Quantitative Empirical Methods (QEM) Exam Reading List (Last updated: June 2021)

Coursework

The primary QEM coursework is a two-course doctoral sequence:

- PLSC 500 (Foundations of Statistical Inference) and
- PLSC 503 (Theory and Practice of Quantitative Methods).

These courses, along with the exam and your future quantitative research, require mathematical maturity, and thus:

- PLSC 529 (Mathematics for Political Science)

is also a core course that serves as a corequisite for PLSC 500 and a prerequisite for PLSC 503. (PLSC 529 is also core for the Formal Theory subfield, for which the coursework is strongly recommended in conjunction with the QEM sequence.) Mastery of the materials covered in PLSC 500 and 503 are essential to passing the QEM exam. PLSC 508 (or equivalent advanced course on research design, e.g., PLSC 511 or S&DS 517) is strongly recommended for students intending to take the exam. Attendance at the MacMillan-CSAP Workshop on Quantitative Research Methods, in addition to at least one other subfield seminar series, is also strongly recommended.

Additional courses in mathematics, formal theory, political economy, probability, statistics, qualitative and archival methodology, and the empirical social sciences in the department and around the university can help you develop your skills. Coursework in the Qualitative and Archival Methodology subfield is recommended as a natural complement to the sequence that will help hone your empirical instincts. Comfort with political economy and formal theory will help you use economic reasoning when engaging with empirical research. More generally, the QEM field spans the breadth of political science, and you are expected to have some familiarity with major debates across the subfields. Your instructors and advisors will be able to help you find suitable courses depending on your interests.

Exam Format

A typical set of instructions for the exam are reproduced below. The instructions may change from exam to exam.

> *You have seven hours to complete the exam. This exam consists of three parts.*
>
> *Back up your assertions with mathematics where appropriate and show your work. Good answers will provide a direct answer that illustrates an understanding of the question, and calculations or statistical arguments to validate the answer. Where applicable, exceptional answers will include all of these as well as proofs that are technically complete, including formally articulating sufficient assumptions and regularity conditions. Questions will not be weighted equally. A holistic score will be assigned to the exam. Therefore, it is important to demonstrate your understanding of the material to the best of your ability.*

Part 1 (Short Answer Section) consists of seven short answer questions. Advice: Note there are multiple correct answers to some questions. We encourage you to give the most complete (but still succinct) solution possible. Do not leave sub-parts of questions unanswered.

Part 2 (Essay Section) contains a recent, well-regarded empirical article. We will ask you to offer an evaluation of its methodological approach and presentation of results. In particular, we will advise you to pay particular attention to the identification conditions (either explicit or implicit), the associated estimation strategy, and possible threats to inference. Your response may be anywhere from 500 to 1500 words.

The only aids permitted for Parts 1 and 2 are (i) one page of double-sided notes, (ii) a word processor on one of the Statlab computers to write up your answers (you may also write up your answers to Part 1 using pencil/pen and paper). After handing in your answers for Parts 1 and 2 of the exam, you may begin Part 3 (though feel free to look ahead). You may hand in Parts 1 and 2 whenever you wish, but we recommend spending no longer than five hours on Parts 1 and 2.

Part 3 (Computer Assisted Section) will involve using statistical software to answer one longer exercise with five associated questions. A complete answer to Part 3 will include code and output, as well as your written answers. Advice: We recommend that you explain what you are trying to do in comments in your code. Even if you are not able to execute your program correctly, you can receive partial credit for explaining clearly what you wanted to do and why.

For Part 3, you are permitted (i) unrestricted use of your own computer with access to the internet or (ii) use of a Statlab computer with access to the internet. The only restriction for Part 3 is that you may not interact with anyone, online or otherwise. For Part 3 (Computer Assisted Portion) of the exam, please turn in a hard copy of your code to Colleen, and also email a digital copy of the code to colleen.amaro@yale.edu.

Reading List

Each section is roughly ordered in increasing difficulty. Some entries are listed more than once across sections.

Probability:

- Freedman, David A. and Philip B. Stark. 2003. What is the chance of an earthquake? NATO Science Series IV: Earth and Environmental Sciences. 32: 201–213.
- Greenland, Sander, et al. Statistical tests, 2016. P-values, confidence intervals, and power: a guide to misinterpretations. Eur J Epidemiol 31, 337–350.
- Freedman, David A., Robert Pisani, and Roger A. Purves. 1998. Statistics, 3rd edn. New York, NY: Norton.
- Peter M. Aronow and Benjamin T. Miller. 2020. Foundations of Agnostic Statistics. Cambridge University Press. Ch. 1-2. (Note: the “Further Readings” sections at the end of each chapter of Aronow and Miller may often be helpful.)

- Wackerly, Dennis D., William Mendenhall III, and Richard L. Scheaffer. 2008. Mathematical statistics with applications, 7th edn. Belmont, CA: Thomson Brooks/Cole.
- Blitzstein, Joseph K. and Jessica Hwang. 2014. Introduction to probability. Boca Raton, FL: CRC Press.
- Wasserman, Larry. 2004. All of statistics: A concise course in statistical inference. New York, NY: Springer Science+Business Media, Inc. Chs. 1-5, 23.
- Hansen, Bruce E. 2021. Introduction to Econometrics. Unpublished manuscript, University of Wisconsin.

