DEPARTMENT OF SOCIAL SCIENCES
Center for Doctoral Studies in the Social
and Behavioral Sciences
Prof. Thomas Gschwend, Ph.D.

UNIVERSITY OF
MANNHEIM

Visitors
D 7, 27 · Room 301/302
68131 Mannheim · Germany
Telephone +49-621-181-2087
Fax   +49-621-181-3699
E-Mail: gschwend@uni-mannheim.de
www.sowi.uni-mannheim.de/gschwend

Assistant (Christel Selzer)
Telephone +49-621-181-2413
E-Mail: methods@uni-mannheim.de

Advanced Quantitative Methods

Spring 2014

Course Details:

	Thomas Gschwend (Lecture)	Lukas Stötzer (Lab)
	gschwend@uni-mannheim.de	lstoetze@mail.uni-mannheim.de
Time:	Wednesday, 8:30–10:00	Thursday, 17:15–18:45
Place:	A 5,6 B 143	A 5,6 B 317
Office Hours:	Wednesday, 10:00–11:00	Thursday, 10:00–12:00

Course Description:

Building on the analytical and theoretical background of the previous course in our MA methods sequence ("Multivariate Analyses"), this course on "Advanced Quantitative Methods" introduces interested graduate students to strategies and tools of how to develop statistical models that are tailored to answer their particular research questions.

You might have noticed by now, the linear regression model is often an inappropriate tool for answering substantive questions in political science. This course serves as an introduction to a multitude of probability models that are appropriate when the linear model is inadequate. After introducing the fundamentals from which statistical models are developed, this course will focus on one specific theory of inference, namely on the statistical theory of maximum likelihood. We will also devote considerable time to statistical programming, simulating and conveying quantities of material interest of such models (using R) in order to encourage students to switch from a consumer-mode into a producer-mode of social science research.

The goal of this course is three-fold: (1) to prepare students to conduct research using appropriate statistical models and to communicate their results to a nontechnical audience;

(2) provide a foundation in the theory of maximum likelihood so students can investigate and implement a wide range of advanced statistical models; and (3) provide students with the tools necessary to fine-tune existing or to develop new statistical models of political phenomena.

Work through the assigned readings ahead of time. We expect everyone to come to class fully prepared. Expect that this will take considerably longer than in a substantive seminar. Do not skip equations! Instead, take notes, prepare questions and team-up with others to answer them, or as last resort, bring them up in class. After every class we expect you to go over the lecture notes and your notes once again. Furthermore, we additionally offer the possibility to send us questions by Wednesday night. We will try to address them in the lab session on Thursday afternoon. There is no point in getting lost — particularly not in an elective class.

In the weekly class meeting the course will be conducted as mix of lecture and practical computer lab sessions. During the lecture sessions we will highlight the central concepts and ideas of the readings (and at that time they should not be new to you!). In the more practical sessions, students will learn to program the statistical models introduced in the lecture. We will also work through examples collaboratively. Our pace will be unpredictable. Nevertheless, understand that the bulk of learning in this course will take place *outside* the classroom, by reading, practicing using statistical software, and doing problem sets.

As an elective class, the current curriculum does not offer an accompanying tutorial for credit, unfortunately. Nevertheless, we managed to offer optional lab sessions this time. We highly recommend that you make use of this opportunity. The lab sessions will be designed as an accompanying tutorial that will provide you with R-scripts and further exercises to help you comprehend the material that is covered in the lecture. Furthermore, in the lab sessions we will discuss issues that came-up in your homework or in the lectures.

Recommended for:

Political science graduate students, including M.A. and interested PhD students (CDSS).

Prerequisites:

M.A. students of political science should have successfully passed the previous course in the MA methods sequence about "Multivariate Analyses" as well as the accompanying "Tutorial Multivariate Analyses". PhD students should have passed equivalent courses. If you can tell me what $(X'X)^{-1}X'y$ is, you have the necessary background to take this class.

Course Registration:

Students who wish to take the course should register for "Advanced Quantitative Methods" at the student portal.

Note that this course is highly demanding and entails a substantial work load for students!

Students who wish to audit this class should notify the instructor 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 (through ILIAS). The following books will be used in the course:

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.

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Newbury Park.: Sage.

Hardware/Software:

Students are welcome to bring their own computers to class. R will be the software package of choice. There will be homework problems that require you to edit and write some R-code. The open-source statistical programming language R is particularly suited for carrying out state-of-the-art computer-based simulations and programming advanced statistical models. It also generates really nice publication-quality graphics. The software runs under a wide array of operating systems. R can be downloaded for free at http://www.r-project.org/. A readable introduction is given by Fox, John. 2002. *An R and S-Plus companion to Applied Regression.* Sage.

