MBR: Quantitative Methods

Harm Schuett

Winter 2016/2017

E-mail: schuett@bwl.lmu.de
Office Hours: tba
Office: Ludwigstrasse 28RG / Room 423

Web: github.com/hschuett/mbr-quanti
Class Hours: 9:00 am - 5:00 p.m.
Class Room: Ludwigstrasse 28RG / Room 416

This Syllabus is preliminary and will be updated continuously (and uploaded to github) till January 9th 2018

Course Dates

Day	Date	Room	Time
1	Tue, 9. Jan, 2018	416	9 am - 5 pm
2	Thu, 11. Jan, 2018	416	11:30 am - 7 pm
3	Wed, 17. Jan, 2018	B-103	9 am - 5 pm
4	Fri, 19. Jan, 2018	416	9 am - 5 pm
5	Mon, 22. Jan, 2018	416	9 am - 5 pm
6	Mon, 29. Jan, 2018	416	9 am - 5 pm

Course Description

This course is intended to teach you the three key pillars of econometrics: Identification (what can be learned in principle?), Estimation (how best to learn it?), Inference (how certain can we be in what we learned?). We will start with linear regression and work our way up to extensions such as fixed effects estimators, and Difference-in-Difference estimators. We'll switch to matching estimators and highlight its link to normal regression analysis, then spend quite some time on instrumental variables and similar approaches to causal inference. We will expand the linear regression framework to so called "general linear models" in order to analyze ordinal and nominal variables (e.g., logistic regression). If time permits, the course will end with a glance at the
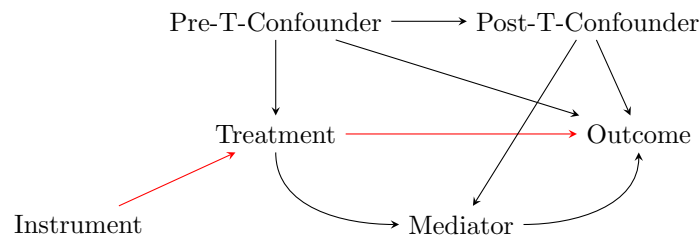

Figure 1: The many headaches of an empirical analysis

advanced topics of General Method of Moments, Bayesian Statistics, and Shrinkage. Throughout the course, the focus will be to develop a solid understanding of the key statistical concepts and tools, underlying modern empirical thinking: Identification and inference in the presence of uncertainty, potential outcomes, and graph theory. My hope is that if you grasp these you will:

1. be able to structure testable hypotheses and classify each challenge of testing those hypotheses as one of the three basic identification problems: confounding bias, endogenous selection bias, and over-control bias.
2. have a clear idea how to setup up an empirical test (e.g., what estimator to choose, which variables to include, which **not** to include),
3. see the connections between all discussed approaches: how they offer different ways of exploiting variation in the data based on what you assume about the underlying data,
4. have gained enough econometric understanding to not shy always form approaching advanced topics on your own.

Required Software

All of the course will be taught in R. You will be given code and learn how to implement all the covered topics in R. The course handouts will be R-Notebooks (In fact the course syllabus is compiled from a R-Markdown document). So please, bring a laptop with the newest R version and RStudio installed to each class. Both is free software, which is part of the reason why I chose to hold the course in R. The other is that R and Python have taken the data and scientific community by storm during the last few years and you will likely need to know both in the future. Try to play around with R a bit before the course starts. Part of the course will be to teach you R. But it helps immensely if you try to find your way around before. In particular, figure out how to install additional packages from within R Studio (check for instance R for Data Science) and install the following packages:

- tidyverse
- lfe
- haven
- igraph
- gridExtra
- scatterplot3d
- *(list will be updated)*

Reading

On a few days, I will post after-class reading. The reading will typically consist of well-written, short review articles or great exemplary empirical studies that will help repeating and understanding the material we covered. Below, please find a list of recommended text books. You don't have to read these if you can follow along just fine with just the handouts. But I encourage you to have access to at least the first two in case you want to check up on material. In a lot of cases, I will refer to the textbooks for proofs and further information. And if you do empirical work later on, you definitely want to have some of them ready at hand. *(Note that Gelman and Hill are significantly revising their book at the moment. A new version is expected some time in 2018)*

We will make heavy use of R Notebooks. This will be my first time trying them out as teaching tool. My idea is to have them be a type of interactive handout/slides crossbreed. This will give you the opportunity to try out the graphs and analysis that you will see in class and adjust the handouts in any way that helps you better remember the stuff. And hopefully it will help you learn R on the way too. Let's see how well this goes.

Great Textbooks

Angrist, Joshua D and Jörn-Steffen Pischke (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton University Press.

Gelman, Andrew and Jennifer Hill (2007). *Data analysis using regression and multilevelhierarchical models*. Cambridge University Press.

