Syllabus for "Applied Quantitative Methods in Democracy Studies and European Studies Research" (33335h)

Felix Lennert

Winter term 2020/21

Contact

As the entire course will take place online, we can only communicate through our laptops and phones. This is fairly unfortunate, especially because I cannot simply do some troubleshooting on your computer or anything similar.

- E-mail: *felix.lennert@liu.se* – please use my LiU mail address for whatever inquiries you have. I have given the Uni Regensburg mail multiple shots, but I simply cannot get it to work with several mail clients on my computer. Hence, to make sure that mails reach me within a reasonable time, use the address stated here.
- Website: Everything will be put on *GRIPS*.
- Office hours: on Fridays, 10–11; will take place online. Please sign up *here* upfront. If nobody has signed up, I will not open the Zoom room. If you cannot make it to one of the office hours, just drop me an e-mail and we set something up. This especially applies to students who live in different time zones. The Zoom Room has the meeting ID 911 0279 5820. The password is "office". Join the meeting using this *link*.
- Class hours: Wednesdays, 10–12; Join meeting *here* – meeting ID is 972 9681 0305, the password is "R".

Course description

When thinking about issues of European integration, a couple of things come to one's mind. Among them is the $64K question: why are some people more in favor of the EU than others? Or, put differently, do the supporters bear features that distinguish them from the opponents? Political scientists with a predilection for statistics call such features predictors. Furthermore, they measure support for European integration on a Likert scale, throw it all into a regression equation, and can then come to a conclusion. How do they know whether different features actually matter? The stars will tell, at least if they indicate p<.05.

But does correlation imply causation? Not necessarily. However, by harnessing smart research designs, scientists can strengthen their claims of causality.

Sometimes, context matters. Countries in the European Union are different, and there might be things investigators simply cannot observe – "unobserved heterogeneity" – and they might miss out on important features, so-called "omitted variable bias." Again, this is no big deal for aforementioned quant-heads: they get themselves some longitudinal data, assume that this heterogeneity remains the same over time – "time-invariant" – and model their artificial control group with fixed effects.

Experiments are seen as the gold standard for causality claims. Roughly, the procedure looks as follows. Scientists try to keep all conditions equal. They split up their sample into two groups. One group receives a treatment, the other group does not. This is for comparison's sake. If there is a change in the treatment group and not in the control group, the researchers will become very excited: now they can claim that they have isolated a causal effect.

But such experiments are difficult undertakings when studying entire societies. Luckily, they sometimes happen more or less naturally. If political scientists want to evaluate what effect different immigrant integration policies had on immigrant integration, they take several countries with diverging policies, consider the policies treatments, and look at the situations before and after their introduction. Then, by comparing the "treated" with the "untreated," they can evaluate the actual effect of the policy. This research design is called difference-in-difference.

Those are the three "applied quantitative methods" I will teach you. Your tool will be R. It is sometimes stubborn but always powerful and you will learn how to use it properly. You will form groups with like-minded people, discuss theories, come up with research questions, derive hypotheses from the aforementioned theories, use the methods mentioned above to test them, and add to the basin of knowledge we scientists are dabbling in.

As you may have already noticed, our language of instruction is going to be English. Since I assume that we are all not going to be native speakers, that should not do any harm and nobody has to feel embarrassed if their command is not top-notch. Former experience in R, Python, or even SPSS is not required yet comes in handy.

Course Objectives

We are going to be a highly diverse bunch in terms of our backgrounds. Some – the political scientists, presumably – might be already familiar with the theoretical content, others will probably be underwhelmed by the statistical methods, and perhaps there will be some students who are proficient coders (maybe even in R!).

Existing research has suggested that groups with higher diversity of prior knowledge show performance and innovation that are superior over homogeneous groups (see, e.g., the brief summary of existing research in van Knippenberg/De Dreu/Homan 2004, pp. 1008–1009). In my opinion, this is an opportunity you should not miss out on and I have therefore decided to strip you of the opportunity to choose your colleagues and have allocated you to groups. The diversity of skills was maximized using your answers in the introductory survey.

As diverse as your skill sets are the intended learning outcomes: at the beginning of every successful research paper stands a puzzling research question. Hence, it is crucial for us to begin with some theoretical groundwork. The emphasis will lie on reading extensive literature reviews. However, I will only remain on the surface – it is your turn to decide where you and your group start to dig deeper until you stumble across a question which is puzzling enough for you to go for answering it.