Course Requirements:

Grading will be based on the following components:

- **Homework Assignments (25%)**

  There will be a series of six homework assignments that will take the form of problem sets, replications, simulations, or extensions of the analysis in class and the lab. The assignments will be handed out at the end of class on Wednesday and you are expected to hand in a solution at the beginning of the next class (unless noted otherwise) a week later. You need to work through every homework assignment. Late assignment will not be accepted.

  We encourage you to work in small groups on the assignments. Usually 3-4 people per group seems to work best. If you have worked with another student, please indicate with whom you did so on your homework. Moreover, you are strongly encouraged to seek advice from both instructors during office hours or by email. Note that instructive discussions about the material are best done during office hours rather than by email.

- **Final Paper (75%)**

  There will be a final draft paper but no final exam. Each student will produce a co-authored manuscript (or a solo-authored manuscript, with permission of the instructor) 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 pic 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 draft paper must include all analysis, tables, figures, and description of the results. A good write-up of the draft 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. To that end, please avoid writing computer-specific lines into your code that will prevent it from running on other machines. We will award partial credit if necessary.

  The final draft paper together with all replication material are due on **June 4th, 2014**. Please submit all files electronically and, additionally, a hard-copy of your draft paper by **10am** that day. Late submissions will not be accepted.

What to do today?

Find a coauthor and start working on the the draft paper very soon.

Other Considerations:

A great website with many R code examples is the UCLA Stat Consulting Site. Another good site that introduces R to SPSS or Stata users is Quick-R.

Learn to use LATEXwhile you can. It is a free typesetting software package and enables you to typeset and print your work at the highest typographical quality, using a predefined, professional layout. The main advantages of LATEXover normal word processors include professionally crafted layouts, support for typesetting of mathematical formulae in a convenient way, a few easy-to-understand commands that specify the logical structure of a document, more complex structures such as footnotes, references, table of contents, bibliographies that can all be generated easily, and free add-on packages for specific tasks (e.g. make a reference list adhere to the exact standards of a scientific journal). A short introduction can be found here.

Detailed Course Outline:

Week 1 (12 February 2014): Introduction. OLS Recap.

Week 2 (19 February 2014): OLS in Matrix Form.

> Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks, CA: Sage. Chapter 1-2.

> Wooldridge, Jeffrey, M. 2009. *Introductory Econometrics. A Modern Approach.* Appendix D & E.

Work through Wooldridge's appendix D first, because it will be heavily used in appendix E. Try to test yourself by doing some concrete examples (2×2- or 2×3-matrices are totally fine) to make sure you understand what's going on. Moreover, I suggest to closely read section E.1 – E.3 and skim the rest of this appendix. Also, in order review the linear model in matrix form take a look at Scott Long's chapter 1-2, in particularly sections 2.1-2.5.

Homework 1 will be assigned.

Week 3 (26 February 2014): OLS in Matrix Form and Probability Theory

> King, Gary. 1989. *Unifying Political Methodology.* Ann Arbor: University of Michigan Press. Chapter $1 + 3$.

> Fox, John. 2007. *Applied Regression, Generalized Linear Models, and Related Methods*, 2nd Edition, Thousand Oaks, CA: Sage. Appendix D.

We will wrap-up our discussion of the linear model in matrix form and do some applications with probability distributions to get more familiarity with them. The core reading will be chapter 1 and 3 from King (1989). Furthermore, a nice overview about probability theory and particular probability distributions (which you also covered last semester) is provided by Fox (2007) in Appendix D. For next week focus merely on sections D.1-D.5.

Note that I provide Jon Fox's appendix in its entirety. If you want to read more about linear algebra take a look at his treatment in Appendix B. It includes nice examples and graphs that provide some more intuition.

Homework 1 is due and Homework 2 will be assigned.

Week 4 (5 March 2014): A first peek at Maximum Likelihood

> King, Gary. 1989. *Unifying Political Methodology.* Ann Arbor: University of Michigan Press. Chapter 4-4.3.

> Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks, CA: Sage. Chapter 2.6.

We will finally start with an introduction of the likelihood theory of inference. Please read chapter 4 (only until section 4.3) of King's UPM book. For a quick peek at MLE I'd like you to refer to Long's chapter 2.6.

Homework 2 is due.

Week 5 (12 March 2014): Maximum Likelihood Estimation and Heteroskedastic Regression

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

Franklin, Charles H. 1991. "Eschewing Obfuscation? Campaigns and the Perception of Senate Incumbents". *American Political Science Review* 85(4): 1193-1214.

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

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks, CA: Sage. Chapter 3.6.1 - 3.6.2.

