

Introduction to Econometrics Final Project

Dr. Soliman
Clemson University
ECON 4050

Guidelines and Advice

This document is intended to clarify what is expected for the final project of Introduction to Econometrics, and also to provide some brief advice on how to approach the paper. As this is the final evaluation for this class the primary aim is to test your knowledge of the material covered in class, and your ability to apply the techniques you have learnt. While credit will be given for original ideas and relevant techniques (used properly) that were not covered in class, it is necessary (and sufficient to gain a good grade) that your paper demonstrate a correct application of one or more of the techniques seen in class and a discussion of any limitations relevant to the context in which you apply the technique.

1. Guidelines

a. Format

Using [R Markdown](#) to write your paper (as you did for the midterm). R Markdown is a document format (and R package) specifically designed for integrating code, text, and figures into a single document (PDF).

b. Length

I do not expect a long paper, i.e. more than 8 pages of pure text, nor something too short that does not allow you to develop your arguments, i.e. less than 3 pages of text.

A good paper would be around 5 pages of pure text. Of course, you will be inserting figures, tables and code making the overall paper longer than 5 pages.

c. Grading

The project is worth 200 points. The Content section below has the number of points in parentheses for each required section of the project.

d. Content

You should have a separate section in your project for each of the four sections below.

i. Literature Review (30 points)

You should use previous work by academic economists, sociologists and other social scientists to inform your study: the underlying process that generated the data; which variables to include etc. Briefly describe the work that has informed your paper.

A comprehensive literature review is not expected and is unnecessary. This section should just demonstrate that you have read a few papers on the topic and have an idea of what are the important implications of the literature.

ii. Data (30 points)

Describe the data, including where you found it, what are its strengths and limitations (discuss its suitability for your research). This section should also present some key features of the data and demonstrate that you know your data. Note it is rare to find the perfect data, so there will likely be limitations.

iii. Methodology (30 points)

Describe (and motivate) the econometric techniques that you use to analyze your data. Remember that the aim of this paper is to evaluate your understanding of the course. Using methods not taught is fine (if used correctly) but it is **necessary and sufficient to use techniques that were demonstrated in class**.

You should aim at describing your method so that it is **replicable**. Note this does not mean describing everything you did, especially things you do not discuss. Instead, aim to give enough information so that someone could reproduce your key results.

iv. Results and Analysis (110 points)

Do not just report your results; discuss them in the context of the limitations of your method and your data (and any other relevant factors). Your paper should address a clear and well-defined question. Not only does this aid the reader (us!) in understanding your aims and findings, but it will also help you to design your study. Make sure you demonstrate that you are able to precisely interpret your results.

- **Things to keep in mind for this section:**
 - **Use different types of visualizations (i.e., histograms, bar charts, line graphs, etc.) and make sure the charts are labeled clearly so that the reader can understand them without reading the whole project first.**
 - **Use some form of linear regression analysis.**

v. Code

Your paper should include all the code you used to produce the results in the paper (including any plots/tables and any cleaning of the data). You do not need to (and should not) include code to produce output that did not make it into the paper. You will be using RMarkdown for this, which should make it easy.

2. Advice

a. Choosing a topic

While the ideal is to choose a topic that is both original and interests you, your primary aim for this paper (especially if it is late in the semester and you have not started) is to choose a topic for which you can find suitable data. Remember that you also need to be able to demonstrate techniques from class, so make sure both the topic and the data allow this.

b. Finding data

There are numerous places to find data, a task made infinitely easier by the internet. A good place to start is the relevant government department of the country you wish to investigate (or the national statistics agency). The American Economic Association [website](#) has a page dedicated to data for economists. Empirical papers will also point you towards useful sources. Lastly, the [Harvard dataverse](#) is an excellent resource.

c. Data analysis

You should use R to carry out all data cleaning and analysis, and it is vital that you keep a record of what you have done. This does not mean just saving your regression outputs. You should save all commands you type in a commented R-script, or ideally write your report using R Markdown. You can also regularly save your history in RStudio, though this will only keep a record of what you type, and not any outputs. This will ensure you can replicate your results easily, and if you find a mistake or want to change something, will greatly simplify finding the relevant code.

d. Writing

Aim to be as concise as possible. Do not list everything you read and all the regressions you carried out. Include just what is necessary to understand what you did and how you made any decisions that were necessary.

Good luck!

Note: it is still possible to write an excellent paper without following any of the advice in this document. Please hand in a PDF though ;)