Georgetown University
Spring 2016
Advanced Applied Econometrics
(PPOL 754-20)

Andrew H. McCallum, Ph.D.

March 3, 2016

Disclaimer

Any opinions and conclusions expressed during this course are solely the responsibility of the author and do not necessarily represent the views of the Board of Governors or any other person associated with the Federal Reserve System.

About me

- International Finance Division at the Federal Reserve Board
- Study firm level international trade
- Finished my Ph.D. in economics in 2013 from the University of Michigan
- Taught 11 courses at Michigan and Georgetown
- Dissertation title: "The Structure and Evolution of Entry Costs in Trade"
- Four Ph.D. econometrics courses, two in advanced theory
- Three Ph.D. statistics courses, two in advanced theory
- Fellowship somewhat like a Fulbright for a year at the European Central Bank before graduate school.
- Research assistant at the San Francisco Fed for two years before Germany
- Undergraduate studies in economics and mathematics at the University of Washington and the London School of Economics
- From a small farming town in rural Washington State
- Raised sheep, chickens, and camped all the time
- I enjoy travel and teaching classes in my spare time

Georgetown University
Spring 2016
Advanced Applied Econometrics (PPOL 754-20)

One half semester module
March 3, 17, 31, April 7, 14, 21, 28
Lecture in Healy 103
Thursdays 6:30pm - 9:00pm

Andrew H. McCallum, Ph.D.
Economist
Board of Governors of the Federal Reserve
Constitution Ave. N.W.
Washington DC, 20551
andrew[dot]mccallum[at]georgetown[dot]edu
Please include "PPOL 754-20" in the email subject line.
Grader: Chenxi Lu cl962[at]georgetown[dot]edu

1 Course Description

This module has two purposes. The first is to delve deeper into some topics from introductory econometrics in order to gain a fuller understanding of how various estimators work and what they are actually estimating, including under cases of misspecification. The second is to learn about some models and estimators that are not typically studied in an introductory econometrics course, including quantile regression and the regression discontinuity design. Throughout the module we will illustrate concepts with actual empirical research on policy-relevant topics. Some of the materials in this course are derived from materials by Joshua Angrist (MIT), Peter Hinrichs (FRB-Cleveland), William Lincoln (JHU-SAIS), Guy Michaels (LSE), and Jeffrey Smith (Michigan).

2 Require Materials

1. Angrist, J.D. and J. Pischke, 2009. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton, NJ: Princeton Press.
2. Wooldridge, J. M., 2012. *Introductory Econometrics.* Cambridge, MA: MIT Press. (You can use any edition).
3. All additional papers and required data sets will be available on the course's Blackboard site.

3 Course Policy

1. Please do not allow mobile phones to ring in class.
2. Academic dishonesty will not be tolerated. You may work with others on the homework but need to list your collaborators and submit your own copy. The final project must be executed on your own. Submitting code you find on the Internet as your own work is plagiarism.
3. Respect the views and learning needs of your fellow students.
4. Bring any problems with the course to my attention.

4 Learning goals

After taking this course you will be able to:

1. Explain that not all social science research is of equal quality and describe what types of "identification strategies" (i.e., ways of going about trying to estimate causal effects of policies and programs) are better than others.

2. Describe what ordinary least squares (OLS) and instrumental variables (IV) estimators are actually estimating and how they are estimating it.

3. Implement and also understand research using the following: fixed effects, differences-in-differences, triple differences, the regression discontinuity design, quantile regression, and clustered standard errors.

4. Derive the OLS and IV regression estimators using matrix algebra.

5 Evaluation

Each of the following items will contribute toward the final grade. Once all points have been tallied, I will apply the McCourt grade distribution policy.

1. (60%) Homework problems assigned at the end of each class and that are due at the beginning of the next class.

2. (40%) Take-home final project which will use the tools from the rest of the course to analyze a new question.