Data Science:

- James, Gareth et al. 2014. An introduction to statistical learning with applications in R. New York, NY: Springer Science+Business Media New York.
- Peng, Roger D., Caffo, Brian, and Leek, Jeff. R Programming. Coursera course at: https://www.coursera.org/learn/r-programming (Note: this is free, choose the "Audit" option.)
- Wasserman. All of statistics. Chapter 24.
- Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. 2009. The elements of statistical learning: Data mining, inference, and prediction. New York, NY: Springer-Verlag New York.
- Hansen, Bruce E. 2021. Econometrics. Unpublished manuscript, University of Wisconsin. Ch. 29.
- Wickham, Hadley. 2019. Advanced R. 2nd Edition. CRC press.

Core Statistics and Econometrics:

- Gelman, Andrew, and Hill, Jennifer. 2006. Data analysis using regression and multilevel/hierarchical models. Cambridge university press. Read with:
    - Efron, Bradley, and Morris, Carl. 1977. Stein's paradox in statistics. Scientific American, 236(5), 119-127.
- Aronow and Miller. Ch. 3-5.
- Wackerly, Mendenhall, and Scheaffer.
- Goldberger, Arthur S. 1991. A course in econometrics. Cambridge, MA: Harvard University Press.
- Chamberlain, Gary. 2010. Econometrics Lecture Notes. Available at: https://github.com/paulgp/GaryChamberlainLectureNotes
- Wasserman. All of statistics. Chs. 6-15, 19-22
- Hansen. Econometrics.
- Cameron, A. Colin and Pravin K. Trivedi. 2005. Microeconomics: Methods and applications. New York, NY: Cambridge University Press.
- Hayashi, Fumio. 2001. Econometrics. Princeton, NJ: Princeton University Press.

Research Design and Causal Inference:

- Freedman, David A. 1991. Statistical models and shoe leather. Sociological methodology, 291-313.
- Sekhon, Jasjeet S. 2009. Opiates for the Matches: Matching Methods for Causal Inference. Annual Review of Political Science 12: 487-508.
- Samii, Cyrus. 2016. Causal empiricism in quantitative research. Journal of Politics, 78(3), 941-955.

- Stuart, Elizabeth A. 2010. Matching Methods for Causal Inference: A Review and a Look Forward." Statistical Science, 25(1) 1-21.
- Broockman, David E., Joshua L. Kalla, and Jasjeet S. Sekhon. 2017. The design of field experiments with survey outcomes: A framework for selecting more efficient, robust, and ethical designs. Political Analysis 25.4: 435-464.
- Holland, Paul W. 1986. Statistics and causal inference. Journal of the American Statistical Association. 81(396): 945–968.
- Gerber, Alan S. and Green, Donald P., 2012. Field experiments: Design, analysis, and interpretation. WW Norton. Read with:
    - Särndal, Carl-Erik, Swensson, Bengt, & Wretman, Jan. 2003. Model assisted survey sampling. Springer Science & Business Media. Chs. 1-3.3.
    - Aronow, Peter M. and Cyrus Samii. 2017. Estimating Average Causal Effects under General Interference, with Application to a Social Network Experiment. Annals of Applied Statistics.
- Cattaneo, Matias D., Idrobo, Nicolas, and Titiunik, Rocio. 2020. A practical introduction to regression discontinuity designs: Foundations. Cambridge: Cambridge University Press.
- Sekhon, Jasjeet S., and Rocio Titiunik. 2012. When natural experiments are neither natural nor experiments. American Political Science Review. 35-57.
- Morgan, Stephen L. and Christopher Winship. 2014. Counterfactuals and causal inference. New York, NY: Cambridge University Press.
- Aronow and Miller. Ch. 6-7.
- Rosenbaum, Paul R. 1999. Choice as an alternative to control in observational studies. Statistical Science. 259-278. Read with discussion.
- Angrist, Joshua D., Imbens, Guido W., and Rubin, Donald B. 1996. Identification of causal effects using instrumental variables. Journal of the American Statistical Association, 91(434), 444-455. Read with discussion. Also read with:
    - Sovey, Allison and Donald Green. 2011. Instrumental Variables Estimation in Political Science: A Readers' Guide. American Journal of Political Science Vol. 55, No. 1. 188-200.
- Freedman, David A. 2008. On regression adjustments to experimental data. Advances in Applied Mathematics. 40: 180–193. Read with:
    - Freedman, David A. 2008. On regression adjustments in experiments with several treatments. The Annals of Applied Statistics. 2: 176–196.
    - Lin, Winston. 2013. Agnostic notes on regression adjustments to experimental data: Reexamining Freedman's critique. The Annals of Applied Statistics. 7(1): 295–318.
- Heckman, James J., and Sergio Urzua. 2009. Comparing IV with Structural Models: What Simple IV Can and Cannot Identify. National Bureau of Economic Research Working Paper 14706. Read with:
    - Imbens, Guido W. 2010. Better LATE Than Nothing: Some Comments on Deaton (2009) and Heckman and Urzua (2009). Journal of Economic Literature, 48 (2): 399-423.
- Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. Mostly harmless econometrics: An empiricist's companion. Princeton, NJ: Princeton University Press.
- Angrist, Joshua D. and Alan B. Krueger. 1999. Empirical strategies in labor economics. Handbook of Labor Economics. 3: 1277–1366.
- Wasserman. All of statistics. Chs. 15-18.

- Hernan, Miguel A. and Robins, James M. 2020. Causal Inference: What If. Boca Raton: Chapman & Hall/CRC.
- Hansen. Econometrics.