Programming Exercises for Lecture 2

Submission deadline: **April-16-2024**

1. The data set of `CEOSAL2` contains information on chief executive officers for the U.S. corporations. The variable `salary` is the annual compensation (in thousands of dollars), and `ceoten` is the prior number of years as CEO of a company.

   a. Calculate average salary and average tenure in the sample.

   b. How many CEO's are in their first year as CEO (that is, `ceoten` $= 0$)? What is the longest tenure as a CEO?

   c. Estimate the simple linear regression model:

   $$\log(salary) = \beta_0 + \beta_1 ceoten + u,$$

   and report the results. What is the predicted percentage increase in salary given one more year as a CEO?

2. Generate data from uniform and normal distributions to complete the following questions.

   a. Randomly generate 500 x_i (independent variable) from $U(0, 10)$ (uniform distribution with range $[0, 10]$) with a function in `scipy.stats`. What are the sample mean and sample standard deviation of the x_i?

   b. Randomly generate 500 u_i (error term) from $N(0, 36)$ (normal distribution with mean zero and variance 36) with a function in `scipy.stats`. Does the sample average of the u_i exactly equal to zero? Why or why not? What is the sample standard deviation of the u_i?

   c. Now generate y_i (dependent variable) as:

   $$y_i = 1 + 2x_i + u_i.$$

   That is, the population intercept $\beta_0 = 1$ and population slop $\beta_1 = 2$. Use the data you generated to run a regression of y_i on x_i with the OLS. What are your estimates of β_0 and β_1? Do they equal to the population values in the above equation? Explain your answer.

   d. Obtain the OLS residuals $\hat{u}_i$ and verify that the properties:

   $$\sum_{i=1}^{n} \hat{u}_i = 0 \text{ and } \sum_{i=1}^{n} x_i \hat{u}_i = 0 \tag{1}$$

hold (subject to rounding error).

e. Compute the same quantities in equation (1) but use the errors u_i you generated in a. in place of the residuals $\hat{u}_i$. What do you conclude?

f. Repeat questions a. to e. with a new sample data. Now what do you obtain for the OLS estimates of β_0 and β_1 and quantities in (1)? Does your conclusion in question e. still hold?

3. Estimate the CAPM for Apple stock. Recall the Capital Asset Pricing Model (CAPM):

$$E\left[R_i - R_f\right] = \beta E\left[R_m - R_f\right],$$

which states risk premium of asset i is proportional to market risk premium. We are interested in estimating Beta (β) in the model. One way to do this is by running a linear regression of a stock's excess return $R_i - R_f$ on market portfolio's excess return $R_m - R_f$, i.e., estimating the following linear regression:

$$R_i - R_f = \alpha + \beta\left(R_m - R_f\right) + u. \tag{2}$$

Now the asset i is Apple stock and the market portfolio is the S&P500 index.

a. Import `data_exercise_2c.csv`. The data include monthly level of the S&P500 index (`SP500`), monthly close and adjusted close prices of Apple stock (`AAPL_close` and `AAPL_Adj_Close`), and monthly U.S. risk free interest rate (`RF_%`, in percentage), from April-2011 to December-2020.

b. Transfer dates in column `Date` to ISO-8601 format.

c. Calculate monthly simple returns of the S&P500 index and Apple stock (with both monthly close and adjusted close prices). Assign the simple returns in a new column called `SP500_sr` (S&P500), `AAPL_sr` (Apple close price) and `AAPL_Adj_sr` (Apple adjusted close price). Note that the first row in these new columns should be `nan`.

d. Calculate monthly excess returns of the S&P500 index and Apple stock (with both monthly close and adjusted close prices). Assign the excess returns in a new column called `MKT_rp` (S&P500), `AAPL_rp` (Apple close price) and `AAPL_Adj_rp` (Apple adjusted close price). Be careful about data of the monthly U.S. risk free interest rate (should be divided by 100 since it is in percentage). Calculate summary statistics of the excess returns with method `describe`.

e. Use the excess return calculated from monthly close price of Apple to estimate the CAPM with the linear regression model of (2) and print the estimation results. What are the estimates of α and β?

f. Use the excess return calculated from monthly adjusted close price of Apple to estimate the CAPM with the linear regression model of (2) and print the estimation results. What are your estimates of α and β?

g. Plot the data and fitted regression lines of both e. and f. The x-axis is for the excess return of the S&P500 index and y-axis is for either the risk premium of Apple from the close price or from the adjusted close price. The plotted figures should look like the following ones.