Syllabus for the PhD course:
Advanced Econometrics IV
("Econometrics of non experimental data")
(A.Y. 2020/2021)

Summary and purpose

Most of the statistical and econometric methods, which constitute core curriculum of Finance and Economics PhDs, are grounded in standard sampling theory statistics.
We have a "population" endowed with a given probability distribution for a vector of random variables. We want to estimate properties of this distribution.
In order to estimate these properties we use a sample from the population.

The basic assumptions on sampling are:

1) The observed dataset and any future dataset on which further estimates shall be based or for which forecasts shall be required are random samples were unit are sampled with known, albeit not necessarily equal, probabilities, out of a given population
2) Unit sampling may be a dependent process, however its structure is completely known
3) At least in principle we can sample any number of time without limits

These assumptions are usually violated in applications of statistics in social sciences and, more in general, in fields where most information comes from observational data (not only Economics but also, e.g., Epidemiology, Demography and Astrophysics).
In particular, the sample may be probabilistic in nature but the probability of observing a given unit in the sample is not known (on this specific point is based the classic definition of "observational data": Cochran, W. G. (1965), "The Planning of Observational Studies of Human Populations," Journal of the Royal Statistical Society, Series A (General), Vol. 128(2): 234–266).
The problem is, first, to understand how to read the results of standard models when data does not come from traditional random sampling and, second, how to modify, if necessary, the said models in order to take into account "selection biases" which may obfuscate the connection between the observed data and the population distribution.
"Selection biases" are possible interactions between the (unknown) observational process and the distribution of variables of interests (a kind of interaction, which, we can avoid in the standard setting by random sampling).
A main objective of the course shall be that of showing, with examples from the literature of several fields, that the existence of selection biases, far from being a nuisance only, is often a relevant component of the information contained in an observational dataset.
With non-experimental, observational, data the proper understanding of the (usually not fully specified) observational process is as relevant as the specification of the population model.
In fact, in most cases the observational process is such a relevant characteristic of the phenomenon under study that a successful description of the way we observe data can result in discoveries even more interesting that those connected with the population model.

Syllabus

- Introduction to the problem of non-sampling/non-experimental statistics. Its connection and its differences with standard sampling statistics. Different classes of problems related with the use of observational data. Intervention and forecast interpretation of statistical results ("causality" vs "forecast").

- The Heckman model. The basic problem: selection of data depending on properties of the phenomenon under analysis. Conditions on the dataset in order to apply the Heckman "correction". Summary of the technical points of the model.
- Censoring and truncation. Definitions. Examples of censoring and truncation. When censoring and truncation are not relevant. When they are relevant and how to deal with them. Corrections for censoring and truncation in a regression model (e.g. Tobit model). The CRSP dataset.
- The equity premium puzzle. Introduction, why it is a puzzle. Possible explanations. Explanation based on survivorship bias. Survivorship bias as an example of censoring/truncation.
- The problem of multiple testing with dependent tests. Why is it common in economics and finance. Why is it necessary to redefine the p-value at a "procedure" level when the procedure consists in running many test. How to deal with dependent tests. How to measure the statistical properties of estimates based on previous tests (pretesting bias). Discussion of H. Campbell et alii paper "…And the cross section of expected returns" about the literature which tries to discover "relevant" factors explaining the different mean returns of stocks.
- Meta-analysis as a quantitative review method. Definition of meta-analysis, what is its purpose. A simple summary of the steps constituting a meta-analysis, with some detail on meta-regression. Summary of one example of meta-analysis. Publication bias and the funnel plot.

Exam
Written exam closed books.
The day of the exam three of the six sections of the syllabus shall be randomly drawn.
Each student shall choose two of the three sections and write a summary of each based on the study materials.

Bibliography

-Paul Rosenbaum "Observational studies" Springer 2002
-A choice of papers and handouts presented during the lectures