Ec 143 - Econometrics: Advanced Methods and Applications

Department of Economics, UC - Berkeley

Spring 2026

Course Description

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 academic research, 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 2026 : This semester we will learn (i) basics of Bayesian analysis and its application to online bandit algorithms (ii) MCMC methods and their application to code-breaking, (iii) proxy variable regression and the Bayesian bootstrap), (iv) model selection via unbiased risk estimation, (v) production function estimation using panel data (to study productivity differences across firms), (vi) quantile regression methods and their application to studying earnings inequality, and (vii) methods for the analysis of duration data (e.g., unemployment spells). Some last minute additions/subtractions to the above list are possible.

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 8AM to 9:30AM, Lewis 9

Office Hours: Tuesdays 2PM to 3:40PM, 669 Evans Hall

Graduate Student Instructor: Jinglin Yang, e-mail: jinglin.yang@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 also 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 and offer it. You will not regret it. I do not grade on a "curve". By helping a classmate you will improve both your own, as well as their, understanding. You will both learn and, also, both do "better" in terms of grades.

Grading: Letter grades for the class will be determined by a weighted average of scores on homework (30%), the two mid-terms (40%), a lecture scribing assignment (10%), and your final project (20%). The in class midterm examinations will be held on March 19th, 2026 and the second on April 30th, 2024 (i.e., the classes immediately prior to spring recess and RRR week). 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 score aggregate is 110. If either of the assigned midterm dates pose a problem (e.g., to athletes traveling those days, observance of a religious holy day), please contact me immediately so we can make alternative arrangements.

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. In the interest of providing timely feedback, homework will not be accept after five days from the assigned due date. You are free, indeed encouraged, to work in groups but each student must submit an individual write-up and/or 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. 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 (exact topics subject to possible change):

Problem Set	**Due Date**	**Topics**
1	February 4th	Probability/Bernoulli Bandits/Thompson Sampling
2	February 25th	Linear regression + Bayesian Bootstrap
3	March 11th	Model selection
4	April 8th	Quantile Regression – Earnings Inequality in Brazil
5	April 22nd	Discrete Hazard Analysis – Recidivism

Additionally, each student must sign up to scribe one lecture. You may work in groups as large as three. A scribing template will be shared on bCourses. Scribing assignments are due no later than two weeks after your assigned lecture and are graded on a 10 point scale. Your final project will be due on the day of the final exam for this class' final exam group (this year the due date May 14th at 10PM). More information about the project will be provided prior to spring break.

Overall numerical course grades will be calculated as follows:

$$\text{Grade} = 30\times\frac{\text{Homework Points}}{45}+40\times\frac{\text{Midterm Points}}{110}+10\times\frac{\text{Scribing Points}}{10}+20\times\frac{\text{Final Project Points}}{100}.$$

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). I do not curve to enforce a certain grade distribution. In past years I have found that 30 to 40 percent of students earn a grade of A- and above, 40 to 50 percent a grade of B- to B+, with the balance scoring lower. If student performance merits it, I am delighted to award more A's, likewise if student learning is less that expected, I will (reluctantly) award fewer As. One thing I want to emphasize is that it is optimal for you to help one another. If you understand the material you will earn a higher grade; helping a classmate will strengthen your understanding and also help them.

Collaboration and AI : You are welcome to collaborate with classmates on homework assignments; indeed this is encouraged. Each student, however, much turn in their own write-up. Please list the names of all students you collaborated with on your turned in homework. Likewise, if you used AI to assist you with your homework, please include a short (one paragraph) description of how you used AI and note which AI you used (e.g., Google Gemini, ChatGPT etc.). Note that a substantial portion of your final grade depends on performance on the in-class midterms. These are closed book, pencil and paper, exams. You are advised to use the homework to help your prepare for these exams. Please keep this in mind as your collaborate with peers and/or consult AI.

Textbook: There is no required textbook for this class, ***good note-taking is essential.*** 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 very useful). While I will occasionally make lecture notes available to students via a course GitHub repository (or on bCourses), 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 (the scribing assignments also help for coverage of missed lectures). 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 and/or bCourses. Although attendance is not mandatory, please be aware that Course Capture is not being used for this class. Students with a structural time conflict with class should not enroll in the class.

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 (2022). 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., extensions are for exceptional circumstances only). 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 honesty *very* seriously. Cheating, of any type, will not be tolerated. Please note that while the use of AI for homework is not prohibited, the disclosure of any such use is required. Non-disclosed use of AI will be considered a violation of academic integrity.

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. However, don't be shy about approaching me with questions immediately before/after class (I'll will try to arrive to class a few minutes early and linger afterwards). For longer questions please make use of my office hours. This is time specifically allocated for your use; please come by! I look forward to getting to know all of you.

