

Empirical project

ECON306 – Summer 2014

Description

Econometrics is a very applied discipline by nature. As such, studying the theory is not the best way to learn it. It is important to practice Econometrics in order to be able to grasp what we can learn from it, what we must pay attention to, and what the potential challenges that may arise in practice are. With this in mind, a big part of your grade for this class will be based on a term project.

You must form teams of 4-5 members and complete an empirical project. Each team must select a data set from the list below and identify a related Economic question. Then you must try to provide an answer to the initial question, based on the empirical findings resulting from the statistical analysis of the data. For this purpose you must read and follow the guidelines from Chapter 11 in the textbook. The analysis must involve at least one multiple regression *and* at least one hypothesis test. The selection of the model(s) and hypotheses must be well justified by a theoretical analysis (it is not required to use a mathematical model, but some Economic intuition must be provided). The project will be graded based on three outputs described ahead: a group report, a group presentation, and an individual executive summary.

Deadlines

You must form your teams and send me an e-mail by the *end of the first week*. Send me an email even if you have not agreed to join any group, in that case I will assign you myself. I expect you to choose a dataset and a general research question by the *end of the second week*. If you are having problems identifying an interesting problem or methodology, come see me before it is too late. The written reports will be due in class on *Monday, June 23th*, and the presentations will be held on the last two days of classes, *Thursday, June 26th* and *Friday, June 27th*.

Datasets

The companion website to Stock and Watson (2011) provides a collection of datasets corresponding to different research papers. If you have problems coming up with research ideas, you may read the corresponding papers, or the empirical exercises in Stock and Watson (2011). However *I strongly encourage you to try to come up with a research question and methodology on your own*. The datasets and the reference to the corresponding papers are available online at wps.aw.com/aw_stock_ie_3/178/45691/11696965.cw/index.html. You must choose one of the following datasets:

- College distance – Cross-sectional dataset containing information on years of education, proximity to universities and cost of tuition, as well as a number of demographic characteristics. The data is taken from a survey of high school students.
- Current population survey – A large random cross-sectional sample containing data on the characteristics of the national labor force in 1992 and 2008.
- Teacher ratings – Course evaluations and instructor characteristics (including a rating of physical appearance) for a number of courses from the University of Texas at Austin.
- Growth – Average growth rates from 1960 to 1995, and macroeconomic characteristics for a sample of 65 countries.
- Guns – Crime incidence, demographic characteristics and gun regulations by state from 1977 to 1999.
- Fertility – Demographic characteristics, family characteristics and number and gender of kids for a large sample of women in fertile ages.

- Names – Call backs to job applications by fake candidates with randomly selected names typically associated with different racial groups.

Grading

- **Report** [50%] – Due in class on Monday, June 23th
The report must be written in the format of a short research paper (no more than 10 pages, 1.5 spacing, 12pt font). The introduction must provide a background for the problem and describe the research question and methodology. You must describe the dataset and present the numerical results using adequate tables and graphs. Every figure and table must be discussed and referred to in the text. The paper must conclude indicating how the analysis of the data helps to answer the proposed question. Be careful not to exaggerate the conclusions that can be drawn from the data.
- **Executive Summary** [25%] – Due in class on Monday, June 23th
Pretend that you work for a very important and busy person (say President Obama), and you must write a short executive summary that will help him/her make an informed decision. Suppose that your boss has little expertise on the subject, and only has a few minutes to read your report between meetings. The executive summary must be short (no more than 1 page), and it must summarize the main findings and implications of your extended analysis, in a concise and assertive way.
- **Presentation** [25%] – Scheduled for the classes of June 21th and June 22th
During the last day of classes, each time will be allocated 10 minutes to present their project in front of the class. The team presentation should convey the main ideas of the paper to an audience that is not necessarily familiar with the topic. You will be graded on your domain of the subject, as well as on your success in transmitting the main points of your project. You should thus be mindful of the contents of the presentation, as well as your delivery.

Computer software

In order to analyze the data you will have to use a computer package capable of performing regression analysis. There is a great variety of software available.

- You may be able to do the project using MS Excel. However, the analysis may be easier if you use more specialized software.
- The leading packages for statistical analysis in Economics are Stata and Eviews. Learning how to use them might be useful for more advanced courses. I will cover some basic Stata commands in class. You may find the following tutorials by Stock and Watson useful tiny.cc/STATA, tiny.cc/EVIEWS.
- Other good programs for statistical analysis are R (open source), SPSS (graphical interface build upon R) and Minitab (developed by Pen State grad students). If you are familiar with Matlab, you can use the statistical toolbox. Finally, a new popular alternative is Python (open source). I have never used it, but Python has a reputation of being very powerful, versatile, and easy to learn.

The only package I will discuss in class is Stata, and the only other package I'm versatile with is Matlab (if you choose to use a different package my ability to assist you will be limited). But you can choose to use whichever software you prefer. All of these packages are available in the computer labs. The datasets are available both in stata and excel formats. The excel format can easily be exported into ASCII (a normal unformatted text file), which can be imported by most software.