

# ECON 435, Term paper: Your own empirical project

December 5, 2016

## The assignment

The goal for this assignment is to replicate the empirical results of a paper that you select.<sup>1</sup> I expect you to select a paper, obtain the data, and replicate the findings in the selected paper. I also need you to summarize the main idea behind the paper in your own words, focussing on the empirical strategy and the interpretation of the results.

## Organization

1. **Deadline** to have your paper **approved**: see course website.. That is not the deadline for submitting it, but for me to **approve** it.
2. **Deadline** to **hand in** your assignment: see course website.
3. Hand in (on Canvas), four files that allow me to completely replicate the analysis:
  - (a) a .zip with the data files (or a single data file, if there is only one that you load in your .Rmd);
  - (b) a pdf of the paper you are replicating;
  - (c) the .Rmd, and
  - (d) the resulting .html.

## Plagiarism

- Do not write down anything you do not understand
- Anything that you copy or paraphrased **must** be attributed to the original source. I take this very, very seriously. Exceptions: our textbook or my notes.
- Refer to **other people's code** as well, if you use it. This could be other students, an online source, or ...
- You are not allowed to use paid assistance, like tutors or proofreaders.
- When in doubt, contact me!

---

<sup>1</sup>If you are passionate about conducting your own research instead of replicating the research done by others, you are welcome to come and pitch your research idea to me.

## Conditions on the selected paper

The paper you are going to replicate must be approved by me by the deadline above. When I decide whether the paper you chose is OK, I take into account the minimal requirements below. I will also use my forecast of the feasibility of the replication (I may judge that it is too easy or too difficult: a possible outcome is that you replicate two papers, or only half a paper.)

- Paper must be published in a top journal in economics after after 2005
  - top journals include: American Economic Review, American Economic Journal: Applied, American Economic Journal: Policy, Review of Economics and Statistics, Quarterly Journal of Economics, Journal of Political Economy, Journal of Applied Econometrics, ...
- The methods must be aimed at uncovering a **causal** effect rather than a correlation. In this course, we do not care about correlations, nor about predictions.
- The data set must have at least 100 observations, and should be a panel or a cross-section. A time series analysis is not admissible as a term paper.
- You must convince me that the data is **available**.
- The methods used in the paper must be **more advanced** than OLS or a comparison of means. The panel data and instrumental variables methods discussed in the course are OK, and anything more advanced is also OK. Program evaluation methods (differences in differences, regression discontinuity design, propensity score) are usually OK. Check with me if you are unsure.

If you decide to conduct your own research instead of doing a replication, a different set of requirements applies. Talk to me.

## Topic

A great way to go is to find existing research that you are interested in, and then go from there. A replication is painful, and that pain will be reduced if you care about the subject matter. The paper you choose has to be empirical in nature, and needs to satisfy the constraints in the previous section . Where do you go and find replicable paper? The following list provides you some starting points:

- Look at the websites of high-quality economic journals that publish empirical papers and require the authors to make the data available
  - A great source of ideas is the “American Economic Journal: Applied”. This journal publishes high-quality papers that often use techniques related to the techniques in this course. Almost all papers have data available. See: <http://www.aeaweb.org/aej/app/index.php>.
  - A similar great source is American Economic Journal: Policy
  - As well as the American Economic Review
  - Journal of Applied Econometrics
  - Review of Economics and Statistics
  - etc.

- Look at the websites of prominent economists.(Acemoglu, Angrist’s data archive)
- Data repositories can provide a good starting point, for example:
  - Harvard’s dataverse
  - For an overview, see this [overview] at the Open Access Directory (Economics is under Social Sciences)
  - Another good starting point is [here], although tracking down a paper that uses a given data set may be challenging
  - Microdata for development economics: [devecondata]
- Some popular books on economics may provide a useful starting point
  - Freakonomics, and Superfreakonomics (or on their Wikipedia pages, where the questions are summarized): e.g. relationship between abortion and crime; parenting and education;
  - Supercrunchers (by Ian Ayres): e.g. how to explain wine prices (suitable for ideas, not for replicable papers);
  - Poor Economics
  - “Mostly Harmless Econometrics” and “Mastering Metrics” (a lot of the data used there is available online, see e.g. Angrist’s website)
- newspapers and magazines (The Economist, Significance magazine)
- blogs;
- an unresolved issue from a different course;

## Format

What I care about is receiving a quality Markdown document with interesting content. There is no minimum/maximum length, but if you want some guidance: 5 pages<sup>2</sup> excluding references, figures, and tables. Explain everything in your own words. Do not write down anything that you do not understand.

Grades will be assigned in relative terms: I will rank your term papers in terms of their quality. Top of that list gets full points, bottom of that list gets half points, and other papers are judged relative to those. If you fail to meet the minimum criteria set forth in this guide, you can receive less than 50%.

### 1. introduction: 0.5 page. **5 points**

- state the question;
- convince me that it is important/interesting;

### 2. data description: 1 page. **5 points**

- describe how you obtained the data;

---

<sup>2</sup>This may not seem like a lot, but the data gathering, processing, and trying of several models, takes a lot of time and only results in a couple of tables.

- describe why this data is relevant for the question you posed in (1)
- report descriptive statistics using tables and figures
- ideally, you replicate the descriptive statistics table in the paper

3. model and results: OLS: 1 pages. **10 points**

- write down the model underlying an OLS in your application, including the assumptions on the error terms, etc.
- present the OLS in a table
- interpret the results
- what is wrong with doing OLS in this context: frame your answer in the context of this application

4. model and results: advanced: 2 pages. **15 points**

- how are you going to solve the problem in the previous bullet point?
- write down the model(s) associated with your solution(s)
- what is the method associated with the previous bullet
- report the result that uses the new solution
- interpret your estimates
- relate them back to the original research question
- attempt to make a strong case that some of your estimates reflect causal effects<sup>3</sup>.

5. conclusion: 0.5 page. **5 points**

- summarize approach and findings.
- discuss strengths and weaknesses.
- attack the causal interpretation in (3)
- suggest avenues for further research, or implications for policy

Some more points about the way I construct your scores:

- You can earn additional points (up to 10) by **extending** the analysis in the paper in a meaningful way. The analysis must be accompanied by a motivation, and an interpretation of the findings.
- You will be penalized if your R code does not work.
- You will be penalized heavily if your analysis and discussion is not centered around causal effects.

---

<sup>3</sup>If you are critical: try to convince me even if you do not believe it. If it makes you feel better, you can destroy your argument in a subsequent paragraph.