FINA 4335-Final Project

Fall 2024-Prof Lingfei Kong

Project Introduction

This is a **tough** and **explorational** project. There are two parts in the final project.

In Part I, you will use eight assets to study technical trading strategies. Five of these assets are the Fama-French 5 industry portfolios, while the remaining assets will be generated by you.

In Part II, you need to construct factors based on multiple investment styles. For each investment style, you want to create 5 long-only equal-weighted quintile portfolios and 1 factor (long-short). You will also create another factor that incorporates all the investment styles.

The necessary datasets are either available in Canvas or accessed via `pandas_datareader`.

Project Submission

After completing the project, submit four well-formatted files by Dec 11, 2024 11:59 pm ET:

- **Two** notebook answer files that shows all the outputs (one for each Part), see Canvas
- **Two** HTML files that shows all the outputs (one for each Part), see Canvas

Grading criteria

The grading will be based on both $\boldsymbol{correctness}$ and $\boldsymbol{effort}$. See the jupyter notebook files for details.

Project guidance

- Most relevant lecture notebooks, HW, and tools
    - Module 7 Stock Returns; Module 10 Portfolio Risk and Return; Module 11 Factor Models; Module 12 Quantitative Trading Strategies
    - HW3; HW4
    - AI tools such as **ChatGPT** and **Copilot** can *assist* your coding and save you a lot of time.
- I will give some general guidance on the group project on Dec 4th class time. I will also hold some office hours (time to be announced).
    - Please note that the instructor will not review or debug your code for the final project.

Part I: Industry trends

In this part, you will use 8 assets to study one type of technical trading strategy (**monthly**). Five of these assets are the Fama-French 5 industry portfolios (accessible via \`pandas_datareader.data\`), while the remaining assets will be generated by you.

1.1 Download Fama-French 5 industry portfolios (2 points)

Download the monthly returns for the Fama-French 5 industry portfolios from January 1, 1995, to November 30, 2024, and save them in a DataFrame named \`ff5ind_monthlyret\`. Convert the percentage returns to decimal format. You will use this dataset, along with other necessary data accessed via \`pandas_datareader\` for Part I.

1.2 Create 3 additional portfolios (15 points)

The 3 additional portfolios are defined as follows:

- 'ew_reb': A portfolio rebalanced monthly to target equal weights across the 5 industries (20% in each industry).
- '$tech\_heavy\_reb$': A portfolio rebalanced monthly with weights targeting 40% in the high technology industry and 15% in each of the other four industries.
- '$consumer\_heavy\_reb$': A portfolio rebalanced monthly with weights targeting 40% in the consumer industry and 15% in each of the other four industries.

Save the portfolio returns in a DataFrame named \`Extra_3assets\`, with columns corresponding to the portfolio names (e.g., 'ew_reb').

Next, merge this DataFrame with \`ff5ind_monthlyret\` and name the merged DataFrame \`assets_monthly\`. Create a bar plot to visualize the average returns for the 8 assets.

2. Simple moving average strategy (SMA) for each asset (15 points)

- **MA3_CPI**: For each asset, take a long position ($position = 1$) when the most recent CPI (The Consumer Price Index for All Urban Consumers, or CPIAUCNS in St Louis Fed) is larger than MA3 of CPI, take a short position ($position = -1$) when the most recent CPI is less than or equal to MA3 of CPI.

Here are a few useful links about CPI:

https://fred.stlouisfed.org/series/CPIAUCNS
https://www.bls.gov/schedule/news_release/cpi.htm

Save the portfolio returns in a DataFrame called \`MAs_CPI_rets\`. Next, create a **bar plot or similar visualization** to display the average returns of the SMA strategy.

3. Portfolio evaluation (8 points)

Report the Sharpe ratios and maximum drawdowns for the SMA strategy. Store the results in DataFrames called \`sharpe_by_asset\` and \`MDD_by_asset\`. Which asset's SMA strategy has the highest Sharpe ratio? Which asset's SMA strategy has the smallest magnitude of MDD? Why do you think this happens? Cite credible sources to support your answer.

Part II: Factor Investing

In this part, you need to construct factors based on multiple investment styles. For each investment style, you want to create 5 long-only equal-weighted quintile portfolios and 1 factor (long-short). You will also create another factor that incorporates all the investment styles. The datasets in Canvas, along with other necessary data accessed via `pandas_datareader`, will be used to explore various factor strategies.

