Homework #2

QMI First-Years

This is the first real-real homework assignment. There is going to be some assigned reading, associated question, code, and question. The expected completion time of this assignment is 2-3 hours. We know that you are busy, but we believe this is the best way to make sure you are retaining the knowledge from the previous week. We also highly recommend that you complete this assignment independently, if you need help of course reach out to your fellow first-years or us!

Finance/Statistic Questions

Read "Understanding Risk and Return, the CAPM, and the Fama-French Three-Factor Model". You can skim to the "CAPM" Section, skip the "Regression Analysis" to "Fama And French" Section, and read completely to the end.

1. A common critique of beta is that it is historical. Critics argue that because it is historical all of the information "included" in beta is already baked into the current stock price (because price(t) of a stock incorporates all publicly available information under EMH semi-strong form). What are some indications from our lecture as well as the reading that this might not completely be the case? Explain fully. [hint: see autocorrelation]
2. Offer a critique, in your own words, of the SMB factor. Provide a brief description of the factor, why it is wrong, and cite two sources not used in lecture or readings.
3. Define r-squared from the paper in your own words.

Read "Choosing Factors". Completely read the first nine pages up to the start of the "Marginal Contributions" Section and the "Conclusion" Section.

1. Why do we use Sharpe Ratio as an optimization function under the select factor models? [if you want further explanation of optimization this starts with a good definition: https://sites.math.washington.edu/~burke/crs/515/notes/nt_1.pdf]
2. What are the reasons for optimizing the Sharpe Ratio? What does it indicate when the Sharpe Ratio is "fully" optimized?
3. Define data dredging. Explain why it is problematic for model testing.

The previous paper uses the phrase, "multiple comparisons issue" (Fama & French 2017). Read this, article to learn more. No questions right now, but it is a very important concept for trading.

https://towardsdatascience.com/the-multiple-comparisons-problem-e5573e8b9578

Computer Science Questions/Problems

1. Google "the Euler Project" and solve the first two problems (IDs 1 and 2). Code this in Jupyter Notebook.
2. Creating a webscraper, requirements:
    a. Must use requests, BeautifulSoup, and Selenium with each other
    b. Choose your own website to scrap
    c. Code must run completely without fail
    d. Clearly comment what you are doing as you see necessary (this is down by inserting a hashtag before a line of text)
    e. The scraper should use, in a robust way, requests (easy one), BeautifulSoup, and Selenium
    f. Create a Pandas dataframe of the data that you collect from the webpage
        i. The target website cannot be overly structured (e.g. Texas Death Row Execution website)
        ii. Target should have some sort of meaningful data to collect, this threshold is low (we just do not want you creating a dataframe of every word in a paragraph)
    g. Acceptable Project: Collecting all of the names and prices of Lego City sets from their website (for this we will provide an example/walkthrough).

This homework is due Wednesday 20^{h}, 2021 at 11:59 p.m. Please submit as a PDF and direct message it to Alex on Slack.