Problem Set 2: Building Factor Portfolios

California Institute of Technology
BEM 114: Hedge Funds
Spring 2025

Due: April 18, 2025

Introduction

In this homework assignment you are asked to replicate simple factor strategies using historical financial data. I have uploaded a copy of the CRSP database to Canvas, called "crsp_1926_2020.csv." This database contains monthly stock returns, and also prices and shares outstanding. The Fama French, Momentum, and Industry factor returns can be obtained from Ken French's website.

Submission Instructions

I recommend you complete the assignment using Python, but you are free to answer the questions in the programming language of your choice. Please submit your assignments on Canvas before 11:59 p.m. on the due date. Assignments should be **typed**, and the first page should include the names and student IDs of all members in your group. Please upload a PDF file with your answers, and a replication package containing your code. You do not need to submit the data.

Questions

1. **Data Cleaning and Summary Statistics**: Financial data is notoriously messy, and the CRSP ("The Center for Research in Securities Prices") database aims to supply a high-quality database of security prices and data. CRSP covers all U.S. stocks but will also list ETFs and other securities that trade on exchanges. We need to be judiciously careful when cleaning our database to ensure we are looking at stocks only.

   (a) (5 points) Clean the CRSP data using the following conditions:
   - Only include stocks that are ordinary/common shares (shrcd = 10 or 11),
   - Only include stocks listed on the NYSE, AMEX, or NASDAQ (exchcd = 1, 2, 3),
   - Set negative prices to NA (prc < 0). Negative prices occur when the stock has no trading on that particular day, and negative prices represent the midpoint of the bid-ask spread.

   (b) (5 points) Plot the number of listed firms per month over the entire sample period.

   Total for Question 1: 10

2. **Replicate Size**: Gene Fama and Ken French published the size and value factors in June 1992. In this question we will replicate the size strategy from 1926 up to 2020.

   At each month t, sort stocks into ten deciles by size. Each month take a long position in the small stocks and a short position in the large stocks, and repeat the following month.

- Use the sample code from the value strategy example in class as a guide.

(a) (5 points) Form the equal- and value-weighted portfolios for the ten size portfolios.

(b) (2 points) Calculate the mean monthly returns for each decile, are these monotonic?

(c) (3 points) Form the long-short small-minus-big portfolio and calculate the mean, volatility, and Sharpe ratio.

(d) (5 points) Estimate the CAPM and FF3 models for both equal- and value-weighted portfolios. How do the alphas change?

(e) (5 points) Does size still work? Investigate the strategy performance after the Fama French 1992 paper was published. How about after the Dot-Com Bubble burst (starting in around 2002)?

Total for Question 2: 20

3. **Replicate Momentum**: Narasimhan Jegadeesh and Sheriden Titman first published the momentum strategy in March 1993. In this question we will replicate the momentum strategy from 1926 to 2020.

   To form these portfolios, at each month t calculate cumulative returns for each stock i over the 11 month period from time $t-12$ to $t-1$. At each month, sort stocks into ten deciles by past returns. Identify the winners as the 10% of stocks that performed the best, and the losers as the 10% of stocks that performed the worst. Each month take a long position in the winners and a short position in the "losers," and repeat the following month.

   - I would group the data by PERMNO and then calculate rolling cumulative returns.
   - The pandas package has a useful function called `.rolling()`. Rolling allows you to apply a function to a data series. This function takes in arguments that allow you to specify the size of the rolling window and the minimum number of periods.
   - You can then apply a function to this rolling window. To do this efficiently you can consider using the lambda function, but you can also apply your own function. For instance, `.apply(lambda x: np.prod(1+x)-1)` will take the product of 1 plus each element of the rolling window, and then subtract off 1. This is useful for calculating rolling cumulative returns.

   (a) (20 points) Form the equal- and value-weighted portfolios for the ten momentum portfolios.

   (b) (2 points) Calculate the mean monthly returns for each decile, are these monotonic?

   (c) (3 points) Form the long-short winners-minus-losers portfolio and calculate the mean, volatility, and Sharpe ratio.

   (d) (5 points) Estimate the CAPM, FF3, and FF5 models for both equal- and value-weighted portfolios. How do the alphas change? Does the FF5 model price momentum?

   (e) (5 points) Recall that alpha is positive either if a hedge fund manager has skill, or if the underlying model does not price all the risks that investors care about. Are the momentum alphas indicative of managerial skill? Why or why not?

   Total for Question 3: 35

4. **Replicate Betting-Against-Beta**: Andrea Frazzini and Lasse Pedersen of AQR published the Betting Against Beta strategy in January 2014. This is a strategy currently in use by AQR.[1] In this question we will replicate the Betting-Against-Beta strategy.

   The full replication is described in Franzzini and Pedersen (2014), but we will take a simpler route. The key to this strategy is to estimate rolling betas for each stock. For each stock i at month t, estimate the monthly CAPM using stock returns from $t-36$ to t and record the CAPM beta. At each month t, sort stocks into ten deciles by their estimate beta. Each month take a long position in the low beta stocks and a short position in the high beta stocks, and then repeat the following month.

   - In this question you need to calculate rolling betas. This will be more difficult because it requires a rolling calculation on a dataframe.
   - I would suggest grouping the data by `PERMNO` first, and the applying a function to the groups.
   - The `statsmodels` package has a rolling regression function called `RollingOLS`. The syntax is slightly different from `OLS`, and you should familiarize yourself with it. The key parameter will be `window`.
   - `RollingOLS` will create a vector of models, which you can estimate using the `.fit()` command. Betas can be accessed from `model.fit().params[['Mkt-RF']].values`.

   (a) (20 points) Form the equal- and value-weighted portfolios for the ten deciles of BAB portfolios.

   (b) (2 points) Calculate the mean monthly returns for each decile, are these monotonic?

   (c) (3 points) Form the long-short BAB portfolio and calculate the mean, volatility, and Sharpe ratio.

   (d) (5 points) Estimate the CAPM, FF3, FF5, and FF5+Momentum models for both equal- and value-weighted portfolios.

   (e) (5 points) This is an important strategy for AQR; it produces alpha, but the Sharpe ratio is quite low. Suppose you were tasked with reducing the volatility of this strategy, what might you try to do? Why?

   Total for Question 4: 35

[1] See: `https://www.aqr.com/Insights/Datasets/Betting-Against-Beta-Equity-Factors-Monthly`.