

Problem Set #5

ECON 833, Prof. Jason DeBacker
Due Thursday, October 2, 1:15 p.m.

This problem set will have you apply the visualization and econometrics tools you've learned to data that is of interest to you.

PART A

Please use Python (or R) to create three visualizations of data related to a topic you would like to research further. These visualizations will be graded based on:

1. The information content communicated from the visualization.
2. The efficiency of the code used to generate the pictures (Did you read in data correctly and transform it with a minimal number of steps? Do you use functions/loops to save on redundant lines of code? Do you have proper documentation?)
3. The aesthetics of the visuals.

The images should be saved to disk into an `images` directory and will be compiled in the `.tex` file that will contain your complete problem set response.

PART B

Please use Python (or R) to estimate a baseline econometric model as well as two variations of this model (e.g., a panel data model with and without fixed effects, an IV model with a different instrument set). You may use an estimator for which there is a package or you may write your own. Your grade will be based on:

1. The relevance of the models you estimate to the economic question you would like to address.
2. Your ability to interpret the results.
3. The accuracy, efficiency, and reproducibility of your code.

The results of this estimation routine should be saved to a summary table in a `.tex` file that can be input into the `.tex` file that will contain your complete problem set response. You can use packages like 'stargazer' or create your own custom table (in Python, you might find dictionaries helpful, flexible containers for your estimation results. These can then be turned into DataFrames and the `.to_latex()` method can save the resulting table.)

When doing the above, I want you to make sure you follow good research practices:

1. Define functions to make your code modular. At a minimum, I want to see three functions in your code: (i) a function that cleans your data (sample selection, variable creation), (ii) a function that creates your visualizations, and (iii) a function that estimates your model. These may be in separate Python/R modules if you feel that adds to the readability of your code.
2. Write unit tests for your functions. Visualization functions may be tested by simply checking that a image object is returned.
3. Use docstrings to document all functions.
4. When reading/writing data from/to disk, please use **relative paths** and make sure you use Python/R functions that specify directories in a way that is compatible across operating systems (i.e., your code should work when anyone else runs it from your repository regardless of them using Windows/Mac/Linux).

5. Automate your process: Write a [bash script](#) to run your Python (or R) programs and compile your TeX file. A person

You will submit your problem set by pushing the files to your GitHub repository that you created from forking the repository for this class. You will place all files for the problem set in the path `/CompEcon_Fall2025/ProblemSets/ProblemSet5/` on your `ProblemSets` branch. These files will include:

1. At least one `*.py` or `*.R` script that creates your visualizations and estimates your model. Note that I want you to commit your data to your repo, or read it in from a URL (NOTE: Github has a maximum file size of 100MB. Do not try to commit a file larger than that.)
2. One `ProblemSet5_LastName.tex` file that contains your problem set summary. The file will read in the images and tables produced with the above. Also include the compiled PDF, but if you do this right, I should be able to recreate it. This file will include a description of the economic question of interest, a description of the data, the three images you produced for Part A, and a tabular or graphical presentation of the results from your econometric model in Part B.
3. One `*.sh` file that will run your Python/R script, update figures and tables, and compile a new pdf of your ProblemSet.

Your overall grade will be determined by the criteria noted in Parts A and B as well as your ability to automate the project in a way that I can replicate when I run your bash script on my computer (after having downloaded your `ProblemSets` branch).