Introduction to Quantitative Methods II
Government 2000

Professor Jasjeet Singh Sekhon
Alexis Diamond, TF

Class: 4-6pm Thursdays
CBRSS Conference Room (34 Kirkland St.)

Associate Professor Jasjeet Singh Sekhon
jasjeet_sekhon@harvard.edu
HTTP://jsekhon.fas.harvard.edu/
Phone: 496-2426, Fax: 496-5149
Office: CBRSS, 34 Kirkland St., Room #4
Office hours: 2:30−3:30pm Wednesdays, or by appointment, or by email anytime

Alexis Diamond, Teaching Fellow
adiamond@fas.harvard.edu
Phone: TBA
Office: TBA
Office hours: TBA
Sections: TBA

Description

The course focuses on how to make causal inferences based on observational data with methods that make few assumptions and tools which allow one to test these assumptions. Topics include philosophy of science useful for causal inference, multivariate and propensity score matching, robust estimation, bootstrap and simulation methods, the general linear family which includes models such as logistic and Poisson regression, instrumental variable and Heckman selection models, and methods applicable for time-series/cross-section data. Prerequisites: Government 1000 or equivalent. Undergraduates are welcome.

Discussion

Government 2000 is the second course in the methods sequence for Government Department graduate students and is a prerequisite for Government 2001.

The primary focus of this course is to teach students enough statistical methods so they may conduct original quantitative empirical work on their own using modern methods and minimal assumptions. Of particlar interest in this course are methods of causal inference useful for working with observational data. Many researchers in the social science think they are making causal

inference when using (linear) regression methods, but few realize the assumptions required when using such limited methods. Alternative quantitative methods, such as propensity and multivariate matching, are discussed in detail, as are issues related to research design. The philosophical underpinnings of various theories of causal inference are also discussed.

We use an email list in this class quite frequently: `mailto:gov2000-list@fas.harvard.edu`. If students have a question about the course material, they are advised to email the question to the entire list and ask for help. Letting everyone see each other's questions and answers improves everyone's work.

Evaluation

Final grades will be based on a series of homework assignments (45% of final grade), a term paper (45%), and class and section participation (10%). There will be no final exam.

It is recommended that students write the term paper jointly with one or at most two other students. Experience has shown that this greatly facilitates learning as well as increases the likelihood that the paper will eventually become a published article.

Weekly readings and class assignments are the norm. Homework will generally be due the day after section. It is highly recommended that students form study groups in order to complete the homework assignments. Although it is recommended that people work together in order to complete the homework assignments, student must hand in their own individual answers. Photocopies and other reproductions of someone else's answers are not acceptable. Students should hand in the answers to the problem set, and all computer code written to find those answers.

Course Software and Books

The programming language for this course is the *R* variant of the *S* statistical programming language. It is installed Harvard-MIT Data Center computers and is available for download from: `http://www.r-project.org/`. *R* is open source software (released under the GNU public license) and is available at no charge. Students may with to use *Splus*, which is a commercial product and another variant of the *S* language, but the code we provide will be written to work with recent version of *R* and some of it will not work with *Splus*.

The three books listed below are required and available at various online bookstores and have been ordered at the campus store. Two of the three books are used in Government 1000 so many students will already own them. The core course material will be communicated in lectures and associated notes and handouts. The textbooks are important reference guides.

- Fox, John. 1997. *Applied Regression Analysis, Linear Models, and Related Methods.* Thousand Oaks, CA: Sage. Fox,
- Fox, John. 2002. *An R and S-PLUS Companion to Applied Regression.* Thousand Oaks, CA: Sage.
- Venables, W.N and Brian D. Ripley. 2002. *Modern Applied Statistics with S.* New York: Springer-Verlag. ISBN: 0387954570

Course Plan and Tentative Outline

1. Causal Inference in the Social Sciences (1 lecture)

   (a) Observational studies versus randomized experiments

   (b) Causal inference versus description

   (c) Methods of inductive inference

   (d) Epistemology vs. ontology

   Additional Readings:

   - Jasjeet S. Sekhon. "Quality Meets Quantity: Case Studies, Conditional Probability and Counterfactuals." *Perspectives on Politics*. June: 281-293. 2004.
   - Rubin, Donald. 1990. "Comment: Neyman (1923) and Causal Inference in Experiments and Observational Studies," *Statistical Science 5*, 472-480.
   - Reiter, Jerome. 2000. "Using Statistics to Determine Causal Relationships." *American Mathematical Monthly*. Pages 24–32.
   - Geddes, Barbara. 1990. "How the Cases You Choose Affect the Answers You Get: Selection Bias in Comparative Politics." *Political Analysis* 2:131–150.