Below are the investment styles you want to explore:

> **Value:** Value is measured by a stock's price-to-sales ratio, where price is the unadjusted price from month $t-1$, and sales are calculated as revenues (already '$lagged$') per share (number of shares should be from month $t-1$). To construct the factor, you should take a long position in stocks that belong to the bottom quantile of price-to-sales, take a short position in stocks that are in the top quantile.
>
> **Momentum**: Momentum is measured by a stock's month $t-12$ to month $t-2$ cumulative return. To construct the factor, you should take a long position in stocks that belong to the top quantile of cumulative return, and take a short position in stocks that are in the bottom quantile.
>
> **Reversal**: Reversal is measured by a stock's cumulative return over the last available 2 weeks in month $t-1$. To construct the factor, you should take a long position in stocks that belong to the bottom quantile of cumulative return, and take a short position in stocks that are in the top quantile.
>
> **Size**: Size is measured by a stock's market capitalization at the end of the previous quarter (Note: there are four quarters a year). To construct the factor, you should take a long position in stocks that belong to the bottom quantile of size, and take a short position in stocks in the top quantile.
>
> **Beta:** Beta is measured by a stock's CAPM beta estimated with 24-month window (month $t-24$ to month $t-1$; min window size =24). To construct the factor, you should take a long position in stocks that belong to the bottom quantile of beta, take a short position in stocks in the top quantile.
>
> The trading directions for the above factors are consistent with theories.

Data preparation (5 points)

1. Download **`weekly_rets.csv`** and **`monthly_rets_funda.csv`** from Canvas, read them into Python as DataFrames named `weekly_rets` and `monthly_rets`, respectively. Parse the 'DATE' column to `datetime` format in both DataFrames. Next, sort both DataFrames in-place by the columns 'PERMNO' and 'DATE'.

Constructing factors (33 points; 15+8+5+5)

You need to create signals, then construct a few factor-investing style portfolios. For each investment style, you want to create 5 long-only equal-weighted quintile portfolios and one factor (long-short). The methodology follows the approach taught in this course (FINA 4335).

2. In a DataFrame called `mrets_with_signals`, store all columns in `monthly_rets` along with the 5 signals (from value to beta) used to construct factors. The signals should be called 'price_to_sales' , 'ret_2to12m', 'ret_2w', 'mcap_quarter', and 'beta'.
   - Note: We should preserve all original values in `monthly_rets`
   - *Each signal is only worth 3 points. Suppose you don't know how to generate a specific signal, then you may use random values (with somewhat reasonable range) to replace the signal. You will lose 3 points but will get most credit for subsequent questions.*
3. For each signal, create equal-weighted quintile portfolios along with the factor (long-short), then save them in DataFrames that are named `value_df`, `momentum_df`, `reversal_df`, `size_df`, and `beta_df`, respectively. There should be 6 columns in each DataFrame, where the first 5 columns are long-only quintile portfolios and the last column is the factor (long-short).
4. Create a well-organized plot to illustrate the average returns of the 25 long-only quintile portfolios from the five DataFrames. Ensure the plot is clear, visually appealing, and not overly crowded. (Example: A nice plot with 3 bar subplots on the top and 2 bar subplots on the bottom)
5. Save the 5 long-short factor returns in a DataFrame called `factors_df` (Do not drop any missing values).

Evaluate the factors you created (14 points; 4+4+6)

6. Estimate Fama-French 5-factor model for the 5 factor portfolios (in `factors_df`). Report the R squares, regression coefficients, and their p values in a single DataFrame called `ff5_factors_reg`.
7. Store the annualized average, annualized median, and annualized standard deviation of returns for the 5 factor portfolios in `factors_df` in a DataFrame named `factors_sumstat`.
8. Summarize at least three takeaways related to investments based on your results in Part II so far. Cite credible sources that help explain and support your findings.

Factor combination strategy ('Mission Impossible') (8 points)

9. Going back to `mrets_with_signals`, let's add a new column called 'index', which ranges between 1-100, is the arithmetic average of its adjusted percentile bucket (1-100) for each of the 5 signals. The bucket labels should be adjusted in a way that 100 represents the theoretically best performing percentile, 1 represents the theoretically worst performing percentile for each investment style.
   - Note: The way to construct the percentile bucket is very similar to quintile bucket except that you need to make adjustment of bucket labels. For example, the smallest size percentile bucket should be labeled as 100 and the largest size percentile bucket should be labeled as 1.

   Next, use $'index'$ as the signal to create five quintile portfolios and an $'index'$ factor. The quintile portfolios are **value**-weighted. You need to figure out the trading direction for this factor. Save the 6 portfolio returns in a DataFrame called `impossible_df`.