MATH 60210: Homework #4

Professor Anthony Sanford

Due: April 15, 2024 by 11:55pm on ZoneCours

Instructions: For this assignment, you are to work in groups of **three**. I will **not** make the groups. It is your responsibility to find team members for your homework. If you have trouble finding a team, you can reach out to me via email and I will try to match you with other classmates. This assignment includes two deliverables: 1) your write-up and 2) your code. Failure to submit one of these **two** components will be considered an incomplete submission. Your code must by written in Python. Your code should run smoothly and be clearly annotated. If you used resources outside of what is assigned for this course, you **must** give credit to the source. For example, if you googled code and used whatever you found, reference the source. Otherwise, you are cheating. Your write-up is meant to analyze, discuss, and interpret your results. Simply submitting a table without clearly discussing that table is not an answer to the question. Consider what I do in class – I present something and discuss it. Think of your write-up in this exact way. You are interpreting your findings for someone who is reading your report. Recall that HEC Montréal has rules regarding plagiarism apply to both your written answers and your computer code.

There are 4 multi-part questions, points for each question/part is written below. For each part, there are four possible grades:

- E: This is the grade if you literally write nothing. Worth 0%
- C: This is the grade if you really don't understand what you're doing...but you wrote something. Worth 50%.
- B: This is the grade if you got the answer mostly correct. Worth 80%. This will likely be the most common grade.
- A: This is the grade if your answer is as good (or better!) than mine. Worth 100%.

Note: Comments in your code that allow me to easily understand what you were trying to do may improve your grade. Particularly ugly or inelegant python code may reduce your grade by 10% (e.g. if you're cutting and pasting a lot because you don't know how to use loops or functions or array operations effectively). You may also lose points if your code is not running correctly or does not provide the same solution as your write-up answer.

Problem 1 [60 Pts], Tests of Asset Pricing Models

a. Write a function retrieves data from Prof. Ken French's Data Library. (You could use the 'pandas-datareader()' or some other method for this.). The function requires no arguments and returns two objects [10 pts]:

- _factors_ : a _DataFrame_ containing six columns of _monthly_ data (from 1980 to 2021) in the following order
  - 'F-F_Research_Data_5_Factors_2x3'
  - 'F-F_Momentum_Factor'
- _portfolios_ : a _DataFrame_ containing
  - '17 Industry Portfolios'

Make sure that all the series in factors and portfolios are excess returns. **Report and interpret the results.**

b. Provide descriptive statistics and appropriate graphs for your data. Make sure to label your graphs correctly and discuss the results (both the descriptive statistics and the graphs).[10 pts]

c. Write a function that accepts [20 pts]:

- a DataFrame containing Factors
- a DataFrame containing excess returns on Assets (or portfolios of Assets)
- a boolean Traded indicating whether the Factors are excess returns on traded assets

and estimates a linear Asset Pricing model, printing a summary of the results and returning:

- a 1D array of the estimated risk premia for each factor
- a 1D array of their tstatistics (based on heteroscedasticityrobust standard errors)
- a 2D array of estimated factor loadings
- a 2D array of their tstatistics (based on heteroscedasticityrobust standard errors)
- a scalar containing the model's R^2
- a 3×1 array containing
  - the test statistic for the H_0 that the model adequately explains the differences in excess returns across portfolios
  - the test-statistic's p-value
  - the degrees of freedom for the test statistic's distribution under H_0

Report and interpret the results.

d. Use your results above to briefly (1-2 lines) answer each of the following questions.[20 pts]

- Which factors (if any) appear to earn a significant risk premium if we assume that the factors are excess returns on traded assets?
- Is the model well-specified, or do some portfolios command a premium that is not explained by the factors?
- Why do the critical values of the Jtest statistic depend on whether or not the factors are traded?
- For the factors that we used above, should we use the results from 'TradedFactorModel' or 'LinearFactorModel'?

Problem 2 [40 Pts], Maximum likelihood estimation

a. Write a function that accepts [10 pts]:

- a DataFrame df
- an integer $dof > 0$

and calculates:

- an OLS regression of the 1st column in *df* on a constant and the other columns
- the regression residuals and their estimated standard deviation.
- the log likelihood of the model under the assumption that the standardized residuals are distributed $t(dof)$.

and returns:

- the log likelihood llf.

Report and interpret the results.

b. Write a function that accepts [10 pts]:

- a DataFrame df
- an integer $dof > 0$

and then:

- uses 'OLSStudentt(df, dof)' to calculate the llf
- calculates the LR statistic testing the H_0 that all coefficients other than the constant = 0.

and returns:

- the likelihood ratio statistic LR
- its p-value pval

Report and interpret the results.

c. Write a function that accepts [10 pts]:

- a DataFrame df

and uses 'OLSStudentt(df, dof)' to create a Profile Likelihood plot for the degrees of freedom of the $t-$distribution:

- You should let $dof = e^{j/12}$ for $j = 0, \ldots, 50$

and returns:

- a 51×2 array with the values shown in the profile plot.

Report and interpret the results.

d. Based on the curve and data shown above, what would be the MLE estimate of the degrees of freedom for the tdistribution in this example? [10 pts]

- roughly 1.37×10^3
- 1
- roughly 15
- something > 60

Report and interpret the results. NOTE: The autotest will accept any of the above four answers, but only one of them is correct.

Problem 3 [20 Pts], Principal Component Analysis and Factor Models

For this question, I want you to think about using the methods of principal component analysis to construct a factor model that would solve a real-world problem. This problem does not require any coding. Rather, you will provide explanations, steps, intuition, and explanations to describe how you would conduct a PCA and a subsequent factor model.

a. Find an article in the news that has a problem that you think could be solved using PCA/factor models. When submitting your solutions for this homework, make sure to include the article in question.
b. From the article, formulate your hypothesis.
c. How would you use PCA to answer your hypothesis?
d. How would you formulate your PCA problem? Here, make sure to include the specifics, include the mathematical model, of the PCA in question. Include the data you would need and what you expect your output to be.

e. How would you convert your PCA into a factor model? Make sure to include the intuition for your factor model, how your variables map to your PCA, and what data you would likely need to empirically test your factor model.