Boston University School of Management
Fusing the Art, Science, and Technology of Business.®

SI915 Course Outline

Models & Methods for Causal Inference in Strategy Research (*Spring 2012)*
Course Meets: BU School of Management

Instructor: Timothy Simcoe
E-Mail: tsimcoe@bu.edu
Homepage: http://people.bu.edu/tsimcoe
Phone (Fax):   617.358.5725
Office Hours: By Appointment, or just knock on my door

Objectives

What should you do when a referee says your paper has an endogeneity problem or a seminar participant asks how your estimates are identified? This course introduces a "toolkit" for attempting to estimate causal relationships using non-experimental data. We will discuss how to establish what relationships exist in the data, when you can interpret these relationships as causal, and how you can convince your audience of your results (without overselling). We will also think carefully about the interaction between large-sample empirical research, qualitative institutional data, and theory, especially the importance of careful theoretical thinking (in the context of the institutional details) for empirical research.

There are two main goals. The first is to develop a clear understanding of the conceptual difficulties associated with establishing causality in empirical work. In particular, I hope to de-mystify several concepts that are frequently invoked as "problems" or "issues" in applied empirical research, but seldom clearly explained. A good grasp of these concepts will lead to improved research design and a sharper understanding of the relative strengths and weaknesses of various statistical methods. The second goal is to describe some of the practical problems that arise in the application of these methods. I will place a special emphasis on testing the key underlying assumptions. The ultimate goal is for you leave prepared to undertake your own empirical research.

Preparation and Prerequisites

This course is designed to complement a graduate sequence in econometrics , but it should be accessible to students with basic knowledge of probability and statistics. We will emphasize intuition and application over proof. However, the readings and class discussion will cover technical material. Problem sets and in-class examples will be taught using the Stata statistical software package

Assignments & Grading

Grades will be based on class participation (10%), a problem set (30%), a replication exercise (30%), and a research paper/presentation (30%).

Class participation (10%): Read the materials; come to class; participate.

Problem Set (30%): I will provide a set of empirical exercise and data that will allow you to practice the methods we learn in class within a fairly controlled environment.

Replication Exercise (30%): Choose an existing empirical paper to replicate, and discuss/critique the robustness of the results using concepts from class. Several journals (e.g. American Economic Review, American Economic Journals, Journal of Applied Econometrics, Journal of Business Economics and Statistics) and researchers (Bronwyn Hall, Bruce Hansen, Justin Wolfers, Dan Trefler, myself) post data from published papers. While there is no page limit, my guess is that the reports will contain 1-4 pages of text, 2-5 tables and/or figures, and a Stata .log output file. One to three tables would likely replicate results from the existing paper and one or two more would perhaps show results that are not shown in the paper. If you can communicate the core ideas in less space, no problem. If you need more space, that's fine too. The key is to show that you could replicate the main results and that you tried some other specifications (informed by what we do in class) to check robustness. Please confirm with me that your chosen paper is appropriate before you work on it, and don't hesitate to ask questions at any point in the process!

Research Paper / Presentation (30%): At the end of the semester, you should submit a written "research design." This 4 to 6 page document will describe how you plan to implement an empirical study. Your research design should read like the "Data and Methods" section of a high-quality empirical paper. I expect to see a description of your data, a specification for the regressions you will perform, and (most importantly) discussion and justification of the assumptions that your reader must maintain in order to believe that your analysis constitutes an answer to the research question.

For this assignment, I strongly recommend that you choose a question you are actually working on. Preferably, you have the data in hand. An alternative is to use a set of hypotheses developed as part of a previous class assignment. In either case, you should submit a 1 or 2 page summary of the theory / hypotheses along with your research design, for a total of 5 to 8 pages.

Class Calendar

	Date	Topic	Assignments Due
1	Jan 25	What is identification?	
2	Feb 1	Field experiments	
3	Feb 8	Selection on observables	
4	Feb 15	Matching etiquette	
5	Feb 22	Instrumental variables	
6	Feb 29	IV Etiquette	
7	Mar 7	Panels and fixed effects	Problem Set
		SPRING BREAK	
8	Mar 21	Diff-in-diffs etiquette	
9	Mar 28	Regression discontinuity	
10	Apr 4	Non-standard data problems	
11	Apr 11	Classic identification problems	Replication
12	Apr 18	Using theory for identification	
13	Apr 25	Quality in empirical research	
14	May 2	Student presentations	Paper & Prest'n

Assigned Reading

Each class will have several assigned readings. There are three types of reading:

1) Conceptual readings deal with tools and methods for causal inference. Some conceptual readings are academic papers, but most are chapters from Mostly Harmless Econometrics: An Empiricist's Companion, by Joshua Angrist and Steve Pischke. This is an excellent handbook for applied empirical research, and I highly recommend that you purchase a copy. Since MHE is not a complete reference, you may also wish to get a copy of William Greene's Econometric Analysis, Jeffrey Wooldridge's Econometric Analysis of Cross Section and Panel Data, or Cameron and Trivedi's Microeconometrics.