Table 1: **Course Outline**

Date	**Topic**	**Readings**
Tu 1/20 Th 1/22	*Logic & Probability, Bayes' rule*	[b] Mitzenmacher & Upfal (2005, Ch. 1) [b] Hacking (2001, Chs. 1 - 7)
Tu 1/27 Th 1/29	*Bayes' rules: Beta- Binomial, Bandits*	[r] Russo et al. (2018)
Tu 2/3 Th 2/5	*MCMC & Ciphers/Code-Breaking*	[r] Lancaster (2004, Ch. 4) [b] Diaconis (2009), Connor (2003)
Tu 2/10 Th 2/12	*Law-of-Iterated Linear Predictors & Proxy Variables*	[r] Wooldridge (2010, Ch. 2 & 4.3) [b] Freedman (1997)
Tu 2/17 Th 2/19	*Bayesian Bootstrap*	[b] Chamberlain & Imbens (2003)
Tu 2/24 Th 2/26	*Model Selection*	[r] Efron (2004) [b] Efron & Hastie (2016, Ch. 12)
Tu 3/3 Th 3/5	*Regression Trees*	[b] Ye (1998, Ch. 8) [b] Efron & Hastie (2016, Ch. 8)
Tu 3/10 Th 3/12	*Productivity*	[r] Syverson (2011) [b] Brynjolfsson & Hitt (2003)
Tu 3/17 Th 3/19	*Transmission Bias* **Midterm 1**	[r] Nerlove (1963) [b] Griliches & Mairesse (1998)
Tu 3/24 Th 3/26	**Spring Recess**	
Tu 3/31 Th 4/2	*Quantile Regression*	[r] Koenker & Hallock (2001) [r] Mood et al. (1974)
Tu 4/7 Th 4/9	*Quantile Regression*	[r] Chamberlain (1994) [r] Fitzenberger & Wilke (2006)
Tu 4/14 Th 4/16	**Strike durations, Recidivism*	[r] Durose et al. (2014) Yang (2017)
Tu 4/21 Th 4/23	**Discrete Hazard Analysis*	[r] Singer & Willett (2003, Chs. 9 - 12) [b] Jenkins (1995), Efron (1988)
Tu 4/28 **Th 4/30**	*Review/Catch -Up* **Midterm 2**	

('' denotes a "tentative topic" which may change, [r] 'required', [b] 'background')*

References

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

Chamberlain, G. (1994). *Advances in Econometrics: Sixth World Congress*, volume 2, chapter Quantile regression, censoring, and the structure of wages, (pp. 171 – 209). Cambridge University Press: Cambridge.

Chamberlain, G. & Imbens, G. W. (2003). Nonparametric applications of bayesian inference. *Journal of Business and Economic Statistics*, 21(1), 12 – 18.

Connor, S. (2003). Simulation and solving substitution codes. Master's thesis, University of Warwick, Department of Statistics.

Diaconis, P. (2009). The markov chain monte carlo revolution. *Bulletin of the American Mathematical Society*, 46(2), 179 – 205.

Durose, M. R., Cooper, A. D., & Snyder, H. N. (2014). *Recidivism of Prisoners Released in 30 States in 2005: Patterns from 2005 to 2010*. Special Report NCJ 244205, U.S. Department of Justice, Office of Justice Programs, Bureau of Justice Statistics.

Efron, B. (1988). Logistic regression, survival analysis, and the kaplan-meier curve. *Journal of the American Statistical Association*, 83(402), 414 – 425.

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.

Fitzenberger, B. & Wilke, R. A. (2006). *Modern Econometric Analysis*, chapter Using quantile regression for duration analysis, (pp. 103 – 118). Springer-Verlag: Berlin.

Freedman, D. A. (1997). From association to causation via regression. *Advances in Applied Mathematics*, 18(1), 59 – 110.

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.

Hacking, I. (2001). *An introduction to probability and inductive logic*. Cambridge: Cambridge University Press.

Jenkins, S. P. (1995). Easy estimaton methods for discrete-time duration models. *Oxford Bulletin of Economics and Statistics*, 57(1), 129 – 137.

Koenker, R. & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4), 143 – 156.

Lancaster, T. (2004). *An Introduction to Modern Bayesian Econometrics*. Oxford: Blackwell Publishing.

McKinney, W. (2022). *Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter*. Cambridge: O'Reilly, 3rd edition.

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

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the Theory of Statistics*. New York: McGraw-Hill Book Company, 3rd edition.

Nerlove, M. (1963). *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld*, chapter Returns to scale in electricity supply, (pp. 167 – 198). Stanford University Press: Stanford.

Russo, D. J., Van Roy, B., Kazerouni, A., Osband, I., & Wen, Z. (2018). A tutorial on thompson sampling. *Foundations and Trends in Machine Learning*, 11(1), 1– 96.

Singer, J. D. & Willett, J. B. (2003). *Applied Longitudinal Data Analysis*. Oxford: Oxford University Press.

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.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2nd edition.

Yang, C. (2017). Local labor markets and criminal recidivism. *Journal of Public Economics*, 147(1), 16 – 29.

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.