Financial Econometrics - Homework 3

Professor Jeroen V.K. Rombouts

May 11, 2023

1 Guidelines

- Deadline: Before next lecture +1 by email to rombouts@essec.edu.
- If you use R, integrate your solutions into R-Markdown allowing you to embed the answers and code in one pdf file. if you use Python, make a Jypyter notebook file.
- Work in groups of 3.
- Explain the code making comments in each step of it.
- Professional presentation and visualisations are part of the evaluation.

2 How well can we explain excess stock returns with FF factors?

We inspire us by the paper of Eugene F. Fama and Kenneth R. French: A five-factor asset pricing model. Journal of Financial Economics 2015; 116:122. The daily data can be found on this website::
https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html
(Fama/French 5 Factors (2x3) [Daily]).

One of the main question in finance is if there exist a model that can explain (not forecast) returns. The model used by Fama and French is:

$$r_t^i - R_t^f = \beta_0 + \beta_1(R_{Mt} - R_t^f) + \beta_2 SMB_t + \beta_3 HML_t + \beta_4 RMW_t + \beta_5 CMA_t + \varepsilon_t \quad (1)$$

where R_t^f is the risk free rate available in the Fama French datasets, the first 3 variables are the former Fama French factors proposed in the 1980s, and RMWt is the difference between the returns on diversified portfolios of stocks with robust and weak profitability, and CMAt is the difference between the returns on diversified portfolios of the stocks of low and high investment firms.

To do this homework, download from Yahoo Finance daily adjusted closing prices (if possible from 2010 onwards) for each stock in the S&P 500, transform them in log returns and express them in percentages. To do this find the ticker list of the S&P 500 index and loop over this list.

QUESTIONS:

1. Do a regression using the Fama french factors for each stock excess returns and store the estimated coefficients, their t-statistics, and the R squared in a 500 x 13 matrix.

2. Compute descriptive statistics for each of the 13 columns of the previous question.

3. Compute nonparametric density estimates for each of the 13 t-statistics (you have about 500 values for each statistic) and make plots. Comment.

4. How differ the estimated factor coefficients according to the 11 sectors of the Global Industry Classification Standard?

5. Is it possible to cluster the stocks in groups in an unsupervised way? How many clusters do you recommend and what are typical characteristics of each cluster.

6. Sort the estimated constants and provide the names of the five companies with the highest constant and the five with lowest constant. Comment.

7. Temporally aggregating your returns to monthly frequency and taking monthly Fama-French data, redo the analysis and compare your findings with the daily frequency results.

8. As in the first question, estimate the daily data regressions but with left hand side variable $r^i_{t+1} - R^f_{t+1}$, that is one day ahead returns. How different are the results?

9. The model considered until now in (1) is a parametric linear regression model. Do you have other algorithms to link the factors to excess returns? If yes, show how they change the fit?

10. Which other factor can you can think of that should be in the model for excess returns? If you find daily or monthly data online, then redo the exercise above and interpret the findings for this factor parameter estimates and comment on how the previous parameter estimates have changed.