

Alternative data sources for research in economics

Syllabus

Èric Roca Fernández

Course's objective

This PhD course revisits recent empirical contributions from a very particular perspective: how the use of alternative data sources may be useful to answer relevant questions in economics. In general, the use of readily available data-sets (for instance INSEE surveys, Banc de France databases or data compilations from other disciplines) eases the burden on data-collection. However, focusing on this data limits the type of questions that can be answered. In contrast, cutting-edge research answering specific questions is turning more and more to using very specific and hand-collected databases that allow researchers to focus on very particular questions that, otherwise, will be hard (or impossible) to answer.

This PhD course has two major objectives:

- a) Analyse some recent uses of alternative data and discuss why the research questions would have been more complicated to answer if traditional data had been used, thus enabling and encouraging PhD students to use such data.
- b) Discuss how to find the relevant data and how to process it, with an emphasis on replicability.

Outline

Introductory session:

Course organization and allocation of the papers to be discussed.

Section 1: Research papers using novel data

In this first part of the course, we shall review recent papers that use alternative data. It will be similar to a reading group, with one student will presenting the data section of a paper and discussing it, answering questions from other students. See "Other References" below for potential papers to be discussed and the topic they covered.

Section 2: Finding data online

Different possibilities to find data online, download it and format it. Focus on documenting the process (e.g. using Python) to facilitate replicability. Examples of data sources:

APIs: Commuting time between cities using the Google Maps API.
Open databases: ex: SPARQL to retrieve information from Wikipedia.
Web-scraping: ex: automating the download of information.

Course materials

During the first part of the course, we will read different research articles, available online. For the second part, the relevant materials will be available on Ametice.

Grading

Students will be graded on the basis of their discussion of the papers (1/3 of the final grade) and on a project (2/3 of the final grade).

The project consists in replicating the data-gathering process and basic results of a research paper. The evaluation will consider the efficiency with which the data is retrieved and processed.

Additional information

Although it is not a formal requisite, students should be familiar with the Python programming language.

Other references

Labour economics:

Arnaud Cheron, Bruno Decreuse, Matching with Phantoms, *The Review of Economic Studies*, Volume 84, Issue 3, July 2017, Pages 1041–1070.

Gender economics:

Alberto Alesina, Paola Giuliano, Nathan Nunn, On the Origins of Gender Roles: Women and the Plough, *The Quarterly Journal of Economics*, Volume 128, Issue 2, May 2013, Pages 469–530.

Comparative development:

Ashraf, Quamrul, and Oded Galor. 2013. "The 'Out of Africa' Hypothesis, Human Genetic Diversity, and Comparative Economic Development." *American Economic Review*, 103 (1): 1-46.

Trade:

Gojko Barjamovic, Thomas Chaney, Kerem Coşar, Ali Hortaçsu, Trade, Merchants, and the Lost Cities of the Bronze Age, *The Quarterly Journal of Economics*, Volume 134, Issue 3, August 2019, Pages 1455–1503.

Migration:

de la Croix, D, Docquier, F, Fabre, A and Stelter, R. 2020. 'The Academic Market and the Rise of Universities in Medieval and Early Modern Europe (1000-1800)'. London, Centre for Economic Policy Research.