In order to make this process more systematic, existing theoretical approaches are presented. They will provide you with different perspectives to look at the questions you have come up with. Starting from there, you can derive and formulate hypotheses. Since the data at hand is an obvious constraint when it comes to testing hypotheses (not for the derivation itself...), I will also familiarize you with several data sources for your project.

The hypotheses must have certain assets. First and foremost, they need to be testable. But how could they be tested? Several strategies exist. We will focus on one of the most common and intuitive approaches: linear regression.
Its foundations – and also some handy variations (e.g., Difference-in-difference) – are presented. Due to your diverse backgrounds, I will be light on mathematics. I rather want to give you an intuition of how you can use these techniques, what they can tell you, and get you up and going as quickly as possible.

Since calculating regression coefficients by hand is a pretty tedious task, an introduction to R is provided as well. The R sessions consist of pre-recorded videos and accompanying scripts with exercises. Thus, students can watch the materials and work through them at their own pace. Though, it is mandatory to hand in assignments on a weekly basis.

To summarize the objected outcomes:

- R

- – Introduction to data wrangling with the `tidyverse`
- – Building regression models with `tidymodels`
- – Visualization using `ggplot2`
- – Writing papers with `RMarkdown`
- Theoretical Aspects
  - – Theories regarding public support
  - – Basics of empirical social scientific research
  - – A roadmap for thinking about European integration
  - – The problem with causality in the empirical social sciences
- Quantitative Methods
  - – Linear Regression
  - – Difference-in-Difference
  - – Fixed Effects

Literature

The course's foundation is three books: for the R content, it is Hadley Wickham's and Garrett Grolemund's *"R for Data Science"*, for the methodological content Kosuke Imai's "Quantitative Social Science." and for everything in between *"Econometrics in R"* by Christoph Hanck, Martin Arnold, Alexander Gerber, and Martin Schmelzer. The former and the latter can be read online, "Quantitative Social Science" can be found in the library and relevant chapters will be uploaded to GRIPS. All things EU-related will be covered using journal articles and the book "European Politics. A Comparative Introduction" by Tim Bale (relevant chapters will be uploaded to GRIPS).

Course Policies

Below you can find some basic rules of behavior for the course and what you will have to do to pass it.

Basic rules of behavior

- Mute your microphone during Zoom lectures at all times – except for when you want to say something.
- Do not hesitate to interrupt me whenever.
- Never discriminate against any of your classmates due to whatever.
- If anything course-related bothers you, send me an e-mail.
- Feel free to copy code from Stackoverflow or whatever resources pop up in your Google searches. I do not consider this cheating. If anybody considers this cheating, they have probably never coded themselves. However, all scripts will be treated and graded as if they were fully written by you, so please make sure that you understand what you copy.

Attendance policy

Right now, attending class is probably easier than ever – you can just stay in bed. However, if you are not able to make it to class due to technical problems, send me an e-mail. I can provide you information on how to join in via your cell phone. **If you, for whatever reason, miss the session, you are obliged to hand in 500 words on the readings so that I can make sure that you have actually worked through the material.**

Everything R-related will happen unsupervised. At the end of the scripts, you will find exercises. Solutions for them will be posted after the due date. I am more than happy to have a look at your solutions and talk about them during office hours (which can, of course, be extended as well). Please contact me beforehand and include your code.

E-mail policy

E-mails will probably be our main means of communication. As my time is limited, consider the following problem-solving strategies before sending an e-mail:

- Give this syllabus a second look.
- If R-related: Google your problem. Try at least three solutions from Stackoverflow.
- If R-related: Give the script I provided you with (and its references) a second look. I take most of the inspiration for the exercises from the resources I use for writing the script. Hence, the solution should be in there. (If you cannot find the solution in the script, please point that out. Each task should be solvable with the script at hand, i.e., the commands and functions you need for solving the problem need to be somewhere in the script. If this is not the case, I will have to adapt the script.)