2) Applied Readings are research papers that use the tools and methods we learn in class. For each Applied Reading, you should arrive in class ready to answer to the following questions: What is the research question? What is the unit of observation (in theory/ideal and in the data)? What are the sources of (exogenous and endogenous) variation? What are the key estimating equation(s)? What are the results and interpretation? In some cases, you will have access to the underlying data used in an Applied Reading, and I strongly encourage you to play around with it.

3) Optional readings provide additional detail on topics related to those covered in class.

C = Required conceptual reading; A = Required application reading

Session 1: What is Identification?

C: Mostly Harmless Econometrics, Chapters 1 & 2.

C: Freedman, D. (1991), "Statistical Models and Shoe Leather," *Sociological Methodology*, 21, 291-313.

C: Manski, C. (1995), Identification Problems in the Social Sciences, Introduction and Chapter 1.

A: Kruger, A. (1993), "How Computers Have Changed the Wage Structure: Evidence from Micro Data" *Quarterly Journal of Economics*, 108, 33-60.

A: DiNardo, J, and Pischke J. (1997), "The Returns to Computer Use Revisited: Have Pencils changed the Wage Structure Too?" *Quarterly Journal of Economics*, 112, 291-303.

Session2: Randomization and Field Experiments

C: Mostly Harmless Econometrics, Chapter 3.

C: Duflo, E., Glennerster and Kremer (2006) "Using Randomization in Development Economics Research: A Toolkit" NBER Technical Working Paper No. 333

A: Bertrand, M. and S. Mullainathan (2003), "Are Emily and Greg More Employable than Lakisha and Jamal?: A Field Experiment on Labor Market Discrimination" NBER Working Paper No. 9873.

A: Bloom, N., B. Eifert, A. Mahajan, D. McKenzie and J. Roberts (2011), "Does Management Matter?:

Evidence From India" NBER Working Paper No. 16658

Blank, R. (1991), "The Effects of Double-Blind versus Single-Blind Reviewing: Experimental Evidence from the American Economic Review," *American Economic Review*, 81, pp. 1041-67.

Manski, C. (1999), Identification Problems in the Social Sciences, Chapter 2.

Boudreau, K. and K. Lakhani. (2011), "Field Experimental Evidence on Sorting of Creative Workers and Performance" (http://www.kevinboudreau.com/papers)

Session 3: Selection on Observables

C: Heckman (1979) "Sample Selection Bias as a Specification Error," Econometrica, 47, 153-161.

A: Shaver, Myles. 1998. "Accounting for Endogeneity in Strategy Performance: Does Entry Mode Affect FDI Survival?" *Management Science*.

C: Rosenbaum, P. and D. Rubin (1983): "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70, 41-55.

A: Sacerdote, B. (2005) "Fixing Broken Experiments Using the Propensity Score" in Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives, Gelman and Line, eds., John Wiley & Sons.

C/A: LaLonde, R. (1986), "Evaluating the econometric evaluations of training programs with experimental data" *American Economic Review*, 76, 604-620.

A: Jaffe, A., M. Trajtenberg and R. Henderson (1993), "Geographic Knowledge Spillovers as Evidenced by Patent Citations" *Quarterly Journal of Economics*, 108(3):577-98.

Session 4: Matching Etiquette

C: Todd, P. (2006), "Matching Estimators" (athena.sas.upenn.edu/~petra/papers/mpalgrave2.pdf)

C: Iacus, S., G. King, and G. Porro (2011). "Causal Inference Without Balance Checking: Coarsened Exact Matching." Political Analysis.

C: Smith, J and P. Todd (2001). "Reconciling Conflicting Evidence on the Performance of Propensity-Score Matching Methods," *American Economic Review*, 91 (2): 112-118.

A: Thompson, P., and M. Fox-Kean (2005): "Patent citations and the geography of knowledge spillovers: a reassessment" *American Economic Review*, 95(1): 450-460.