2. Review of linear least squares and its assumptions (1 lecture)

   (a) Data Generating Process

   (b) Derivation and properties of OLS (BLUE)

   (c) OLS under asymptotic assumptions

   (d) Problems with OLS

   (e) Aliasing, error-in-variables, misspecification

   (f) Simple but limited fixes such as Huber-White standard errors

3. The Bootstrap, other Simulation Methods, and Bayesian Alternatives (1 lecture)

   (a) Standard errors and hypothesis tests

   (b) Quantities of interest

   (c) Answering questions without analytical formulas

   (d) Multivariate-t regression

4. Foundations for Robust Estimation (2 lectures)

   (a) Why least squares?

   (b) Influence curves

   (c) Types of robustness

   (d) High break-down point estimators

   (e) Fisher consistency

   (f) Measures of location and their influence functions

      - Quantiles

- Winsorized mean
- Trimmed mean
- M-Measures of location
- Least median of squares
- Least trimmed squares
- S-Measures of location
- MM-estimator

(g) Measures of scale

- Median absolute deviation (MAD)
- Biweight midvariance
- Percentage bend midvariance
- Least quartile difference (LQD)

Additional Readings:

- Western, Bruce. 1995. "Concepts and Suggestions for Robust Regression Analysis." *American Journal of Political Science* 39: 786–817.
- Walter R. Mebane, Jr. and Jasjeet S. Sekhon. forthcoming. "Robust Estimation and Outlier Detection for Overdispersed Multinomial Models of Count Data." *American Journal of Political Science*.

5. Problems related to Time Series Cross Section Data (1 lecture)

   (a) Spurious Regressions
   (b) Distributed Lags
   (c) Serial correlation
   (d) ARMA processes
   (e) Tests of serial correlation

6. Extensions to Generalized Linear Models (logistic regression and friends) (2 lectures)

   (a) The unification of the linear model
   (b) Exponential dispersion models
   (c) The three components of GLM
      - Response distribution or "error structure"
      - Linear predictor
      - Link function
   (d) Example models
      - Models of proportions (logit, probit)
      - Models of counts (Poisson)
      - Multinomial choice models (MNL, MNP)
   (e) Quasi-likelihood

   Additional Readings:

- Jonathan N. Wand, Kenneth W. Shotts, Jasjeet S. Sekhon, Walter R. Mebane, Jr., Michael C. Herron and Henry E. Brady. 2001. "The Butterfly Did It: The Aberrant Vote for Buchanan in Palm Beach County, Florida." *American Political Science Review* 95 (4): 793–810.

7. Experimental Manipulation (1 lecture)
   Additional Readings:

   - Holland, Paul. 1986. "Statistics and Causal Inference (with comments by Rubin, Clark Glymour, Clive Granger)." *Journal of the American Statistical Association* 81(396): 945–970
   - Gerber, Alan S. and Donald P. Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *American Political Science Review* 94(3): 653–663.
   - Imai, Kosuke. "Do Get-Out-The-Vote Calls Reduce Turnout? The Importance of Statistical Methods for Field Experiments." *American Political Science Review*. Forthcoming.
   - Rubin, Donald. "The Design of the New York School Choice Scholarship Program Evaluation." in *Research Designs: Inspired by the work of Donald Campbell*, L. Bickman, eds., Thousand Oaks, CA: Sage. Ch 7, 155–180.

8. Multivariate Matching and Propensity Scores (1 lecture)
   Additional Readings:

   - Rosenbaum, Paul R. and Rubin, Donald B. (1985) "Constructing a control group using multivariate matched sampling methods that incorporate the propensity score" *American Statistician* 39: 33–38.
   - We revisit Gerber and Green (2000) and Imai.

9. Instrumental Variables and Heckman Selection Models (1 lecture)
   Additional Readings:

   - Joshua D. Angrist and Alan Krueger. Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments. *Journal of Economic Perspectives* 14:4 69–85.
   - Guido Imbens, Joshua Angrist, and Donald Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of Econometrics* 71(1-2), 145–160.
   - Steven D. Levitt and James M. Snyder, Jr. 1997. "The Impact of Federal Spending on House Election Outcomes." *Journal of Political Economy* 105(1): 30–53.

10. Other approaches to causal inference: Diffs-in-Diffs and Pseudo-Control units (1 lecture)

    - David E. Card and Alan B. Krueger. October 1993. "Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania." *NBER Working Paper #W4509*
    - Alberto Abadie and Javier Gardeazabel. 2003. "The Economic Costs of Conflict: A Case-Control Study for the Basque Country." *American Economic Review* 93(1), 113–132.

Supplementary Reading Material

The following books are certainly not required, but they may be of interest during the course. It is often very useful to read the same material covered by a variety of authors. Within each section, books are approximately ordered by increasing sophistication.