6 Schedule

6.1 Research Design, Randomized Controlled Experiments, and Terminology and Notation from Probability and Statistics (all of 3/3)

1. Chs. 1 and 2 of *Mostly Harmless Econometrics*

2. Krueger, A. B., (1999) "Experimental Estimates of Education Production Functions," *The Quarterly Journal of Economics*, 127:3, pp. 1057-1106.

6.2 Ordinary Least Squares: What It's Actually Estimating and Why You Should Use It and a Brief Introduction to Matching Estimators (all of 3/17 and first-half of 3/31)

1. Ch. 3 of *Mostly Harmless Econometrics*.

2. Dehejia, R. and S. Wahba. 1999. "Causal effects in nonexperimental studies: reevaluating the evaluation of training programs," *Journal of the American Statistical Association*, 94:448, pp. 1053-1062.

6.3 Instrumental Variables Estimation: What Is Being Estimated and How to Use It (second-half of 3/31 and all of 4/7)

1. Ch. 4 of *Mostly Harmless Econometrics*.

2. Angrist, J. D. and A. B. Krueger. 1991. "Does Compulsory School Attendance Affect Schooling and Earnings?" *The Quarterly Journal of Economics*, 106:4, pp. 979-1014.

6.4 Fixed Effects, Differences-in-Differences, and Triple Differences (first-half of 4/14)

1. Ch. 5 of *Mostly Harmless Econometrics*.

2. Card, D. 1992. "Using Regional Variation in Wages to Measure the Effects of the Federal Minimum Wage," *Industrial and Labor Relations Review*, 46:1, pp. 22-37.

3. Card, D. and A. B. Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania," *The American Economic Review*, 84:4, pp. 772-793.

4. Besley, T. and R. Burgess. 2004. “Can Labor Regulation Hinder Economic Performance? Evidence from India,” *The Quarterly Journal of Economics*, 119:1, pp. 91-134.

6.5 The Regression Discontinuity Design (second-half of 4/14)

1. Ch. 6 of *Mostly Harmless Econometrics*.
2. Lee, D. S. 2008. “Randomized experiments from non-random selection in U.S. House elections,” *Journal of Econometrics*, 142:2, pp. 675-697.
3. Angrist, J. D. and V. Lavy. 1999. “Using Maimonides’ Rule to Estimate the Effect of Class Size on Scholastic Achievement,” *The Quarterly Journal of Economics*, 114:2, pp. 533-575.

6.6 Quantile Regression (first-half of 4/21)

1. Ch. 7 of *Mostly Harmless Econometrics*.
2. Angrist, J., V. Chernozhukov, and I. Fernández-Val. 2006. “Quantile Regression under Misspecification, with an Application to the U.S. Wage Structure,” *Econometrica*, 74:2, pp. 539-563.
3. Abadie, A., J. Angrist, and G. Imbens, 2002. “Instrumental Variables Estimates of the Effect of Subsidized Training on the Quantiles of Trainee Earnings.” *Econometrica*, 70:1, pp. 91-117.

6.7 Clustered Standard Errors (second-half of 4/21)

1. Ch. 8 of *Mostly Harmless Econometrics*.
2. Bertrand, M., E. Duflo, and S. Mullainathan. 2004. “How Much Should We Trust Differences-in-Differences Estimates?” *The Quarterly Journal of Economics*, 119:1, pp. 249-275.

6.8 Econometrics with Matrix Algebra (all of 4/28)

1. Appendixes D and E of Jeffrey M. Wooldridge’s *Introductory Econometrics*.

6.9 Assignment schedule

Homework submissions should include the *.do file, the *.log file created by that *.do file and a lucid written document addressing each section of the homework questions. They should be emailed to the grader with me cc’d and include “PPOL 754-20 Homework X” in the subject line before the start of class at 6:30pm. Your lowest score will be dropped so only the top 5 scoring homeworks will count towards your grade.

The final project will be similar to the homeworks but longer and more difficult. It will cover all of the main topics in the course. You must execute the final project on your own and without the help of your classmates.

