

Advanced Statistical Models in Political Analysis using R

Fall 2017

Course Details:

<i>Lecture:</i>	Mon, 12.15 - 13.45 AFL-F-121	
	Lukas F. Stoetzer	Benjamin Schlegel
E-mail:	lukas.stoetzer@uzh.ch	kontakt@benjaminschlegel.ch
Office:	H-341	H-343
Office Hours:	Tue, 14.00 - 15.30	Mon, 14.30 - 16.30 ¹

Course Description:

Building on the analytical and theoretical background of the ("Statistical Models in Political Analysis"), this seminar on "Advanced Statistical Models in Political Analysis" introduces interested Master students and graduate students to strategies and tools of how to develop statistical models that are tailored to answer their particular research questions.

This course serves as an introduction to a multitude of probability models that are appropriate when the linear model is inadequate. In the first half we will focus on the statistical theory of maximum likelihood. The second half of the course discusses models that are particularly relevant for comparative research, where the independence assumption of generalized linear models is often unmet. Throughout the seminar will also devote considerable time to statistical programming using R.

The course language is English. This includes lecture, lab sessions, discussions and assignments.

Course Registration:

The course is divided into a lecture and a lab session. During the lab sessions, students will apply the statistical models introduced in the lecture. You should bring your own Laptop to the sessions and attend the same session every week.

Note that this course is highly demanding and entails a substantial work load, in the form of six homework assignments and a final paper. Students who wish to audit this class

¹Please register online for Benjamin's R-Help Desk at https://politikwissenschaften.ch/rhelpdesk_registration.php

should notify the instructors in advance (participation is subject to free room capacity). Please note that only registered students will receive feedback on their written work.

Readings:

We will not use a single textbook for this course. Selected readings are available on the course website. The following books will be used in the course and provide further material on the topics covered:

Eliason, Scott R. 1993. *Maximum Likelihood Estimation: Logic and Practice*. Newbury Park: Sage.

King, Gary. 1989. *Unifying Political Methodology*. Ann Arbor: University of Michigan Press.

Gelman, Andrew, and Jennifer Hill. 2006. *Data analysis using regression and multi-level/hierarchical models*. Cambridge University Press.

Box-Steffensmeier, Janet M., et al. 2014. *Time series analysis for the social sciences*. Cambridge University Press.

Wooldridge, Jeffrey 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press.

Software:

For all calculations, we will support and use the open-source statistical programming language R. The course requires a sound working-knowledge of R. The lab sessions will be devoted to deepen students R-programming skills. Within the first four weeks we will learn how to program own estimators in R, implement them in an R-package and apply them to data sets from major fields in political science.

Prerequisites:

There is formally no prerequisite. This, however, does not mean that this is introductory class! Students should have a good working knowledge of R, passed at least one graduate level lecture on introduction to statistical modeling, and possess a good command of high school algebra.

Course Requirements:

Grading will be based on the following components:

- **Six Homework Assignments (25 %)**

The homework assignments will take the form of problem sets, programming, replications, simulations, or extensions of the analysis in class and the lab. The date at which the problem sets are handed out can be found in the course outline below. Generally, the assignments will be handed out at the end of lab sessions and you are expected to

hand in the solution at the beginning of the next lecture, Monday 12pm (unless noted otherwise). Late assignment will not be accepted.

Homeworks need to be handed in in groups of 2 students.² You should find one coauthor with whom you will work on the weekly homeworks. Please indicate the group members on your answer sheets. Moreover, you are strongly encouraged to seek advice from the instructors during office hours or through the online forum. For each homework you will be able to get up to 20 points. With the worst mark being canceled, the maximum point overall is 100.

- **Final Paper (75 %)**

There will be a final paper. Each student will produce a manuscript that applies or develops an appropriate statistical model to an important substantive problem. Students will choose their own topics. What works particularly well is to start replicating an already published article in order to develop it further for the final paper. My advice is to pick an article that interests you, was published within the last few years in a good journal, and uses methods we have or will talk about in class (or uses different methods at about the same level of sophistication).

The paper must include all analysis, tables, figures, and description of the results. A good write-up of the paper should read like the third quarter of a journal article. The rest of the draft may be in detailed outline form, although it would be better to have it fully written.

You also need to provide all necessary information to replicate your analysis. The replication material must include your dataset and computer code to be able to reproduce all tables and figures that make it in the paper. We expect you to comment your computer code heavily to explain what you are doing. Your code must be neatly formatted and run cleanly.

The final draft paper together with all replication material are due on **January 3th, 2017**. Please submit all files electronically by **5pm** that day. Late submissions will not be accepted.

Course Outline:

- **Introduction**

- **Week 1 (18.09.): Lecture**

- **OLS in Matrix Form**

- **Week 2 (25.09.): Lecture**

Wooldridge, Jeffrey, M. 2009. Introductory Econometrics. A Modern Approach. Appendix D and E.