If you were not able to solve the problem by following the aforementioned strategies, please stick to the following rules:

- Choose a meaningful subject that starts with "advanced-methods:" – that's how I know whose e-mails need to be answered first.
- If it is an R-related problem, try to provide me with a *reprex*.
- Please try to be concise about your problem.
- If it will take me more than 15 minutes to type a meaningful answer, sign up for office hours or ask for an additional slot.

Examination

I will not distinguish between what program you are in or what module you chose the course for. There are three things your group will have to hand in:

- A tentative exposé. It should contain the research question, a brief motivation for it, and what other researchers have written about it so far. You are going to present it to the class in the session before Christmas.
- The final exposé. Same as the tentative one but including the hypotheses and a reasonable testing strategy (method(s), data sources). Each group has to present another group's exposé and give extensive feedback.
- The paper. Should resemble an actual research paper. More on how to write one and the expectations will be announced in class.

Class Schedule

Classes are going to take place on Zoom on Wednesdays, 10–12 c.t. (i.e., 10.15–11.45). Some sessions might be shorter, depending on the material we are going to cover. Everything R-related is going to be taught in an asynchronous manner using pre-recorded material. The assignments are to be handed in on GRIPS.

Given the remote nature of this semester and the isolation you might face, I tried to incorporate as many group elements as possible, so that you can get into contact with your peers.

The readings assigned to the session should be read **upfront**.

2020-11-04

Let's get it started!

- Getting to know each other
- Course overview – what's to expect:
    - Going through the course plan
    - The chores
    - Why methodological/statistical skills?
    - Why R?
- The preparatory survey
- After the session: assigning you to groups

Readings

No readings.

Homework

- Install R and RStudio
- Get in contact with your group members

2020-11-11

An introduction to R

- Setting up R and RStudio
- RStudio Projects
- Using it as a calculator
- Installing packages
- Types of vectors

Readings

No readings.

Homework

- Exercises due before session on 2020-11-18.

2020-11-18

European integration

- A brief history of the EU – mile-stones and their effects on studying the EU
- One way of thinking about European integration

Readings

- Bale, Tim 2005: *European Politics. A Comparative Introduction.* New York: Palgrave Macmillan. Chapters 1 and 2.
- Stone Sweet, Alec/Sandholtz, Wayne 1997: European integration and supranational governance. *Journal of European Public Policy* 4:3, pp. 297–317.

Homework

- R script + video; subject: **data wrangling I**; exercises due before session on 2020-11-25.

2020-11-25

How is it studied?

- The EU – unprecedented in a couple of ways
- Fields of study
- Discussion: potential explanations for support of, and skepticism towards, European integration in breakout rooms:
    - Group 1: the utilitarian approach – who profits most?
    - Group 2: identity
    - Group 3: national-political contexts

–> You will prepare a brief presentation (~10 minutes) in your respective group and lecture your peers in the second part of the session

Readings

- Hobolt, Sara/De Vries, Catherine 2016: Public Support for European Integration. *Annual Review of Political Science* 19(1), pp. 413–432.
- For group 1: Anderson, Christopher/Kaltenthaler, Karl 1996: The Dynamics of Public Opinion toward European Integration, 1973–93. *European Journal of International Relations* 2(2). pp. 175–199.
- For group 2: Carey, Sean 2002: Undivided Loyalties. Is National Identity and Obstacle to European Integration?. *European Union Politics* 3(4), pp. 387–413.
- For group 3: Sánchez-Cuenca, Ignacio 2000: The Political Basis of Support for European Integration. *European Union Politics* 1(2), pp. 147–171.

Try to obtain the readings yourself first. If you fail, drop me a mail.

Homework

- Get together with your group, think about research questions that lend themselves to the investigation.
- R script + video; subject: **data wrangling II**; due before 2020-12-02

2020-12-02

Empirical Political Science: a primer/refresher

- The quali-quanti schism
- The Coleman boat
- The logic behind empirical political science
    - Understanding the scientific method
    - Accordingly, how to design a research paper?
    - How is the exposé supposed to look like?
- A few notes on data
    - Scale levels
    - Potential data sources (and how you can access them)

Readings:

No mandatory readings, everybody should be familiar with the content. However, if you are interested, here is the list of resources I used for preparing the session:

- Rich, Richard/Brians, Craig/Manheim, Jarol/Willnat, Lars 2018: *Empirical Political Analysis. Quantitative and Qualitative Research Methods.* London: Routledge. – Chapters 2 and 3.
- Ylikoski, Petri 2016: Thinking with the Coleman Boat. *The IAS Working Paper Series* 2016(1). Linköping: The Institute for Analytical Sociology.
- Coleman, James 1986: Social theory, social research, and a theory of action. *American Journal of Sociology* 91, pp. 1309–1335.

