

Object Oriented Programming in Python

In completing the tasks specified below, you are expected to provide solutions through an object oriented programming design as much as you can.

1. Use Python to read the two csv files containing information of companies traded on NYSE and NASDAQ. Generate security objects which contain all information given in the files.
2. Create a set of companies consisting of those whose market capitalization are greater than US\$500 million based on your results in 1. Denote this set with A. Download the "Open", "High", "Low", "Close", "Adj Close", "Volume" of ALL securities in A from 2000-1-1 to 2017-11-17 as a Pandas data panel. Save the downloaded data into a file so it can be read by Python later.
3. For any given date T between 2000-1-1 and 2017-11-17, say, $T = 2017-4-6$, use data obtained in 2 to generate two pandas dataframes X and Y. Both X and Y have identical indexes which consists all the tickers in set A obtained in 2.
 - a. Y contains three columns: 1-day log-return, 5-day log-return and 22-day log-return. The n-day log-return of a security S on day T is computed as $r(n, T) = \log(P_T/P_{(T-n)})$ where P_T denotes the adjusted close price of S on day T.
 - b. X contains 6 columns: MA_5, MA_22, MA_200, PM_5, PM_22, PM_68. MA_m is the simple moving average of the adjusted close price over the past m-day window counting back from day (T-1). PM_n is $r(n, T-1)$ as defined in part a. X is a Nx6 matrix where N is the number of tickers in set A.
4. Use linear regression method in Pandas or Scikitlearn to fit linear models to explain stock returns over 1-day, 5-day and 22-day with factors contained in X, respectively. Namely,
For Z = each columns of Y
Fit $Z = \alpha + \beta * X$, where alpha is a scalar and beta is a 6x1 vector.

End of loop over 3 columns of Y

Discuss the statistical significance of the linear regressions coefficients based on the t-statistics value of the regression coefficients provided by the fitting package. Run this regression to get the beta vector daily for 60 days starting from 2 different dates: $T_1 = 2016-6-1$, $T_2 = 2017-7-1$. Compute the average of the 60 regression coefficient vectors and the standard deviations of each coefficient over the two respective periods and report the results using a table.