1. Homework 1 assigned on 3/3 and due on 3/17.
2. Homework 2 assigned on 3/17 and due on 3/31.
3. Homework 3 assigned on 3/31 and due on 4/7.
4. Homework 4 assigned on 4/7 and due on 4/14.
5. Homework 5 assigned on 4/14 and due on 4/21.
6. Homework 6 assigned on 4/21 and due on 4/28.
7. Final project assigned on 4/21 and due on 5/10 at 5pm ET.

Grades

I calculated your numerical score for the course according to the syllabus. With your numerical percentage in hand, I then round up every student to the nearest full percentage point.

- A 100-96
- A- 95-91
- B+ 90-86
- B 85-81
- B- 80-75

Distribution last year:

- 3 students received As, 11 had A-, 9 had B+, 4 had B, and 2 had B-.
- No one failed the course.
- The average numerical score was 89% and earned a B+.
- A grade of B+ is one you can fairly represent as having a good grasp of the material.
- Students that earned A's have been hired by Mathematica Policy Research and the Urban Institute.

Homeworks

- Some theory but mostly data work
- Focused on understanding the material covered in the lecture
- Data portion frequently includes replicating results seen in the book
- Because results in the book often do not include details and are from papers, I also include the necessary paper in each assignment.
- They are difficult, time consuming, and graded rigorously

Final project

- Similar to homeworks, mostly data and a bit more theory
- Draws from all topics discussed in the course
- Approaching Ph.D. level difficulty
- About a week and a half to work on it
- After finishing it, you'll be better at econometrics than many economists

Thesis presentations

- Present your thesis idea, data, and specification
- 15 mins. at the end of lecture
- Chance to get feedback on your thesis
- Graded by me with less emphasis on results and presentation but more on clarity of the idea and theory
- Would replace 2nd lowest homework grade

Chapter 1: Questions about Questions

Four main questions for each research project

1. What is the causal relationship of interest?
2. What is the ideal experiment which could be used to identify the causal relationship?
3. What is the identification strategy absent an experiment?
4. What is the method of statistical inference?

1. What is the causal relationship of interest?

- Does more formal education result in higher wages?
- Do democratic institutions cause faster economic growth?
- Does exporting to Mexico make it easier to export to Spain?
- Do hospitals make people healthier?
- Does smoking while pregnant cause babies to be born with lower birth weights?

2. What is the ideal experiment which could be used to identify the causal relationship?

Does smoking while pregnant cause babies to be born with lower birth weights?

The ideal experiment

- Prohibit a woman from smoking during her pregnancy then measure her baby's birthweight.
- Go back in time and force the same pregnant woman to smoke during the same pregnancy then measure the baby's birthweight.
- The difference in the two birthweights is the causal effect of smoking on birthweight.
- Do this for lots of women to get an average effect.

Clearly, this approach is problematic for many reasons.

3. What is the identification strategy absent an experiment?

In other words, how are you going to use data that was not generated in a randomized trial to approximate one?

Exploit "natural experiments"

- Most US states require students to enter school in the calendar year in which they turn 6 years old.
- Imagine school starts on Sep. 1.
- Ahn is born on Dec. 31 and would be 5 years 8 months old.
- Beth is born on Jan. 1 and would be 6 years 8 months old.
- Most states also require students to remain in school only until their 16th birthday
- If both students drop out on their birthdays, Ahn received a year more of schooling than Beth
- Birthdate is essentially random so Ahn and Beth were "randomly" assigned different amounts of education

Politically impossible to run a randomized trial to assign years of education that mimics this natural experiment

4. What is the method of statistical inference?

In other words, what method is used to learn about unobserved quantities (parameters, standard error, etc.) using what is observed (data)?

- What is the population to be studied?
- What are the characteristics of the sample of the population to be used?
- How is the outcome related to the determinants?
- What assumptions are used to construct parameters from the data?
- How are standard errors calculated?
- How robust are the estimates to alternative assumptions?

