

Econ 375
Applied Econometrics
Early winter 2014 AD
Project Guidelines

KEY DATES:

PRESENTATIONS - THURSDAY 4 DEC 2014 IN LAB

PAPER AND DOCUMENTATION MATERIAL DUE - FRIDAY 12 DEC 16:30

The purpose of the project is for you to demonstrate to me your understanding of the basic principles of empirical research. Projects will be completed in groups of two to three or individually. This is appropriate since much modern empirical work is coauthored.

You will formulate a question about an economic issue based upon economic theory and use data to investigate and (possibly) answer your question. This should involve all of the steps of empirical research we have discussed. I am intentionally leaving some room for you to formulate an interesting question. You will collect the data you need to answer your question. Michael Haines and Peter Rogers gave us several ideas for data sources that you can look at. And Peter Rogers at the library is happy to help you find what you are looking for, within the limits of his time. However, stick to datasets that are already well constructed. One can get bogged down in cleaning data and be lost forever to the human world...

The project consists of three components:

- A presentation of your research to the class
- A written paper that will be delivered in hard copy to my mailbox
- A set of electronic documentation (described below) to be deposited in the class project folder on the **econdata** server

PRESENTATION

Each team or individual will present their findings to their peers. This will be done during lab section on Thursday 4 December 2014 AD. The purpose of this exercise is that:

- Students will have to defend what they have done, assumptions they have made, etc.
- Peers can offer suggestions on directions to take and how to improve the paper.

Overall I hope that this exercise will improve the quality of the final papers.

The other students present will be expected to be participating and providing questions and comments about the work being presented. I will take notes on your involvement and it will be counted as part of your presentation grade.

Your presentations should include:

- Motivation for your project. Why do we care? Why is this an interesting or important question?
- Brief mention of previous work in this area. What did others find? Do not dwell on this or allow yourself to get dragged into detailed discussion of the work of others.
- Model specification: What is the form of your estimation model and why did you choose it? You do not need to show all specification tests, but be prepared to defend your specification if challenged. Did you consider how your data is structured in the three dimensions DMH?
- Results: present your results in a clear, easy to read table. Make it clear what we are supposed to see and why. Visuals are also always good. A graph is worth 1000 numbers. However, remember that a graph should always make the data EASIER to interpret...

- Conclusions: what do your results say about your theoretical question? How strong is the evidence? What are the limitations of what you have done?

Participation in the presentation of others

The point of the presentations is for you to act as colleagues and help others to improve their work. Therefore, you should not be passive during the presentations of others. Challenge them. If you are not convinced by their arguments, explain why and offer suggestions for improvement. Of course, this should be done in a constructive manner. No heckling.

Some things to think about as you watch your colleagues present:

- Is it clear how this work fits into the existing literature? Does it contrast with or extend previous results?
- Does the model seem like a reasonable specification? If you are skeptical, challenge the person on whether they have considered this.
- Results: is it clear what you are supposed to see and why? If something is presented that is not clear, ask. Is there anything about the results that is surprising or seems counterintuitive? If so, ask for an explanation.
- Conclusions: do these results relate to the theoretical question? Are the variables in the data the same ones as the theory discusses? How strong is the evidence? Are there other avenues for research that might shed more light on the issue?

PAPER

Final papers will be due on the last day of class 12 December 2014 AD by the end of the day (16:30). The week in between the presentations and the final due date should be used to incorporate the feedback you get during the presentations. I will take notes on suggestions and expect to see them in the final work.

Your paper must include the following (though this does not mean that you need to have your paper written in sections with these titles):

- Abstract: A short paragraph explaining what your paper does and the main results. You need not go into details, but give a flavor of the results you have found. This is an excellent exercise, since it makes you think very carefully about what you are actually trying to tell us in the paper.
- Below the abstract on the title page, list two or three JEL codes (use 2-digit codes) that relate to your literature. Codes can be found here: http://www.aeaweb.org/jel/jel_class_system.php Also list a few keywords that relate to your work. These are used when people search for your work.
- Introduction
 - Discuss the economic situation you are investigating. Explain to us why we should be interested, and tell us the results you have found.
 - After reading the introduction, it should be clear to the reader what you have done in the paper, in what way this contributes to previous literature (broadly, the specifics can be discussed in the lit review), and the main results you have found. Then, if the reader is still interested, he can move on to see the details in the rest of the paper.
- Literature review
 - You will discuss AT LEAST three papers relevant to your issue. your own investigation will be informed by this previous work. These papers should provide you with a model structure, ideas of what variables are important, etc. Papers must be cited in proper economic style. You can learn this by looking at how things are cited in the papers you read.

