Empirical assignment: Please from a group of at least three students to complete this empirical assignment. Feel free to use any programming language that you prefer. Here I will give you a few hints if you use Stata. Your submissions would be:

* Stata/R script(s) that produce the empirical report.
* One empirical report (up to 4 pages, excluding references, tables, and figures; use “insert Figure/Table. # here” to indicate positions for Tables and Figures while putting all figures and tables at the end of the report).
* Times New Roman, font size 12, double spacing
* The grade will be based 50% on your code writing and 50% on your written report.
  + Code will be evaluated based on the clarity ( such as indent, and comments) and whether it can generate all the figures and tables in your report.
  + Written report grades will be based on the correctness of your results and whether you clearly explain and interpret your results.

Files:

Paper: [cost of secrecy-Gross-MS2022.pdf](file:////courses/706/files/102347%3fwrap=1)

Original replication package: [replication\_package\_MS-ENI-20-00293.zip](file:////courses/706/files/102349%3fwrap=1)

Data: [data\_\_assignees.dta](file:////courses/706/files/102350%3fwrap=1)

In the attached files, you will find the papers and two datasets. Besides, you are given the original replication package offered by the author to replicate all empirical results in his paper. Feel free to explore his do files.

Read *Gross (2022)* and answer the following questions based on the two datasets mentioned.

1. Effect of compulsory secrecy on follow-up innovations by the same firms

Take the data\_\_assignee\_\_nber\_cat.dta dataset, and then limit the observations to years between 1930 and 1960 (years >= 1930 & year <= 1960).

* Briefly introduce the data structure (eg. What identifies one unique observations, how many unique companies are there, whether one variable has any missing values etc).
* Conduct two t-tests, comparing the average number of patents per time period and average number of patents with secrecy order respectively, by whether the company is a OSRD contractor or not in the sample. Report the results and explain your findings
* Hint: ttest patents, by(osrd\_contractor)).
  + Calculate total number of patent classes one firm patented per year, and present a binned scatter plot exploring how total number of patents with secrecy possibly affects total number of patent classes of a firm

- Hint:

- bysort assignee anypatent year: egen nclass = nvals(nber\_cat)

- keep assignee year nclass  secret\_patents if anypatent == 1

- duplicates drop

- binscatter nclass secret\_patents

(Please ponder upon why “bysort” should include “anypatent”, and feel free to explore alternative ways of operationalizing this question)

* Replicate the main regression results reported in Table 4, and more importantly, instead of using OLS estimators, please replicate the results using MLE estimators based on logistic models. Please report formatted tables in your empirical reports (instead of screenshots of Stata/R outputs).
* Hint:
* Note that in each column, the sample used is slightly different. Column (1) covers the most comprehensive sample, which could be defined via:
* global sample if self\_incat\_pre1939 >= 1 & self\_incat\_midwar>=1 & self\_incat\_secrate>0 & (year<=1939 | year>=1946)
* Logistic regression with controlling for company fixed effects and standard errors clustering at the unit of firms would be
* clogit any\_patent self\_incat\_secrate\_t\* self\_incat\_secrate period\* $sample, group(assignee\_cat) cluster(assignee)
  + Explain, according to the Stata/R routines, how many observations are dropped because they because of all positive or all negative outcomes and why these are dropped by your statistical package.

1. Effect of compulsory secrecy on follow-on innovation

Take the dataset data\_\_patent\_citations.dta, and answer the following questions.

* Plot the distribution of total number of forward citations received by whether patents has ever been ordered compulsory secrecy in the same figure, and report proper statistics to compare these two distributions. Explain your findings.
* Hint: plot kernel density figures using command “twoway kdensity”.
  + Replicate the results reported in Table 6 with replacing the dependent variable as total number of forward citations received and estimating the same equation using MLE estimator based on poisson regressions. Report your results, compare your results with Table 6 presented in the paper, and articulate your findings.
* Hint: dependent variable would be changed to fcites instead.
* Commond you can use for poisson regression can be:
* ppmlhdfe fcites secret secret\_19\* secrecy\_eval secrecy\_eval\_19\* , absorb(grant\_year class\_yr) cluster(uspto\_class)

1. Run any regression that interest you using the same dataset provided by the replication package of the paper. Explain your model specification and your results.

References:

Gross, Daniel P. "The Hidden Costs of Securing Innovation: The Manifold Impacts of Compulsory Invention Secrecy." *Management Science* (2022).