Goldberger, Arthur Stanley (1991). *A course in econometrics*. Harvard University Press.

Wooldridge, Jeffrey M (2010). *Econometric analysis of cross section and panel data*. MIT Press.

Great Additional Reading

McElreath, Richard (2016). *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC Press.

Tufte, Edward R (1990). *Envisioning information*. Graphics Press.

Course Policiy

Attendance Policy

> *Showing up is 80 percent of life* – Woody Allen, [via Marshall Brickman](#)

Class participation is necessary to understand the presented methods and to use them properly. MBR participants are required to attend at least 75% of the course hours to receive a passing grade. At the end of this course there will be an exam which is mandatory for all MBR participants.

Grading Policy

The grade will be determined by an exam at the end of the course. *Details on the exam will follow in December*

E-mail Policy

You can always write me an e-mail; I am usually quick to respond. If you have longer questions however, I would prefer if you would come to my office hours. It is a more efficient use of your and my time in that case. As a general rule, if you have an administrative question, please look at the syllabus first. I will not answer emails with questions that should be clear from just looking at the final syllabus at the end of December.

Class Schedule

Tue, 9. Jan, 2018 - Day 1: Inference in the Presence of Uncertainty

Part 1: Statistical Basics

- Introduction to econometric and statistical thinking
- The goal of most of scientific analysis: inference (vs. say prediction)
- Identification as a population concept versus inferring from samples
- Intro to graph theory

Excercise part: Intro to R + simulation and real data excercises to understand inference under uncertainty

Part 2: Regession Analysis I

- Introduction to linear regression models
- Basic assumptions and properties of linear regression models

Excercise part: Paper case 1: Replicating a study, reporting, interpreting, and analyzing regression output.

Thu, 11. Jan, 2018 - Day 2: Identification

ATTENTION: Because of a conflict with another MBR course, we will start at 11:30 am and finish at 7:00 pm on Tuesday the 11th.

Part 1: Causality, Identification

- How to structure an empirical test using the potential outcomes and causal diagram frameworks
- Confounders, endogeneity, and correlated omitted variable bias
- The three basic identification problems: confounding bias, endogenous selection bias, and over-control bias

Excercise part: Simulation and real data excercises to understand causes of consequences of confounding variables

Part 2: Regession Analysis II

- Regression model properties in the presence of confounders.
- Post-treatment and Pre-treatment variables.
- The problem with controlling for post-treatment variables.
- How to design a regression test. How not to test for mediation.
- What variables to include and which **not** to include.

Excercise part: simulations

After-class Reading:

Elwert, Felix and Christopher Winship (2014). "Endogenous selection bias: The problem of conditioning on a collider variable". In: *Annual Review of Sociology* 40, pp. 31–53.

Wed, 17. Jan, 2018 - Day 3: Matching Estimators and Panel Models

ATTENTION: This day will be held in a different room: Edmund-Rumpler-Str. 13, B-103. Sorry for the inconvenience.

Part 1: Panel Fixed-Effecs Models

- Introducing advantages and disadvantages of panel models
- Using panel models to control for unobserved confounders.

Excercise part: Replicating a study and computing fixed-effects models.

Part 2: Potential Outcomes and Control Groups

- The potential outcomes framework
- Selection-on-observables via matching
- Modern matching procedures
- Comparison to linear regression
- Synthetic control groups

Excercise part: Replicating a study and comparing matching and regression estimates.

Fr, 19. Jan, 2018 - Day 4: Instrumental Variables

Part 1: Instrumental Variables - The Claim to Fame of Econometrics

- Instrumental Variables as a way to circumvent confounding variables
- The logic of instrumental variables via the causal path framework
- Estimation issues with instrumental variables.
- Why you should always look for instrumental variables and why you shouldn't

Excercise part: tba.

Part 2: Stretch Block.

- Reserved if some topics need more time
- Question session. Chance to review key parts of the course. Chance for you to ask questions.
- Bigger paper replication.

Excercise part: tba.

Mon, 22. Jan, 2018 - Day 5: Natural Experiments

Part 1: Diff-in-Diff

- The ideal case: finding a natural experiment. Almost as good as randomly assigned.
- Pre-/post treatment for treatment and control group in a natural experiment.
- The all-important common trends assumption.

Excercise part: tba.

Part 2: Regression Discontinuity

- Searching for distribution kinks and defendable comparisons
- Problems with RD designs.
- Examples.

Excercise part: tba.

Mon, 29. Jan, 2018 - Day 6: General Linear Models and Advanced Topics

Part 1: General linear models

- Extending the regression framework to include ordinal and binary data.
- Logit and Probit models.
- Complications when interpreting Logit and Probit coefficients.
- Poisson Regression.
- Over-dispersion.

Excercise part: tba.

Part 2: GMM and Bayesian Statistics

- What to do if one equation is not enough
- Intro to structural equation modelling

Excercise part: tba.