Make sure you closely (re)-read the entire King's UPM, chapter 4. For those of you who appreciate a slightly different take on MLE take a look at Eliason (1993). Please also read a short section in Long (1997) chapter 3.6.1 and 3.6.2 in order to get a sense of how to actually estimate standard errors using maximum likelihood. For an nice application on how to set-up a heteroskedastic regression model take a look at the "classic" Franklin (1991) paper.

Homework 3 will be assigned.

Week 6 (19 March 2014): Models for Binary Dependent Variables

Esarey, Justin, and Andrew Pierce. 2012. "Assessing Fit Quality and Testing for Misspecification in Binary-Dependent Variable Models." *Political Analysis* 20(4), 480–500.

Greenhill, Brian, Michael D. Ward, and Audrey Sacks. 2011 "The Separation Plot: A New Visual Method for Evaluating the Fit of Binary Models." *American Journal of Political Science* 55(4): 991–1002.

King, Gary. 1989. *Unifying Political Methodology.* Ann Arbor: University of Michigan Press. Chapter 5.1-5.3.

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks, CA: Sage. Chapter 3.

We will take a closer look at models for dichotomous dependent variables. Please take a close look at Chapter 5.1-5.3 of King (1989) and Chapter 3 in Long (1997). Also skim

the Esarey/Pierce (2012) as well as the "separation plot" paper of Greenhill et al. (2011) for new strategies of how you could evaluate your model in terms of model fit.

Homework 3 is due.

Week 7 (26 March 2014): Interpretation and Simulation

> Hanmer, Michael J., and Kerem Ozan Kalkan. 2013. "Behind the Curve: Clarifying the Best Approach to Calculating Predicted Probabilities and Marginal Effects from Limited Dependent Variable Models." *American Journal of Political Science* 57(1), 263–277.

> King, Gary, Michael Tomz and Jason Wittenberg. 2000 "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44(2): 347–361.

Please read closely King et al (2000). If you have read it before - read it again! I read it several times myself after I digested it first (it was a prominent working paper at that time). This piece is definitely on the "Top 10" list of papers every MA student has to digest. Also take a look at the Hanmer paper to understand the difference between average-case and observed-value approaches. Which one do you prefer?

Homework 4 will be assigned.

Week 8 (2 April 2014): Ordered Choice Models & How to write a publishable Paper

> Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks, CA: Sage. Chapter 5.

In the first part we will discuss ordered choice models. Focus on Long's chapter 5. In the second party I will provide you with some strategies on how you can write a publishable paper in this class.

Homework 4 is due.

Week 9 (9 April 2014): Multinomial Choice Models

> Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables.* Thousand Oaks, CA: Sage. Chapter 6.

We will cover "multinomial choice models". Please take a look at Long's Chapter 6.

Easter Recess: No class on 16 & 23 April 2014!

Week 10 (30 April 2014): Conditional Logit Model

> Alvarez, R. Michael, and Jonathan Nagler. 1998 "When Politics and Models Collide: Estimating Models of Multiparty Elections." *American Journal of Political Science* 42(1): 55-96.

> Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage. Chapter 6.

We will discuss conditional logit models. Please closely read Alvarez and Nagler's 1998 AJPS piece and re-read Long's Chapter 6.

Homework 5 will be assigned.

Week 11 (7 May 2014): Selection Bias and Multi-Equation Models

> Dellmuth, Lisa Maria, and Michael F. Stoffel. 2012 "Distributive politics and intergovernmental transfers: The local allocation of European Union structural funds." *European Union Politics* 13(3): 413-433.

> Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage. Chapter 7.

> Timpone, Richard J. 1998. "Structure, Behavior, and Voter Turnout in the United States". *American Political Science Review* 92(1): 145-158.

We will discuss selection bias model, particularly tobit and heckman models. For tobit models please closely read Long's chapter 7 and for a nice application take a look at the recent award-winning Dellmuth and Stoffel paper. In order to better understand the Heckman model browse through Rich Timpone's APSR piece.

Homework 5 is due and Homework 6 will be assigned.

Week 12 (14 May 2014): Multi-level Models

> Gelman, Andrew, and Jennifer Hill. 2006. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. New York: Cambridge University Press. Chapter 11–13

Make sure to closely read all three chapters.

Homework 6 is due.

Week 13 (21 May 2014): Baby Bayes - a primer

Martin, Andrew D. 2008. "Bayesian Analysis." In *Oxford Handbook of Political Methodology*, eds. Janet M. Box-Steffensmeier, Henry E. Brady, and David Collier. Oxford: Oxford University Press, 494–510.

Finally, we will cover some ground in terms of Bayesian Analysis. Please read the intro article by Andrew Martin.

Week 14 (28 May 2014): Student Presentations

In order to provide you with feedback on your final papers we will have short (5 min.) presentations of your hypothesis and the key results of your paper. Please email me your presentations until **8am** that day.