- Good sources for literature are:
 - * EconLit (http://cu-cel.org/licensedb/licenses/index_alldb.php?titlecode=e)
 - * Google Scholar (<http://scholar.google.com/>)
 - * RePEc (<http://econpapers.repec.org/>)
 - * Social Science Research Network (SSRN) (<http://www.ssrn.com/>)
- You can then trace literature by looking at sources cited by these papers. Stick with papers that are at a reasonable level for you to understand, at least the idea behind it. You can always do a simplified version of a model if someone did something really fancy.
- REMEMBER: The point of the lit review is to put what you are doing in context of what has been done before, not to review everything that has been written about the topic since the big bang (or Creation... up to you).
- Data
 - Describe the data that you are using, where it came from, and why it is the right data to use to answer your question.
 - Thou shalt summarize the properties of thy data to show that thou knowest them. Tables of summary stats are good to have. If you have time series, it is often more useful to show a graph of the variable over time. DO NOT INCLUDE HISTOGRAMS AND SCATTERPLOTS OF ALL YOUR VARIABLES HERE, THIS WILL GO INTO THE DATA APPENDIX.
 - Use this data to form an estimable model relevant to your theoretical question. Sometimes, you may want a separate section describing the model. If the model is very simple, this may not be necessary. Use your judgment.
- Results and Inference
 - Report your estimation results. Results should be organized in an easy-to-read table such as those you have (hopefully) seen in your literature review.
 - Use your results to make inference related to your theoretical question.
 - Relate your results to those of the literature you read in preparation. If your results disagree, discuss why. If they agree, discuss what you have learned that is new.
- References (proper citations)
 - In proper economic journal style (APA) form.
 - All works must be cited properly within the text in economic style. We use the last name and year. No footnotes or any of that nonsense. For the most part, we use APA style: <http://www.apastyle.org/>

ELECTRONIC DOCUMENTATION

The main electronic files that you should turn in are of three types: data files, do-files, and a read-me file. In addition, you should submit an electronic version of the printed paper (described above) that you turn in. (Please note that the electronic version of the printed paper that you turn in with your electronic files is not a substitute for the printed paper itself. As described below, you must turn in a printed copy of the paper at my office.)

Data files:

1. Raw data file or files. You should turn in all of the data you use, in the format in which you initially downloaded it. You should add a note to each raw data file indicating the source from which you obtained the data, and the date on which you obtained it. You should specify the source in enough detail that someone else could find and download the same data and come up with the exact same numbers you have. Other than adding a note with the source of the data and the date you downloaded it, you should not make any other changes to the raw data. If

all of the data you used came in a single file, then you will have just one raw data file. But if you downloaded several different raw data files, you should keep and turn in a copy of each of the originally downloaded files (each with an embedded note about the source and the date of download). Even if you downloaded several files from the same source, if you downloaded them separately you should keep separate copies of all the files you downloaded, in their original formats. Give every original data file you turn in a name that has raw in it; e.g., raw_income.xls or raw_census.htm.

2. Importable versions of all raw data files. If any of the raw data files you downloaded are in Stata format, or a format that can be imported to Stata using the `insheet` command (e.g., tab-delimited text), you do not need to create or turn in any other versions of those data files. For every raw data file that is not in Stata format or a format that can be imported using the `insheet` command, you need to create a new version in a format that can be imported to Stata using the `insheet` command. To create these files, you will have to delete everything in the original data files except the data you want to import, and perhaps a first row with names for the variables in the file. These files should be saved as tab-delimited text or comma-separated values, with names such as `importable_income.txt` or `importable_census.csv`. When you create the importable versions of your data files, you should only change the original data files in ways necessary to make it possible to import them into Stata using the `insheet` command. Any other changes you want to make to the data should be done in Stata (and the commands that make the changes should be included in the do-files you turn in, as described below).
3. Final data files used in the production of the results

Do-files

The number, names, content and organization of your do-files will depend upon the number and structure of the files from which you obtain your data, and the steps of data cleaning and management required to create a dataset in the final form you want for your analysis. The instructions below describe the do-files that would be required for many typical cases. If some other way of organizing the do-files would be more appropriate or convenient for your project, you should make appropriate adaptations to these instructions. Here, I refer to STATA but if you are using SAS, then adjust accordingly.

