griliches, Z. & Mairesse, J. (1998). *Econometrics and Economic Theory in the 20th Century*, chapter Production functions: the search for identification, (pp. 169 – 203). Cambridge University Press: Cambridge.

Guttag, J. V. (2013). *Introduction to Computation and Programming Using Python*. Cambridge, MA: The MIT Press.

Loomis, J. & González-Cabán, A. (2008). *The Economics of Forest Disturbances*, volume 79 of *Forestry Sciences*, chapter Contingent valuation of fuel hazard reduction treatments. Springer: Dordrecht.

McKinney, W. (2017). *Python for Data Analysis*. Cambridge: O'Reilly.

Mitzenmacher, M. & Upfal, E. (2005). *Probability and Computing*. Cambridge: Cambridge University Press.

Newey, W. K. (1984). A method of moments interpretation of sequential estimators. *Economics Letters*, 14(2-3), 201 – 206.

Pirog, R. (2011). *The Role of Federal Gasoline Excise Taxes in Public Policy*. Technical Report R40808, Congressional Research Services.

Puller, S. L. (2009). Estimation of competitive conduct when firms are efficiently colluding: addressing the corts critique. *Applied Economics Letters*, 16(15), 1497–1500.

Rauch, J. E. (1993). Productivity gains from geographic concentration of human capital: evidence from the cities. *Journal of Urban Economics*, 34(3), 380 – 400.

Shapiro, J. M. (2006). Smart cities: quality of life, productivity, and the growth effects of human capital. *The Review of Economics and Statistics*, 88(2), 324 – 335.

Shen, H. (2014). Interactive notebooks: sharing the code. *Nature*, 515(7525), 151–152.

Subramanian, S. & Deaton, A. (1996). The demand for food and calories. *Journal of Political Economy*, 104(1), 133 – 162.

Syverson, C. (2011). What determines productivity. *Journal of Economic Literature*, 49(2), 326 – 365.

VanderPlas, J. (2017). *Python Data Science Handbook*. Boston: O'Reilly.

Wasserman, L. (2004). *All of Statistics*. New York: Springer.

Ye, J. (1998). On measuring and correcting the effects of data mining and model selection. *Journal of the American Statistical Association*, 93(441), 120 – 131.