Causal Inference

- Rosenbaum, Paul R. 2002 (2nd ed). *Observational Studies.* New York: Springer-Verlag.

Bootstrap Methods

- Efron, Bradley and Tibshirani, Robert J. 1994. *An Introduction to the Bootstrap.* Chapman & Hall. ISBN: 0412042312.
- Hall, Peter. 1992. *The Bootstrap and Edgeworth Expansion.* New York: Springer-Verlag.

Computer Books

- Krause, Andreas and Melvin Olson. 2002. *The Basics of S-PLUS.* 3rd ed. New York: Springer-Verlag. ISBN: 0387954562.
- Spector, Phil. 1995. *An Introduction to S and S-Plus.* Wadsworth Publishing Company. ISBN: 053419866X.

Generalized Linear Models

- Dobson, Annette J. 2001 (2nd ed). *An Introduction to Generalized Linear Models.* Chapman and Hall. ISBN: 1584881658.
- Lindsey, James K. 1997. *Applying Generalized Linear Models.* New York: Springer-Verlag. ISBN: 0387982183.
- McCullagh, Peter and J. A. Nelder. 1989 (2nd edition). *Generalized Linear Models.* Chapman-Hall. ISBN: 0412317605.

Econometrics

- Greene, William H. 1990. *Econometric Analysis.* New York MacMillan Publishing Company.
- Maddala, G. S. 1994. *Limited-Dependent and Qualitative Variables in Econometrics.* Cambridge University Press. ISBN: 0521338255.
- Davidson, Russell and James G. MacKinnon. 1993. *Estimation and Inference in Econometrics.* 1st ed. New York, NY: Oxford University Press. ISBN: 0195060113.
- White, Halbert. 1996. *Estimation, Inference and Specification Analysis.* Cambridge University Press. ISBN: 0521574463.

Matrix Algebra Review Books

- Namboodiri Krishnan. 1986. *Matrix Algebra: An Introduction*. SAGE Publications. ISBN: 0803920520.
- Searle, Shayle R. 1982. *Matrix Algebra Useful for Statistics*. Wiley-Interscience. ISBN: 0471866814.

Regression Diagnostics

- Belsley, David A., Edwin Kuh, and Roy E. Welsch. 2000. *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. Wiley, John and Sons. ISBN: 0471058564.
- Tukey, John Wilder 1977. *Exploratory Data Analysis*. Addison-Wesley: Reading MA.
- Mosteller, Frederick and John Wilder Tukey. 1977. *Data Analysis and Regression*. Addison-Wesley: Reading, Ma.

Robust Estimation

- Cook, R. Dennis and S. Weisberg. 1982. *Residuals and Influence in Regression*. Chapman and Hall. ISBN: 041224280X.
- Hoaglin, David C., Frederick Mosteller, and John W. Tukey (editors). 2000. *Understanding Robust and Exploratory Data Analysis*. John Wiley and Sons. ISBN: 0471384917.
- Carroll, Raymond J. and David Ruppert. 1988. *Transformation and Weighting in Regression*. CRC Press. ISBN: 0412014211.
- Staudte, Robert G. and Simon J. Sheather. 1990. *Robust Estimation and Testing*. Wiley, John and Sons. ISBN: 0471855472.
- Huber, Peter J. 1981. *Robust Statistics*. John Wiley and Sons. ISBN: 0471418056.
- Hampel, Frank R., Elvezio M. Ronchetti, Peter J. Rousseeuw, and Werner A. Stahel. 1986. *Robust Statistics: The Approach Based on Influence Functions*. Wiley, John and Sons. ISBN: 0471829218.

Probability and Statistics

- DeGroot, Morris H. 1986. *Probability and Statistics*. Addison-Wesley.
- Hogg, R. V. and A. T Craig. 1994 (5th ed). *Introduction to Mathematical Statistics*. Simon and Schuster. ISBN: 0023557222
- Rice, John A. 1995. *Mathematical Statistics and Data Analysis*, 2nd Ed. Belmont, CA: Duxbury Press.
- Casella, G. and R. L. Berger. 1990. *Statistical Inference*. Duxbury Press: Belmont, CA.
- Cox, D. R. and Hinckley, D. V. 1979. *Theoretical Statistics*.
- Lehmann, E. L. 1998 (REV with Casella). *Theory of Point Estimation*. Springer-Verlag. ISBN: 0387985026.

- Lehmann, E. L. 1997 (2nd REPRIN). *Testing Statistical Hypotheses*. Springer-Verlag. ISBN: 0387949194.

Time Series

- Mills, Terence C. 1990. *Time Series Techniques for Economists*. New York: Cambridge University Press.
- Hamilton, James D. 1994. *Time Series Analysis*. Princeton University Press.