**BEM 106: Homework #2 Description**

**Goal:** The goal of this homework is twofold: (a) to introduce basic financial company data (albeit historical) and (b) to get you working with regressions in Python.

The excel file that accompanies this homework contains financial information for a large set of companies for the period 1980-2000s. The appendix lists the labels for each variable.

The main task of this homework is to investigate which variables explain (predict) financial leverage, i.e. the amount of debt that companies take on. The secondary task is to illustrate how standard errors relate to sampling.

Tasks:

1. Install the statsmodels package into Jupyter (if not already installed)
   1. You can find more information here: https://scikit-learn.org/stable/install.html
2. Import data
   1. Datafile name: hw2-dataset.xlsx
3. Create variables
   1. From “datadate”, create a new variable called “year” (as four-character integer, i.e. 1983)
   2. Using stock price and the number of stocks, compute the total dollar value of stock (product of stock price and number of shares) and create a new variable called “equity”
   3. Create a variable called “assets,” which is the sum of “equity” and long-term debt. This is the total value of the company. (Hint: long-term debt variable is in thousands of dollars.)
   4. Create a variable called “leverage” = long-term debt / “assets” – this is the extent to which a company uses debt to finance its operations. That is, higher leverage means more debt relative to company assets.
4. Questions about the sample
   1. How many unique companies are in the dataset?
   2. How many observations, on average, are there per company?
   3. Plot a histogram of leverage in 1990
5. Conditional expectation function and regression

You hypothesize that the “model of the world” (CEF) can be linearly approximated by the following:

Estimate this model using Ordinary Least Squares using heteroskedasticity-robust standard errors in four different ways:

1. No controls
2. Control variables: Year fixed effects
3. Control variables: Company fixed effects
4. Control variables: Company and Year fixed effects

Report and interpret the beta coefficients on log(capital expenditure) and log(operating income) and their standard errors under all four scenarios. Report the R-squared of each of the four regressions.

1. Finally, draw 1000 random samples of 100 companies (note: each sample should contain all years of the data for each company selected), and re-estimate regression on each sample. Do NOT copy/paste the code 1000 times – rather, use a loop to randomly select the sample, run the regression, and store your beta coefficients. Plot the distribution, separately, for each beta on log(capital expenditure) and log(operating income). How do the standard deviations of these distributions compare to your standard errors reported in Part (5)?

Appendix: Variable names and labels in HW2 Datafile

datadate – Date of observation

gvkey – Company identifier

capxv – Capital expenditure on property, plant and equipment (in thousands of dollars)

dltt – Long-term debt (in thousands of dollars)

dpc – Depreciation and amortization – cash flow (in thousands of dollars)

invt – Inventories (in thousands of dollars)

oancf – Operating activities – net cash flow (in thousands of dollars)

oiadp – Operating income after depreciation (in thousands of dollars)

ppeveb - Property, plant, and equipmrnt – End Balance (in thousands of dollars)

ppevbb - Property, plant, and equipmrnt – Beginning Balance (in thousands of dollars)

sppe – Sales of property, plant, and equipment (in thousands of dollars)

pstkrv – Preferred stock – redemption value

naics – North American Industry Classification Code

sic – Standard Industry Classification Code

prc – Stock price at the end of day on datadate

shrout – Number of stocks outstanding