The point of this assignment is 1) to make sure that you don’t get out of this class without knowing how to download and organize the basic CRSP and Compustat data; 2) to get you looking at real data before you actually have to produce a paper; and 3) to basically lower the bar to doing empirical work by giving you a starter library of code that can be modified and expanded in the future to suit your needs. The assignment will consequently be a lot more work, but a lot more valuable, to those early in the program. If you have your own code and are comfortable with your way of doing things, there’s no need to use mine.

The assignment: Please replicate some known result from the literature on the cross section of expected stock returns (see below for a list of possible topics—but please feel free to choose a result that is not on the list). Please adhere to the following guidelines:

- Be skeptical!
  - There are at least a couple topics on the list where I think the data only provide very weak support (at best) for the authors’ claims.
- Please go beyond what was done in the literature in some dimension. I want you to (try to) replicate the basic empirical facts—not replicate the paper.
  - How persistent are the strategy’s abnormal returns (does the “signal” predicting returns only have highly transitory power)?
  - Is the result there outside the really small stocks?
    - Almost everything is much stronger in the small caps.
  - Sub-sample (or out-of-sample) results?
  - Is it stronger in industries, or across industries?
    - You do not need to do all of these. Please just pick some further dimension to explore that you think is particularly relevant to your topic (and it doesn’t have to be one of those listed above).
- As data checks along the way my code looks at some very basic results (e.g., value and momentum). If you choose one of these topics, then I’ll expect you to do more above and beyond the basic results.
- If you already have experience with these data, the please pick something you haven’t looked at before.
I’ll also expect you to be a little more ambitious in your analysis.

- You don’t need to write a paper. Turn in a few tables showing the basic results and the additional dimension you investigate. Include a brief summary (few paragraphs) summarizing what you found, and in what ways it’s both consistent and inconsistent with the prominent literature.
  - Be concise! I do not want to see the results of every test you ran. It is as important to figure out what not to include as what to include
    - Early in your career writing papers you’ll have a hard time leaving out any table that has anything mildly interesting (they’re your babies)—but if a table isn’t central to your story, then drop it!

If you don’t have your own code, I have posted a zip file (“Matlab Package”). It has detailed instructions, and corresponding code, that will walk you through getting and organizing the data. Start with the word files “CRSP data exercise” and “Compustat data exercise.”

**Some topic suggestions**

The following topics are simply meant as possibilities (they are relatively easy, and can be done with the data from the exercises), and are no means meant to be exhaustive. Please feel free to pick a topic that does not appear on this list.

**Topic: Performance of the CAPM / Betting Against Beta**

Data Issues: Requires rolling estimates of stock betas—the Make_betas script in the project folder helps with that.

**Primary references**

**Topic: Value and Long Run Reversals**

**Primary references**
Topic: **Characteristics vs. Covariances**

Data Issues: Requires rolling estimates of stock betas—the Make_betas script in the project folder helps with that.

**Primary references**

**Basic momentum**

**Primary references**

**Industry and style momentum**

**Primary references**

**Momentum Refinements**

**Primary references**
Topic: **Profitability**

**Primary references**

Topic: **Post Earnings Announcement Drift**

Data Issues: Requires quarterly earnings announcements and announcement dates—see “Getting the Quarterly Data” in the project folder.

**Primary references**

Topic: **The Accruals Anomaly**

**Primary references**

Topic: **Mutual Fund Performance Persistence**

Data Issues: Requires mutual fund return data—see “Getting the Mutual Fund Data” in the project folder.

**Primary references**
Topic: **Investment Anomalies (asset growth and capital expenditures)**

**Primary references**

Topic: **Distress anomalies (e.g., the under-performance of troubled firms)**

Data Issues: Uses quarterly data— see “Getting the Quarterly Data” in the project folder.

**Primary references**

Topic: **Equity issuance and under-performance**

Data Issues: Works much better with quarterly data— see “Getting the Quarterly Data” in the project folder.

**Primary references**

**Other possibilities**
- The Stock Market Valuation of Research and Development Expenditures (Chan, Lakonishok and Sougiannis, JF 2001)—also looks at advertising.
- Industry Concentration and Average Stock Returns (Hou and Robinson, JF 2006).
- Evaluate the performance of investment strategies from the popular literature (for example, Greenblat’s “Little Book That Beats the Market).