

Institut for Political Science—University of Zürich

INTRODUCTION TO QUANTITATIVE METHODS

Lucas Leemann

June 29.&30., 2010

Contact Information:

Lucas Leemann, Institut für Politikwissenschaft, Uni Zürich. Email:
l1eemann@gmail.com. Office hours: Daily, TBA.

Course Information:

Lecture times: Every day 9.30-12.00 and 14.00-16.00. Lecture location:
Oerlikon. Language of instruction: English or German (depends on stu-
dents), slides will be in English.

Overview

This course intends to provide an advanced introduction into quantitative methods. Students are expected to have had previous exposure to the material. The class will mainly cover the linear regression model and serve as a refresher. The second part is concerned with more complex models which arise frequently in many political science applications. We will cover these models (logit, probit, ordered binary models, multinomial logit) in less depth and more try to enable a general understanding. This second part will rely heavily on examples.

Goals

Students will learn the to use the linear regression model and be aware of possible problems. They will learn how to detect and test for possible shortcomings and what the consequences are. Finally, students will master the standard strategies to overcome such obstacles.

The second part of the class will enable students to recognize which model is necessary for which question. They will know the basic elements, such as interpretation, test theory, and problems. The exercises and examples will provide a rich pool of Stata code which will prove helpful later in their own research.

Books - Manuscripts

Although students are not expected to prepare for this class, I will frequently quote different sources.¹ The main book is a standard textbook, *Introductory Econometrics* by Jeffrey Wooldridge. I encourage everybody to buy a copy as it will prove helpful in the next couple of years. A slightly easier treatment of mostly the same materials is offered by Gujarati.

- **Books:**

Wooldridge, Jeffrey. 2003. *Introductory Econometrics: A Modern Approach*. USA: Thomson.

Gujarati, Damodar N. 2003. *Basic Econometrics*. Boston: McGraw-Hill.

- **Manuscript:**

Steenbergen, Marco R. 2008. *Discrete Choice Models for Political Analysis*. Bern: Department of Political Science, University of Bern. (pdf)

- **You might want to take a look at:**

Stata Web Books: Regression with Stata. University of California, Los Angeles. URL:
<http://www.ats.ucla.edu/stat/stata/webbooks/reg/default.htm>.

Wawro, Greg. 20XX. *Lecture notes and other stuff*. New York: Columbia University. URL:
<http://www.columbia.edu/~gjw10/w4912.html>.

- **Paper:**

Freitag, Markus. 2006. "Bowling the state back in: Political institutions and the creation of social capital." *European Journal of Political Research*. 45: 123–152.

¹The only exception is a paper by Freitag (2006) which I expect students to read before the first day of class. It shows a particular way of dealing with many of the problems which we will discuss. Although not every provided solution is ideal the paper is a great starting place.

PART I: Refresher Linear Regression

Class:

The linear regression model is the work horse of social science quantitative methodology. We will cover the following topics:

- Research Design (experiments, quasi-experiments, and observational data). Why do we need statistical methods and what implications does this have?
- The model and its assumptions, interpretation of marginal effects, model quality (R^2 , adjusted R^2 , and F-test)
- Test Theory I; what is a statistical test? How do we test joint hypotheses?
- Dummy variables; what changes when we introduce dummy variables? What happens when we interact dummy and explanatory variables?
- Test Theory II; How do we test if the difference of significant estimates is itself significant?

Lab:

Some examples in **Stata** with code and data provided to briefly illustrate the ideas of this first part.

Further Readings:

Wooldridge, Jeffrey. 2003. *Introductory Econometrics: A Modern Approach*. USA: Thomson. (Chapter 2, 7)

Gelman, Andrew and Hal Stern. 2006 “The Difference Between “Significant” and “Not Significant” is not Itself Statistically Significant” *The American Statistician* 60 (4): 328-331.

PART II: Potential Problems of the Linear Regression Model

Class:

In this part we will discuss common obstacles and their remedies. We will start with a brief description of each problem, how to detect it, and which strategies can be employed to overcome it. We will focus on issues arising when cross sectional data is used. Therefore, we will forego topics of time series analysis, such as autocorrelation and panel heteroscedasticity. We will focus on the unholy trinity of pitfalls; omitted variable bias, multicollinearity, and heteroscedasticity.

- Endogeneity
- Heteroscedasticity...when we make mistakes in the assumption of our errors...

- Omitted Variable Bias; is the model complete?
- Multicollinearity - ...or why more data is better

Lab:

We will replicate some of the findings of Freitag (2006). Thereby we will see certain aspects which leave room for debate. Code and data for **Stata** will be provided. This example will prove especially helpful since we have an array of possible explanatory variables while having only 24 observations to estimate our models on.

Further Readings:

Wooldridge, Jeffrey. 2003. *Introductory Econometrics: A Modern Approach*. USA: Thomson. (Chapter 2, 3)

Williams, Richards. 2010. *Multicollinearity*. Download available:
<http://www.nd.edu/~rwilliam/xsoc63993/l11.pdf>

PART III: Binary Models and other challenges

Class:

The linear regression is fine if we are interested in explaining an outcome which is theoretically continuous. But often the focus of political science questions lies on discrete outcomes. Examples of this are questions, whether a MP votes yes or no, if a democracy fails or not, for which party an individual votes, how important a person evaluates job security. All these are examples of questions which cannot always be answered by relying on the linear regression model.

In this part we focus on models which have an outcome variable that can only take on two values. We will see the model, gain an intuition of the statistical background, and learn to interpret its estimation results.

- LPM - a solution within the LRM framework
- Logit and Probit. Intuition and estimation - glance at MLE.
- How do we interpret these results?

Lab:

Some examples in **Stata** with code and data provided to illustrate the ideas of this third part. We will rely heavily on the add-on package provided by Scott Long.²

We will use data on female labor participation and also data of the passengers of the Titanic. This lab will cover step-by-step what we saw in the lecture.

²For those of you who will bring their own computers (highly encouraged), please start **Stata** and type `net search spost` then click on `spost9_ado` from <http://www.indiana.edu/~jslsoc/stata> and install it. If you get an error message that **Stata** is busy, make sure that your window is displaying all information and if necessary click on the `-more-` sign at the bottom.

Further Readings:

Steenbergen, Marco R. 2007. *Discrete Choice Models for Political Analysis*. Bern: Department of Political Science, University of Bern. (Chapter 2)

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage. (Chapter 3, 4)

IF WE HAVE ENOUGH TIME:

PART IV: Ordered Outcomes and Multinomial Models

Class:

In this second part of non-continuous outcome variables we are focussing on ordered and unordered outcomes. If you ask respondents to express how much they value or support something, their answers usually are on an ordered scale. Standard examples are how strongly someone supports a specific measure and usually we have three, five, or seven answer categories. If we would like to build a statistical model explaining those answers, we need a model which suits such outcome variables. The ordered logit or ordered probit model are possible solutions.

Another important category of models allows us to explain vote choice. In Switzerland we usually vote for one of several parties and these parties can not be ranked. The vote variable is therefore nominal and requires a different set of models. We will briefly discuss the multinomial logit model and its strong assumptions.

Steenbergen, Marco R. 2007. *Discrete Choice Models for Political Analysis*. Bern: Department of Political Science, University of Bern. (Chapter 2)

Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage. (Chapter 5, 6)

Lab:

If time permits we will replicate the results from the lecture and learn how Stata and especially the add on package SPOST facilitates interpretation of these models greatly.

Further Readings:

Steenbergen, Marco R. 2007. *Discrete Choice Models for Political Analysis*. Bern: Department of Political Science, University of Bern. (Chapter 2)

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