Preliminaries

Scalar random variable x

$$\text{Discrete } x\text{: } E[x] \equiv \sum_z z \Pr(x = z)$$
$$\text{Continuous } x\text{: } E[x] \equiv \int z f_x(z)\, dz$$
$$\text{Variance: } V[x] \equiv E\left[(x - E[x])^2\right]$$

Are these random or fixed?

Joint distribution of two scalar random variables x and y:

$$\text{Discrete } y\text{: } E[y \mid x] \equiv \sum_z z \Pr(y = z \mid x)$$
$$\text{Continuous } y\text{: } E[y \mid x] \equiv \int z f_{y|x}(z)\, dz$$
$$\text{Covariance: } \operatorname{Cov}[x, y] \equiv E[(x - E[x])(y - E[y])]$$

Are these random or fixed?

Uncorrelated vs. Mean-Independent

Uncorrelated

- x and y are uncorrelated when $\mathrm{Cov}\left[x, y\right] = 0$

Mean-Independent

- y is mean-independent of x when $E\left[y \mid x\right] = E\left[y\right]$

Does one condition imply the other?

- Mean independence implies uncorrelatedness:

$$\begin{aligned}
E\left[(x - E\left[x\right])(y - E\left[y\right])\right] &= E\left[E\left[(x - E\left[x\right])(y - E\left[y\right]) \mid x\right]\right] \\
&= E\left[(x - E\left[x\right])(E\left[y \mid x\right] - E\left[y\right])\right] \\
&= E\left[(x - E\left[x\right]) \cdot 0\right] \\
&= 0
\end{aligned}$$

In general: independent $\Rightarrow$ mean-independent $\Rightarrow$ uncorrelated

Chapter 2: The Experimental Ideal

Main points of chapter 2

1. The selection problem is the main problem
2. Random assignment solves the selection problem
3. Experiments can be written as regressions

2.1 The selection problem is the main problem

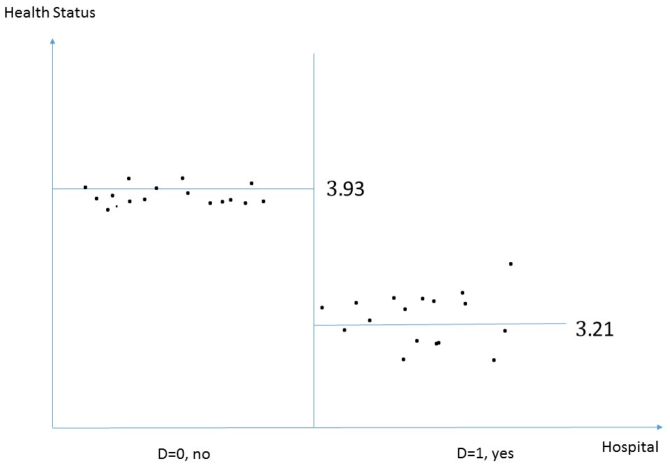
Do hospitals make people healthier?

Each individual i has one of two outcomes Y_i depending on if they go to the hospital $D_i = 1$ or not $D_i = 0$

$$Y_i = \begin{cases} Y_{1i} & if\ D_i = 1 \\ Y_{0i} & if\ D_i = 0 \end{cases}$$

Ideally, we would observe both outcomes for the same person (like the smoking mother example in the beginning). Absent a time machine, this is impossible. We can never see Y_{1i} and Y_{0i} directly.

Actual data on health and hospitalizations



Write the outcome as

$$Y_i = Y_{0i} + (Y_{1i} - Y_{0i})\, D_i$$

Observe average health outcomes given treatment

$$\begin{aligned} E\left[Y_i \mid D_i = 1\right] &= E\left[Y_{0i} + (Y_{1i} - Y_{0i})\, D_i \mid D_i = 1\right] \\ &= E\left[Y_{0i} + (Y_{1i} - Y_{0i}) \mid D_i = 1\right] \\ &= E\left[Y_{1i} \mid D_i = 1\right] \\ &= 3.21 \end{aligned}$$

