

# Big Data in Finance - Part III

## CRSP and Compustat: an Application to Quant Finance

Ritt Keerati\*

February 11, 2022

During this part of the course, you will gain familiarity with two of the most widely used datasets in finance: CRSP and Compustat. The course will be based on hands-on applications in quantitative investing and is designed to give you exposure to up-to-date empirical work in asset pricing and its applications in quantitative finance.

## Class Material

All class material is available in a GIT repository in Bitbucket. You can clone or download [here](#).

## Schedule

- Lecture 5: Introduction, CRSP and Compustat basics
- Lecture 6: Application to asset pricing: Factor replication
- Lecture 7: Quantitative investing

## Prerequisites

### 1. Install Anaconda

- I use **Anaconda**, one of the most popular Python distributions.
- Anaconda is easy to install and comes with a great package management system called *conda*.
- In class, I am going to use **Python 3.8.12**.
- Follow this link to install Anaconda. Make sure you choose the correct operating system.

### 2. Working knowledge of Python

---

\*I am greatly indebted to Lira Mota for sharing her teaching materials for this course with me. If you find typos, or have any comments or suggestions, then please let me know via [ritt.keerati@gsb.columbia.edu](mailto:ritt.keerati@gsb.columbia.edu).

- Choose an IDE in which you can run python code (.py) and notebooks. I recommend using Jupyter Lab or PyCharm. You can install it using Anaconda or, for PyCharm, it can be downloaded for free [here](#).
  - Student and faculty members license is for free, you only need to apply at PyCharm license.
3. WRDS direct connection with Python
    - WRDS has built a Python module that allows direct download of datasets from WRDS services in Python. This is very convenient and we are going to use this tool in class.
    - In order to use the direct download you need to setup your connection beforehand by following the instructions [here](#).
    - The WRDS support is very responsive, so make sure to email them if you need help to setup your connection.
  4. Working knowledge of GIT
    - All course material will be available in Bitbucket repository. Access [here](#).
    - Make sure to setup a Bitbucket account [here](#) and study the GIT basics [here](#).
    - Using GIT will change the way you collaborate in research projects, making it much easier to organize and keep track of changes made by you or your colleagues.
  5. Optional: power-up your Jupyter Notebook.
    - Notebooks are great to produce documents you intend to present.
    - We are going to use notebooks during class.
    - Here you can find a description of very useful plugins for Jupyter Notebooks. I highly recommend that you install the suggested plugins.

## Homeworks

There will be **two** homeworks.

1. Due Thursday, 03/24: Exploring CRSP and COMPUSTAT
2. Due Thursday, 03/31: Quantitative investing

## Lectures

### Lecture I: CRSP and Compustat Basics

1. Introduction
  - (a) WRDS basics
  - (b) How to download data into Python
2. CRSP
  - (a) Securities File Monthly
  - (b) Securities File Daily

- (c) Events Table
  - (d) Stock Header Info
3. Compustat
- (a) Fundamentals Annual
  - (b) Fundamentals Quarterly
  - (c) Pension Annual
  - (d) Names Table

## Lecture II: Asset Pricing Factor Replication

1. CRSP and Compustat merge
2. An overview of Fama and French factor construction technology
3. Characteristics Construction: Fama and French (2015) + Momentum
  - Size (CRSP)
  - Book to Market (Compustat)
  - Profitability (Compustat)
  - Investment (Compustat)
  - Momentum (CRSP)

4. Replicate Fama and French (2015) five factors and momentum factor.

### References:

- **Fama, Eugene and Kenneth French, “Common Risk Factors in the Returns on Stocks and Bonds,”** 1993, *Journal of Financial Economics*, 33, 3-56.
- **Jegadeesh, Narasimhan and Sheridan Titman, “Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency,”** 1993, *Journal of Finance*, 48, 65-91.
- **Fama, Eugene and Kenneth French, “A Five-Factor Asset Pricing Model,”** 2015, *Journal of Financial Economics*, 116, 1-22.

## Lecture III: Quant-Investing

1. Performance evaluation: Alpha evaluation and Fama-MacBeth regressions
2. A word on back testing
3. A word on trading costs
4. Performance analyses
  - (a) Value investing
  - (b) Momentum
  - (c) Characteristic efficient portfolios

### References:

- Harvey, Campbell, Yan Liu and Heqing Zhu, “... and the Cross-Section of Expected Returns”, *Review of Financial Studies*, 2016, 29, 5-68.
- McLean, David and Jeff Pontiff, “Does Academic Publication Destroy Stock Return Predictability?” 2016, *Journal of Finance*, 71, 5-32.
- Fama, Eugene and Kenneth French, “A Five-Factor Asset Pricing Model,” 2015, *Journal of Financial Economics*, 116, 1-22.
- Korajczyk, Robert and Ronnie Sadka, “Are Momentum Profits Robust to Trading Costs?,” 2004, *Journal of Finance*, 59(3), 1030-1082.
- Novy-Marx, Robert and Mihail Velikov, “A Taxonomy of Anomalies and Their Trading Costs,” *Review of Financial Studies*, 2017.
- Daniel, Kent and Tobias Moskowitz, “Momentum Crashes,” 2016, *Journal of Financial Economics*, 122(2), 221-247.
- Daniel, Kent, Lira Mota, Simon Rottke and Tano Santos, “The Cross-Section of Risk and Returns,” 2020, *The Review of Financial Studies*, 33(5), 927–1979.