1. One file with a name like `import.do` should contain commands that import each of your importable data files and save the data in new Stata-formatted files (with `.dta` suffixes).
2. One or more files with names like `cleaning.do` or `merging.do` that take the data in the Stata-formatted files created by `import.do` and create the final version of the dataset that you use for your analysis. The last of these do-files should end with a command to save your final dataset in `.dta` format, with a name such as `final.dta`.
3. One file with a name like `results.do` that begins by opening up your final dataset, and then generates each statistical result, table or figure that is reported in the paper. Before the command that generates each piece of output, put a comment in the do-file stating what piece of output the command is going to produce (e.g., The following command generates Figure 1, or The following command generates Table 3.) Note: It is important to include comments in all your do-files that explain the purpose of each command or group of commands in the do-file; e.g., importing the data, changing coding of the variables, creating new variables, taking logs to correct for skew, etc.

Data Appendix

The data appendix should contain an entry for every variable in your final data file or files. For each variable you should give the name that you are using for the variable. State the source from which the variable was taken. (This should be one of the sources described in the data overview. If all your data came from one source, you may indicate this fact just once at the beginning of the data appendix, and then not repeat it in the entry for every variable.) Indicate the number of observations of the variable

that are missing (out of the total number of observations in your data set), and give a brief description of what the variable measures.

Give the name or label for the variable that was used in the source from which you obtained it.

For every categorical variable, indicate the coding scheme for the variable. Present a frequency distribution for all the non-missing observations of the variable. If there is an ordering to the categories of the variables, be sure that they are listed in order in the frequency distribution, and include the cumulative percent frequencies (as well as the frequencies and percent frequencies). If there is no ordering to the categories, do not include the cumulative percent frequencies. Finally, present a bar graph showing the number of observations in each of the categories (unless for some reason that is infeasible, e.g., if the number of categories of a variable is very large).

For every quantitative variable, indicate how the variable is defined and what the units of measurement are. For example, if the variable measures income, is it the income of an individual, or the total income earned by everyone in the individual's household? Is it measured in dollars, pesos or yen? Corrected for inflation (and if so, what is the base year)? Pre-tax or post-tax? Including government transfers like social security or any form of welfare payment? Report the mean, median, standard deviation, minimum and maximum of the variable, and finally present a histogram.

In addition, your data appendix should contain results of specification tests such as RESET test, tests for heteroskedasticity, and whatever else is appropriate to demonstrate that your model is statistically adequate. **THIS IS NOT A CHECKLIST.** You do not need to perform every test we learn in class if it is not appropriate to your project. The point here is to show that, to the best of your ability, you have constructed a statistically adequate model that will provide reliable inference.

A read-me file

One file in .pdf format called *readme.pdf*. This file should list all the other files included in your electronic documentation, and explain how they are organized into folders and subfolders. You should give a brief explanation of what each file contains. Your read-me file should also give step-by-step instructions on how to use the files in your electronic documentation to replicate the results of your study. A sample read-me file for the electronic documentation that would accompany a hypothetical senior thesis (that we invented for purposes of illustration) is posted with this assignment. Please refer to that sample read-me file for guidance on constructing a read-me file for your project.

An electronic copy of the printed paper you turn in. Please create a single document containing an electronic version, in pdf format, of the final printed paper you turn in

Organizing your files

- Make a folder named Raw Data, and put all your raw data files in it.
- Make a folder named Importable Data, and put all your importable data files in it.
- Make a folder named Data, and put your Raw Data folder and your Importable Data folder in it.
- Make a folder named Do-files, and put all your do-files in it.
- Make a folder named Team X Final Project (where X is actually the last name of the person in your group whose last name comes first alphabetically). In that folder, put:
 - your Data folder (with the Raw Data and Importable Data folders inside it)
 - your Do-files folder
 - your readme.pdf file
 - the electronic version of your printed paper
 - the data appendix