Dehejia, Rajeev and Sadek Wahba. 1999. "Causal Effects in Nonexperimental Studies: Reevaluating the Evaluation of Training Programs," *Journal of the American Statistical Association.*

Imbens, G. (2004), "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review" *Review of Economics and Statistics*, 86, 4-29.

Session 5: Instrumental Variables

C: Mostly Harmless Econometrics, Chapter 4.

C: Angrist and Kruger (2001) "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments," Journal of Economic Perspectives, 15, 69-85.

A: Acemoglu, D., S. Johnson and J. Robinson (2001) "The Colonial Origins of Comparative Development: An Empirical Investigation" American Economic Review, 5, 1369-1401.

A: Albouy, D. (2004) "The Colonial Origins of Comparative Development: An Empirical Investigation: Comment" American Economic Review, forthcoming. (http://www-personal.umich.edu/~albouy/)

James H. Stock & Francesco Trebbi (2003), "Who Invented Instrumental Variable Regression?," *Journal of Economic Perspectives*, 17(3), pages 177-194.

Session 6: IV Etiquette

C: Murray, Michael (2006). "Avoiding Invalid Instruments and Coping with Weak Instruments," Journal of Economic Perspectives.

C: Imbens, G. and J. Angrist (1994). "Identification and Estimation of Local Average Treatment Effects" Econometrica, 62:2, pp. 467-475.

C: Heckman, J., S. Urzua and E. Vytlacil (2006) "Understanding Instrumental Variables in Models with Essential Heterogeneity," Review of Economics and Statistics.

A: Galasso, A., M. Schankerman and C. Serrano (2011). "Trading and Enforcing Patent Rights", NBER Working Paper No. 17367

Hoxby, C. (2000) "Does Competition Among Public Schools Benefit Students and Taxpayers?" *American Economic Review*, 90, 1209-1238.

Rothstein, J. (2007) "Does Competition Among Public Schools Benefit Students and Taxpayers? A Comment on Hoxby (2000)" *American Economic Review*, 97, 2026-2038.

Hoxby , C. (2005) "Competition Among Public Schools: A Reply to Rothstein (2004)" NBER Working Paper No. 11216.

Session 7: Panel Data and Fixed Effects

C: Mostly Harmless Econometrics, Chapters 5 and 8.

C: Heckman, J. And J. Smith (1999) "The pre-programme earnings dip and the determinants of participation in a social programme: Implications for simple programme evaluation strategies" Economic Journal, 109(457), 313-348.

C: Bertrand, M., E. Duflo and S. Mullainathan (2004), "How Much Should We Trust Differences-in Differences Estimates?" *Quarterly Journal of Economics*, 119, 249-76.

A: Mas, Alex "Labour Unrest and the Quality of Production: Evidence from the Construction Equipment Resale Market" Review of Economic Studies (2008) vol. 1 pp. 1-30.

Griliches, Z. and J. Mairesse (1995) "Production Functions: The Search for Identification" NBER Working Paper 5067.

Hausman, Jerry A., and William E. Taylor. 1981. Panel Data and Unobservable Individual Effects. Econometrica 49(6), 1377-1398.

Session 8: Difference-in-Differences Etiquette

A: Agrawal, A., and A. Goldfarb (2008). "Restructuring Research: Communication Costs and the Democratization of University Innovation." *American Economic Review*, 98(4): 1578–90.

A: Azoulay, P., J. Graff-Zivin and J. Wang (2010). "Superstar Extinction." *The Quarterly Journal of Economics, 125 (2): 549-589.*

C: Donald, S. and K. Lang (2007), "Inference with Difference in Differences and Other Panel Data" *Review of Economics and Statistics*, 2, 221-233.

Abadie, A. (2005) "Semiparametric Difference-in-Differences Estimators" *Review of Economic Studies*, 72, 1-19.

Athey, S. and G. Imbens (2006). "Identification and Inference in Nonlinear Difference-In-Differences Models," Econometrica, 74(2): 431-497.

Session 9: Regression Discontinuity

C: Mostly Harmless Econometrics, Chapter 6.

C: Imbens, G. and T. Lemieux (2007) "Regression Discontinuity Designs: A Guide to Practice" NBER Technical Working Paper No. 337.

A: Lee, D.S., E. Moretti and M. Butler (2004) "Do Voters Affect Or Elect Policies? Evidence from the U. S. House" *Quarterly Journal of Economics*, 119, 807-859.

A: Oreopoulos, P. (2006) "Estimating Average and Local Average Treatment Effects of Education when Compulsory Schooling Laws Really Matter." American Economic Review.

Lee, D., and T. Lemieux (2010). "Regression Discontinuity Designs in Economics." Journal of Economic Literature, 48(2): 281–355.

