FBE 551 Homework #4

Due on 10/29/2025 by 11:59 pm

This is a group homework assignment. You should work in your teams, include all student names in a markdown cell at the top of the Jupyter Notebook that you submit.

To complete this homework assignment, you will need to use the following data sets

- crsp.monthly_stocks_HW4.feather: monthly stock data from CRSP.
- compustat_annual_HW4.feather: annual financials data from Compustat.
- FamaFrenchMonthly_HW4.csv: Fama-French 3 Factors, monthly returns.

Carry out the following calculations in a single Jupyter notebook. Make sure that I will be able to run it. For example, read in the files from the current directory. I.e., do not write the full path of the file in your code. Do not have your code change directories to a pathway that is probably not valid on my computer. This will allow me to run your notebook in any directory on my own PC.

Please use markdown cells to help me understand what you are doing. (I know it may be obvious, but I want to see that you know how to use markdown.) Clearly highlight what parts of your notebook are used to answer each question below.

Start

Read through the attached paper **'Fundamental Indexation'** which details the construction of a long-only "smart beta" strategy. Your task will be to recreate the core findings in Table 1 and Table 2 of the paper, with a couple of small adjustments, and then extend to an out-of-sample test.

Start by merging the CRSP and Compustat data sets in the usual way. Recall that you may need to install pyarrow to read the feather files. (Reviewing the '7. Momentum Investing and Fama MacBeth.ipynb' code from Week 7 may be helpful.)

You will then create indexes of the top 1000 stocks each month ranked and then weighted by the following metrics from Compustat:

- Book Value (SEQ in Compustat),
- Income (IB),
- Dividends (DVT)
- Sales (SALE).

You will also create a similar portfolio of the top 1000 stocks each month weighted by Market Cap (calculated from CRSP as PRC * SHROUT).

Note that in the paper they use 5-year averages of most of these accounting metrics; to keep it simpler, you will only need to use the most recently available value. Be sure to lag the Compustat variables by 6 months, as standard procedure to avoid look ahead bias.

After merging the CRSP and Compustat data, lag and fill forward the Compustat data so that each month in the resulting dataset will have a value for the Compustat variables, if available. You will then need to create a list of the top 1000 stocks each month ranked by the given metric, and produce portfolio weights. I recommend you define a function that will take the accounting column name as input and return a vector or series of monthly portfolio returns.

You will first need to rank the stocks each month by that metric, and select the top 1000. You will then create weights based upon that metric. Recall that each stock's weight in month t (by book value, for example) will just be:

$$W_{i,t} = \frac{Book\ Value_i}{\sum_{j=1}^{n} Book\ Value_j}$$

You can then multiply the stock return each month by that weight (*from the prior month, of course – avoid look-ahead bias).* Recall that a portfolio return is

$$r_{p,t} = \sum_{i=1}^{N} w_{i,t-1} r_{i,t}$$

So, the portfolio return for that month will be the sum of all those partial returns. Once you have the time series of portfolio returns, you can answer the questions below. Note that these questions ask you to produce tables such as in the paper. While it's helpful to see easy-to-read output, this does not have to be a consolidated, well-formatted table. I recommend you write a short function that will take the monthly returns for a portfolio and output the columns asked for in a single line. You can then run this function for each one of your portfolios. It is easiest then to use pd.concat() to put all 5 portfolios into one dataframe.

Question 1 (20 points)

Once you have all the portfolios in one dataframe, replicate the entries below to produce a version of Table 1 from the paper. This should include dates ONLY from 1962 through 2004, just like the paper. You should use the Fama-French MKT factor instead of the S&P 500 (add the RF rate to the MKT-RF to get the full market return). Make a table with the following 5 rows

- "Reference" (Market Cap weighted portfolio. Rank stocks by lagged market cap, keep the top 1,000 and weight them by lagged market cap to create the reference portfolio.)
- Book
- Income
- Sales
- Dividends

and the following columns:

- annualized *arithmetic* return (not geometric, unlike the paper, also multiply by 100 to put in percentage terms).
- annualized volatility (multiply by 100 to put in percentage terms)

- annualized Sharpe ratio (use annualized return and annualized volatility to compute this)
- Excess return vs. Reference (multiply by 100 to put in percentage terms).
- t-stat on the Excess return vs. Reference.

Please round all entries in the table to two decimal places. Note the excess return over the "Reference" and its t-stat for the "Reference" portfolio (first row of the table) should be NaN or zero since you are taking the excess return of the portfolio with itself.

Question 2 (20 points)

Produce a version of Table 1 from the paper, as in Question 1, but this time for the period 2005-2024. How well have the returns and characteristics held up in this out-of-sample test? Make some comments in a markdown cell.

Question 3 (20 points)

Create a version of Table 2 from the paper that includes the same portfolios as in Questions 1 and 2. Again, this should include dates only from 1962 through 2004. For columns, produce

- Arithmetic annualized return (multiply by 100 to put in percentage terms)
- CAPM beta vs. the MKT-RF factor (NOT vs. "Reference" as in the paper)
- CAPM alpha (annualized and multiply by 100 to put in percentage terms)
- t-stat on CAPM alpha

Do this for both the 1962-2004 period and the 2005-present period. Please round all entries in the table to two decimal places. Make some comments in a markdown cell – how well have the results held up out of sample?

Question 4 (20 points)

Create a table that includes the same portfolios as in the previous 3 questions. This time, for columns, produce

- Arithmetic annualized return (multiply by 100 to put in percentage terms)
- A Fama-French 3-factor beta on the MKT-RF factor
- A Fama-French 3-factor beta on the SMB factor,
- A Fama-French 3-factor beta and the HML factor
- The FF 3-factor alpha from the same regression
- t-stat on the FF3 alpha.

Do this for both the 1962-2004 period and the 2005-present period.

Question 5 (20 points)

Make some comments about your FF3-factor analysis from Question 4 in a markdown cell.

- How do the CAPM results change out-of-sample?
- What happens to the CAPM results in the paper's sample once we include other factors in the FF3 model?
- What does this imply about the source of return being captured by the factors book, income, sales, and dividends?
- What has this analysis taught us about the nature of returns to “Fundamental Indexation”?