Observe average health outcomes of those not treated

$$\begin{aligned} E\left[Y_i \mid D_i = 0\right] &= E\left[Y_{0i} + (Y_{1i} - Y_{0i})\, D_i \mid D_i = 0\right] \\ &= E\left[Y_{0i} + (Y_{1i} - Y_{0i})\, 0 \mid D_i = 0\right] \\ &= E\left[Y_{0i} \mid D_i = 0\right] \\ &= 3.93 \end{aligned}$$

Observed difference in average outcomes

$$
\begin{aligned}
\underbrace{E\left[Y_i \mid D_i = 1\right] - E\left[Y_i \mid D_i = 0\right]}_{\text{Observed difference in average}} &= E\left[Y_{1i} \mid D_i = 1\right] - E\left[Y_{0i} \mid D_i = 0\right] \\
&= \underbrace{E\left[Y_{1i} \mid D_i = 1\right] - E\left[Y_{0i} \mid D_i = 1\right]}_{\text{Average treatment effect}} \\
&+ \underbrace{E\left[Y_{0i} \mid D_i = 1\right] - E\left[Y_{0i} \mid D_i = 0\right]}_{\text{Selection bias}}
\end{aligned}
$$

We need a way to see $E\left[Y_{0i} \mid D_i = 1\right]$, the average outcome of not being treated, Y_{0i}, after the individual was treated, $D_i = 1$.

2.2 Random assignment solves the selection problem

We need to make the selection bias zero

$$\begin{aligned} E[Y_{0i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0] &= 0 \\ E[Y_{0i} \mid D_i = 1] &= E[Y_{0i} \mid D_i = 0] \end{aligned}$$

If Y_{0i} is independent of D_i, then $E[Y_{0i} \mid D_i = 1] = E[Y_{0i}] = E[Y_{0i} \mid D_i = 0]$.

Randomly assigning treatment ensures that D_i is independent of everything.

Once we have random assignment, we don't need a model, we can just compare means.

If experiments are so great, why don't we use them exclusively?

- Expensive
- Impossible (politically, physically, etc.)
- Unethical
- Takes too long (might have to wait 50 years for outcomes)

2.3 Experiments can be written as regressions

Write health outcomes as a linear model

$$\begin{aligned} Y_i &= Y_{0i} + (Y_{1i} - Y_{0i})\, D_i \\ Y_i &= E[Y_{0i}] + (Y_{1i} - Y_{0i})\, D_i + Y_{0i} - E[Y_{0i}] \\ Y_i &= \alpha + \rho D_i + \eta_i \end{aligned}$$

The observed average outcome after treatment is

$$\begin{aligned} E[Y_i \mid D_i = 1] &= E[\alpha \mid D_i = 1] + E[\rho D_i \mid D_i = 1] + E[\eta_i \mid D_i = 1] \\ &= \alpha + \rho + E[\eta_i \mid D_i = 1] \end{aligned}$$

and observed average non-treatment outcomes is

$$\begin{aligned} E[Y_i \mid D_i = 0] &= E[\alpha \mid D_i = 0] + E[\rho D_i \mid D_i = 0] + E[\eta_i \mid D_i = 0] \\ &= \alpha + E[\eta_i \mid D_i = 0] \end{aligned}$$

where the conditional expectations just leave parameters ρ and α.

Model errors and selection bias

The following steps connect our discussion of selection bias to conditions for the model errors

$$
\begin{aligned}
& E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0] \\
= \quad & \alpha + \rho + E[\eta_i \mid D_i = 1] - \alpha - E[\eta_i \mid D_i = 0] \\
= \quad & \rho + E[\eta_i \mid D_i = 1] - E[\eta_i \mid D_i = 0] \\
= \quad & \underbrace{\rho}_{\text{ATE}} + \underbrace{E[\eta_i \mid D_i = 1] - E[\eta_i \mid D_i = 0]}_{\text{Selection bias}}
\end{aligned}
$$