Session 10: Data problems

Count Data

Wooldridge, J. (1999) "Distribution free estimation of some non-linear panel data models" Journal of Econometrics, 77-97.

Mullahy, J. (1997), "Instrumental-Variable Estimation of Count Data Models: Applications to Models of Cigarette Smoking Behaviour," Review of Economics and Statistics, 11, 586-593.

Santos Silva, J. M. C. and Silvana Tenreyro. 2006. "The Log of Gravity." Review of Economics and Statistics 88(4): 641-658.

Endogenous Dummies

Angrist, J. (1999) "Estimation of Limited Dependent Variable Models with Dummy Endogenous Regressors: Simple Strategies for Empirical Practice" Journal of Business and Economic Statistics, 19(1), 2-16.

Non-linear interactions

Ai, Chunrong, and Edward C. Norton. 2003. "Interaction Terms in Logit and Probit." Economics Letters 80, 123-129.

Puhani, Patrick A. 2008. "The Treatment Effect, the Cross Difference, and the Interaction Term in Nonlinear "Difference-in-Differences" Models." IZA Discussion Paper # 3478.

Age-Year-Cohort Problems

Mehta, A., M. Rysman and T. Simcoe (2007) "Identifying the Age Profile of Patent Citations: New Estimates of Knowledge Diffusion" manuscript.

Bronwyn H. Hall & Jacques Mairesse & Laure Turner, 2005. "Identifying Age, Cohort and Period Effects in Scientific Research Productivity: Discussion and Illustration Using Simulated and Actual Data on French Physicists," NBER Working Papers 11739

Duration Data

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

Van den Berg, G. J. (2005). Competing Risks Models. Working Paper.

Session 11: Classic identification problems

Heterogeneity vs. State-dependence

Heckman, J. (1991), "Identifying the Hand of Past: Distinguishing State Dependence from Heterogeneity" American Economic Review, 81(2), 75-99.

Simultaneity

Manski, C. (1999), Identification Problems in the Social Sciences: Chapter 6.

Angrist, J., K. Graddy and G. Imbens (2000) "The Interpretation of Instrumental Variables Estimators in Simultaneous Equations Models with an Application to the Demand for Fish" Review of Economic Studies, 67, 499-527.

Identifying Clusters

Rysman, Marc, and Shane Greenstein. 2005. "Testing for agglomeration and dispersion." Economics Letters 86: 405-411.

Ellison, Glenn D., Edward L. Glaeser, and William R. Kerr. Forthcoming. What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. American Economic Review.

The Reflection Problem

Manski, C. (1999), Identification Problems in the Social Sciences: Chapter 7. OR Manski, C. (1993) "Identification of Endogenous Social Effects: The Reflection Problem," Review of Economic Studies, 60(3), 531-542.

Waldinger, Fabian. 2009. Peer Effects in Science: Evidence from the Dismissal of Scientists in Nazi Germany. Working Paper, University of Warwick.

Sacerdote, Bruce. 2001. Peer Effects with Random Assignment: Results for Dartmouth Roommates. Quarterly Journal of Economics 116(2), 681-704.

Tucker, Catherine. 2008. "Identifying Formal and Informal Influence in Technology Adoption with Network Externalities", Management Science 55(12): 2024-2039.

Session 12: Structural Models – Using theory for identification

Reiss, Peter C., and Frank A. Wolak. 2007. "Structural Econometric Modeling: Rationales and Examples from Industrial Organization." in Handbook of Econometrics, Vol. 6A. Eds. James Heckman and Edward Leamer. Chapter 64, pp. 4277-4415.

Olley, Steve and Pakes, A (1996) "The dynamics of productivity in the telecommunications equipment industry." Econometrica".

Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2006. "Structural Identification of Production Functions." Working paper, University of Toronto.

Bresnahan, Timothy F., and Peter C. Reiss. 1991. "Entry and Competition in Concentrated Markets." Journal of Political Economy 99(5), 977-1009.

Goettler, Ronald, and Brett Gordon. 2009. Does AMD spur Intel to innovate more? Working paper.

Session 13 : Views on Quality in Empirical Research

Heckman, J. (2000), "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective" *Quarterly Journal of Economics*, 115, 45-97.

Journal of Economic Perspectives (Spring 2010). Symposium: Con Out of Economics

Journal of Economic Literature (June 2010). Forum on the Estimation of Treatment Effects

Section I of the Strategy Research Initiative "Strategy Reader" available at (http://strategyresearch.net/strategy_reader)