Assignment - Momentum, Value, and Size Premiums

AI tools are becoming part of everyday data analysis, and you will increasingly be asked to answer questions like the one in this assignment using AI. That said, AI can't be trusted blindly. Humans still matter (for now!) because we can tell when results make sense and can push the AI in the right direction.

For this assignment, you will use Claude Code to do the analysis. Claude will write, debug, and iterate on the Python code. You are not expected to understand the code it generates. What you are responsible for is understanding the economic results and making sure the analysis is producing sensible answers by asking the right questions.

Together, you and Claude will test whether classic asset pricing factors show up in a large sample of U.S. stocks using firm-level regressions. This is an individual assignment. Your deliverable is a short write-up (3–4 pages) that answers the questions below, includes regression tables, and explains what the results mean. Claude can do that as well.

Do not dig into the code Claude produces. Just work through the Claude Code interface, feed the entire pdf, and ask any questions you think are necessary. Below are the instructions.

Data

Download the dataset from this link. The file contains merged CRSP–Compustat monthly data covering U.S. public firms.

Question 1: Load and Explore the Data

Load the data. Report the number of observations, the date range, and the number of unique firms.

Question 2: Define the Variables

Construct the following factor characteristics for each firm-month:

(a) **Market Beta**

(b) **Size**

(c) **Value** (Book-to-Market)

(d) **Momentum**

Question 4: Single-Factor Model

Run a pooled OLS regression of next-month firm returns on market beta only. Make sure you correct for correlation in the error term. Interpret the coefficient.

Question 5: Full Factor Model

Run the full model with all four factors: market beta, size, value, and momentum. Use pooled OLS and correct for correlation in the error term.

(a) Report the coefficient estimates and t-statistics.

(b) Interpret the signs of the coefficients.

(c) Are your results consistent with the findings in:

- Jegadeesh and Titman (1993), "Returns to Buying Winners and Selling Losers," *Journal of Finance*
- Fama and French (2015), "A Five-Factor Asset Pricing Model," *Journal of Financial Economics*

Question 6: Are the Effects Linear?

Add a quadratic term for each of the four factors. Are any of the quadratic terms statistically significant? What does this imply economically?

Question 7: Subsample Analysis

Re-run the full linear model from Question 5 on different subsamples of the data. Have the premiums changed over time? Which ones have weakened? Which ones remain robust?

Question 8: Out-of-Sample Backtesting

Test whether your model works out of sample. Estimate the full linear model on a training set, predict returns on a test set, and form long-short portfolios. Report the average return and Sharpe ratio of the strategy.

Question 9: Factor Farming

The dataset contains over 900 accounting variables. Many of these could, in principle, predict returns.

(a) Select 20 plausible accounting variables as candidate factors (e.g., profitability, leverage, investment, cash flow, etc.). Normalize them cross-sectionally and run the same regression specification for each one.

(b) Split the sample into a training period and a test period. Identify the factors that are significant in-sample, and check whether they survive out-of-sample.

(c) Report how many candidate factors are significant in-sample, how many survive out-of-sample, and the survival rate.

(d) Discuss the implications for factor investing. Why do most in-sample factors fail out-of-sample?