Using the notation introduced earlier

$$
\begin{aligned}
& E[Y_i \mid D_i = 1] - E[Y_i \mid D_i = 0] \\
= \quad & E[Y_{0i} + (Y_{1i} - Y_{0i}) D_i \mid D_i = 1] - E[Y_{0i} + (Y_{1i} - Y_{0i}) D_i \mid D_i = 0] \\
= \quad & E[Y_{0i} \mid D_i = 1] + E[(Y_{1i} - Y_{0i}) D_i \mid D_i = 1] \\
- \quad & E[Y_{0i} \mid D_i = 0] - E[(Y_{1i} - Y_{0i}) D_i \mid D_i = 0] \\
= \quad & \underbrace{E[(Y_{1i} - Y_{0i}) D_i \mid D_i = 1]}_{\text{ATE}} + \underbrace{E[Y_{0i} \mid D_i = 1] - E[Y_{0i} \mid D_i = 0]}_{\text{Selection bias}}
\end{aligned}
$$

Our obsession

Econometrics is obsessed with

$$E\left[\eta_i \mid D_i\right] = 0$$

the error term in our models must be mean-independent, and hence uncorrelated, with the regressors.

Essentially all the work we do is to ensure this condition holds.

Tennessee STAR experiment

Do smaller classes cause higher student achievement?

- Cost $12mn
- Cohort of kindergartners in 1985-86.
- Study ran for 4 years, until original cohort was in 3rd grade
- Included 11,600 children
- Treatment 1: small classes with 13-17 children
- Treatment 2: 22-25 children and a part-time teacher's aide (default)
- Treatment 3: 22-25 children and a full-time teacher's aide

Is the random assignment actually random?

Table 2.2.1: Comparison of treatment and control characteristics in the Tennessee STAR experiment

	Students who entered STAR in kindergarten				
	Variable	Small	Regular	Regular/Aide	Joint *P*-value
1.	Free lunch	.47	.48	.50	.09
2.	White/Asian	.68	.67	.66	.26
3.	Age in 1985	5.44	5.43	5.42	.32
4.	Attrition rate	.49	.52	.53	.02
5.	Class size in kindergarten	15.10	22.40	22.80	.00
6.	Percentile score in kindergarten	54.70	48.90	50.00	.00

Regression on experimental data

$$y_i = \alpha + \rho D_i + X_i'\beta + \varepsilon_i$$

where

- $y_i \in [0, 100]$ percentile score
- $D_i = \{0, 1\}$ assigned to small class or not
- X_i' child specific covariates like race, gender, income proxy, white teacher, teacher experience, teacher education

How do we interpret the coefficients?

What effect does the treatment have?

$$y_i = \alpha + \rho D_i + X_i'\beta + \varepsilon_i$$

Table 2.2.2: Experimental estimates of the effect of class-size assignment on test scores

Explanatory variable	(1)	(2)	(3)	(4)
Small class	4.82	5.37	5.36	5.37
	(2.19)	(1.26)	(1.21)	(1.19)
Regular/aide class	.12	.29	.53	.31
	(2.23)	(1.13)	(1.09)	(1.07)
White/Asian (1 = yes)	–	–	8.35	8.44
			(1.35)	(1.36)
Girl (1 = yes)	–	–	4.48	4.39
			(.63)	(.63)
Free lunch (1 = yes)	–	–	-13.15	-13.07
			(.77)	(.77)
White teacher	–	–	–	-.57
				(2.10)
Teacher experience	–	–	–	.26
				(.10)
Master's degree	–	–	–	-0.51
				(1.06)
School fixed effects	No	Yes	Yes	Yes
R^2	.01	.25	.31	.31

robust s.e. in parentheses

Wrapping up

- One can always regress something on something a get a number for β
- Whether that number means what you hope it means depends on your ability to ensure $E\left[\varepsilon_i \mid X_i\right] = 0$ i.e. your solution to selection/endogeneity
- The rest of the course primarily focuses on controling for selection via different methods
- Next lecture, 3/17, will discuss in detail why simple linear regressions that are correctly estimated are so powerful
- Please read Ch. 2 and 2/3 of Ch. 3 of MHE
- Please do the homework due on 3/17