MECON6102
Problem Set 1 – Factor Model

Version: 2024/02/20

Due Date: 2024/04/16

Feel free to ask about algorithm-related problems. Please learn to debug your code using Google, GitHub, Stackoverflow, and the package manual.

In this assignment, you will run regressions to understand how the factor model explains asset return and construct an investment strategy based on the factor model.

1. Datasets

- Monthly Stock Returns: Monthly Stock File (upenn.edu)

- Fama-French 5 Factors: Kenneth R. French - Data Library (dartmouth.edu)

- Hou, Xue, Zhang q-Factor: Factors (global-q.org)

- Pastor Stambaugh Liquidity Factor: Robert Stambaugh's Home Page (upenn.edu)

Please first understand how these factors are constructed so that you will know what they represent.

2. (Navie) Factor Regression

For each stock, each month, use the data from 60 months ago to the previous month (it is okay if the sample has less than 60 months as long as it can be estimated in OLS), and run regression ("rolling-window regression"):

$$R_{i,t} = \alpha + \sum \beta_{i,j} F_{j,t} + \varepsilon_{i,t}$$

- $R_{i,t}$: Stock i's return at month t.

- $F_{j,t}$: Factor j's return at month t. Please use all factors provided (FF5, HXZ5, PS). However, both FF5 and HXZ5 contain market returns. Therefore, please only use market return in FF5 ("Mkt-RF") and drop market return in HXZ5 ("R_MKT").

- α: Constant term of the regression.

- $\beta_{i,j}$: Coefficients of factor j on stock i.

For each month, you can estimate the α and $\beta_{i,j}$. Therefore, iterating the process, you will have a time series of α and $\beta_{i,j}$ for stock i. Plot the time series of the beta of "Mkt-RF" and "α" and explain whatever you find for stock PERMNO=10324. Discuss the fluctuations of the time series of alpha and beta.

Rolling window means that, for example, for α and $\boldsymbol{\beta}$ in 201212, you use data from 200801 to 201212 to estimate.

200801	200802	200803	200804	...	201212	201301	201302	201303
				...	$\alpha_1, \boldsymbol{\beta_1}$			
				...		$\alpha_2, \boldsymbol{\beta_2}$		
							$\alpha_3, \boldsymbol{\beta_3}$	
								$\alpha_4, \boldsymbol{\beta_4}$

Then plot the time series of α and $\boldsymbol{\beta}$.

3. Fama-MacBeth Regression

If you do not know this regression, refer to this wiki: Fama–MacBeth regression - Wikipedia. Furthermore, you may search for materials on the Internet, which is relatively abundant.

By running Fama-MacBeth Regression, you can estimate the risk premium of each factor over time. You will obtain time series of $\gamma_{t,j}$ for factor j (notation in the wiki) by iterating rolling-window regressions. <u>Plot the time series of "Mkt-RF"'s risk premium and explain whatever you find.</u>

4. LASSO Regression

In running naïve factor regression, you can replace the estimation method (OLS) with LASSO (set $\alpha = 0.001$). Also, replace the factors with a 0- to 6-month lagged version. We are still using the previous 60 months (including the current month's) data to do the regression. For example, suppose we simply do CAPM by LASSO with lagged factors. It will be like this:

$$R_{i,t} = \alpha + \beta_0 Mkt_{t-0} + \beta_1 Mkt_{t-1} + \beta_2 Mkt_{t-2} + \cdots + \beta_6 Mkt_{t-6} + \varepsilon_{i,t}$$

The candidate factors are those provided (FF5, HXZ5, and PS). Then for each stock each month, you will have a group of factors that LASSO chooses. Suppose there is 60% of stocks choose Mkt_{t-2} factor (2-monthly lagged market return, $\beta_2 \neq 0$). That means this factor has good explanatory power to 60% of stocks in this month. <u>Please show the top 5 prevalent factors for each month. Given the limitation of computing power, please simply calculate the results for YYYYMM={200501, 200601, …, 202001}.</u>

You will produce a table like ("_l3" represents factor with a 3-month lag.):

YYYYMM	1st	2nd	3rd	4th	5th
200501	Mkt-RF_l0	AggLiq_l0	AggLiq_l3	LIV_Q_l0	LIV_Q_l1
200601	…	…	…	…	…
…	…	…	…	…	…
202001	…	…	…	…	…

You can also try various α to see how the results will be affected.

5. Mean-Variance Portfolio

Recall lecture note "Lec 4 Multi-Factor Model" page 12. For every year starting 2005, estimate the mean and variance of each stock. Mean and variance is calculated by factor model using data in the previous five years, which will be covered in Lecture 5. Please be reminded that choosing appropriate factors is important in calculating variance.

After compiling the mean and variance, calculate the weight to construct the most efficient mean-variance portfolio (assuming $\gamma = 1$ and no transaction costs). Please note that you can have long or short stocks in this practice. You will hold this portfolio for the following twelve months and then adjust your portfolio based on the calculation of the next iteration (i.e., rebalancing your portfolio annually). Finally, plot the cumulative returns of this investment strategy and calculate the annualized Sharpe ratio and maximum drawdown.

6. Deliverable

The final report should contain results generated by your program. It would help if you properly visualize them and provide interpretations of the results, for example, explaining why the coefficient of a factor change over time.

You should submit a ZIP file containing ONE code file (.py or .ipynb) and ONE analysis report (.pdf).