In this lecture we derive the Ordinary Least Square estimator using matrix notation. To be able to follow the discussion please carefully work through

²Single author assignments will also be accepted.

Wooldridge's appendix D and E beforehand.

– **Week 3 (02.10.): Lab-Session**

The aim of this session is learn how to program your own OLS-function in R.

Homework 1 will be assigned.

• **Maximum Likelihood Estimation of Generalized Linear Models**

– **Week 4 (09.10.): Lecture**

Eliason, Scott R. 1993. Maximum Likelihood Estimation: Logic and Practice. Newbury Park: Sage. Chapter 1-4.

King, Gary. 1989. Unifying Political Methodology. Ann Arbor: University of Michigan Press. Chapter 4.

The aim of this lecture is to give a condensed introduction to maximum likelihood estimation. We work through the derivative of the linear regression model from maximum likelihood principals and underscore the power of this estimation technique by discussing a heteroskedastic regression model. Please carefully work through the readings, as these will provide you with a vibrant source for further applications of these models.

– **Week 5 (16.10.): Lab-Session**

In this session we implement our own maximum likelihood estimators in R. We discuss implementation of the heteroskedastic regression model from the lecture, but to get a sense of the usefulness of this approach apply it also the estimation to logit/probit models.

Homework 2 will be assigned.

• **Multilevel Hierarchical Models**

– **Week 6 (23.10.): Lecture**

Gelman, Andrew, and Jennifer Hill. 2006. Data analysis using regression and multilevel/hierarchical models. Cambridge University Press, Chapter 12-13.

Steenbergen, M., and Jones, B. 2002. Modeling Multilevel Data Structures. American Journal of Political Science, 46(1),218-237.

In week 6 we discuss models for data that are ordered hierarchically. An example for this are survey respondents in different countries. Our discussion of multi-level models centers on two-level models with fixed predictors for the level specific effects. For the lecture you should read Gelman and Hill Chapter

12-13, as well as Steenbergen and Jones Article. Gelman and Hill is an excellent introduction book with lots of applications and further advanced material.

– **Week 7 (30.10.): Lab-Session**

The lab session discusses the estimation of multi-level models in R, drawing on different examples from the literature.

Homework 3 will be assigned.

• **Time-series Models**

– **Week 8 (06.11.): Lecture**

Box-Steffensmeier, Janet M., et al. 2014. Time series analysis for the social sciences. Cambridge University Press.

In week 8 we turn our attention to dynamic processes by discussing univariate time series models. Our discussion centers around autoregressive processes and moving averages. The aim is to understand implications of time series dynamics. Please read chapter 2 of Box-Steffensmeier et al.. The textbook serves as a reference book that covers insights in modeling dynamic processes beyond the lecture.

– **Week 9 (13.11.): Lab-Session**

In the lab session we apply time series models to applications from political science, with a focus in estimating the correct dynamics.

Homework 4 will be assigned.

• **Panel Models**

– **Week 10 (20.11.): Lecture**

Wooldridge, Jeffrey 2002. Econometric Analysis of Cross Section and Panel Data. MIT Press.

In week 10 we discuss panel data analysis. Panels here refer to studies that contain repeated measurements of the same individuals over time. We compare different estimation strategies to consider unobserved unit heterogeneity: fixed and random effects, and first difference. Chapter 10 of Wooldridge books serves as the basis of our discussion, but will be a detailed and companion for further investigations in this direction.

– **Week 11 (27.11.): Lab-Session**

The lab-session employs examples to get familiar with the estimation of panel models for unit heterogeneity in R.

Homework 5 will be assigned.

- **Models for Time-series Cross-section**

- **Week 12 (04.12.): Lecture**

Beck, Nathaniel, and Jonathan N. Katz. 1995. "What to do (and not to do) with time-series cross-section data." *American Political Science Review*, 89(3), 634-647.

Pluemper, Thomas, Vera E. Troeger, and Philip Manow. 2005. "Panel data analysis in comparative politics: Linking method to theory." *European Journal of Political Research*, 44(2), 327-354.

Last but not least, we take a look a common data-source in political science: Time-series cross-section data. In this, a relatively small number of fixed units is observed over time, e.g. West European countries from 1970-2010. Our discussion centers around the de facto Beck-Katz standard in the literature, paying particular attention to the case when (and why) this model specification might not be appropriate. Please read Beck & Katz article to be well-prepared for the discussion.

- **Week 13 (11.12.): Lab-Session**

In the lab session we will replicate Pluemper et al re-analysis of Garrett and Mitchells study on government spending to see how the different estimation procedures can be implemented in R.

Homework 6 will be assigned.

- **Wrap-up**

- **Week 14 (18.12.): Lecture**