Homework

- Get together with your group, rethink your ideas, start working on the exposé
- R script + video; subject: **data visualization**; due before session on 2020-12-09

2020-12-09

Introduction to linear regression

- Basic principles
- Mathematical principles
- Hypothesis tests

Readings:

- Imai, Kosuke 2017: *Quantitative Social Science.* Princeton: Princeton University Press, pp. 123–160, 342–369. (can be found on GRIPS)
- Hanck, Christoph/Arnold, Martin/Gerber, Alexander/Schmelzer, Martin 2020: *Introduction to Econometrics with R.* Chapters 4, 5.1–3, and 7. Currently in Open Review, find it here.

Homework

- R script + video; subject: **regression I**; assignments due before 2020-12-16

2020-12-16

More on linear regression

- Assessing the quality of results
    - RSS
    - Heteroskedasticity and homoskedasticity
- Correlation $\neq$ causation
    - Some more or less funny examples
    - How to avoid this pitfall
    - Multiple predictors ("control variables")
- Again: some notes on the data available

Readings:

- Hanck, Christoph/Arnold, Martin/Gerber, Alexander/Schmelzer, Martin 2020: *Introduction to Econometrics with R.* Chapters 5 and 6. Currently in Open Review, find it here.
- Imai, Kosuke 2017: *Quantitative Social Science.* Princeton: Princeton University Press, pp. 156–181, 370–389. (can be found on GRIPS)

Homework:

- R script + video; subject: **regression II**; assignments due before 2021-01-13
- Come together with your group and discuss the following questions:
    - What is the research question and why is it relevant?
    - Which theoretical angle can be taken?
    - What have other researchers done so far?
    - (Optional: your hypotheses)
- Put together a brief presentation (max. 10 slides, pdf format) and send it to me before 2020-12-23

2020-12-23

Recapitulating before Christmas

- Every group presents their project (10 minutes)
- Feedback from others is crucial!
- I will run you through the coming weeks

Off you go into your well-deserved winter break. You are neither expected nor supposed to work on anything during your weeks off.

2021-01-13

Variations of regression

- More things on how to go about causality
    - Harnessing panel data: fixed effects
    - Seizing the occurrence of (more or less) natural experiments: Difference-in-Differences

Readings:

- Hanck, Christoph/Arnold, Martin/Gerber, Alexander/Schmelzer, Martin 2020: *Introduction to Econometrics with R*. Chapter 10. Currently in Open Review, find it here.
- Imai, Kosuke 2017: *Quantitative Social Science*. Princeton: Princeton University Press, pp. 54–69. (can be found on GRIPS)

Homework:

- R script + video; subject: **Fixed Effects and Difference-in-Differences**; assignments due before 2021-02-03
- Prepare yourself for the upcoming group discussion.

2021-01-20

Another group discussion

- Various groups, various papers
    - First: discussion in breakout rooms
    - Second: groups present their assigned papers to the rest of the course
- Focus lies on the methodology: how have the methods enabled the researchers to study things the way they did it and make the claims they made?

Readings:

- Neureiter, Michael 2019: Evaluating the effects of immigrant integration policies in Western Europe using a difference-in-differences approach. *Journal of Ethnic and Migration Studies* 45(15), pp. 2779–2800.
- Gabel, Michael/Whitten, Guy 1997: Economic Conditions, Economic Perceptions, and Public Support for European Integration. *Political Behavior* 19(1), pp. 81–96.

Homework:

- Prepare the final exposé. Deadline: evening before 2021-01-27

2021-01-27

Exposés are circulated

- Each group gets another group's exposé
- Will have to present it (~15 minutes) and give feedback on it (~5–10 minutes) on 2021-02-03 and 2021-02-10

2021-02-03

Present your peers' final projects 1/2

2021-02-10

Present your peers' final projects 2/2

Also: our last session

- Summary
- Feedback
- Off you go.