Big Data in Finance

Part II - CRSP and Compustat

Lira Mota *

February 17, 2020

Schedule

- Lecture 7: Introduction and setting up your environment (326 Uris)
- Lecture 8: CRSP basics (306 Uris)
- Lecture 9: Compustat basics (326 Uris)
- Lecture 10: Factor Investing Part I (306 Uris)
- Lecture 11: Factor Investing Part II (326 Uris)

Prerequisites [1]

1. Working knowledge with Python.
    - I recommend using PyCharm as Python IDE for Python 3. It can be downloaded for free at download PyCharm.
    - Student and faculty members license is for free, you only need to apply at PyCharm license.
2. WRDS direct connection with Python.
    - WRDS has built a Python module that allows direct download of data sets 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 first time you try to setup your connection it might not work, unfortunately. The WRDS support is very responsive, so make sure to email them if you need help to set up your connection.
3. Working knowledge with GIT.
    - All course material will be available in Bitbucket repository. Access here.

*Columbia Business School. lmota20@gsb.columbia.edu

[1] Luckily enough, I know Kriste covered most of the material here. Make sure to have everything up to speed by Lecture 2.

- It is mandatory that you create your own git repository for this class. Part of your homework will be to send us your git log - we will go over it during class.
- Make sure to set up 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.

4. 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 three homeworks. Only homework 2 and 3 are going to be graded.

1. Due 02/19: Set up your environment;
2. Due 02/26: Playing with CRSP and Compustat;
3. Due 03/02: Factors replication.

Lectures

Lecture 8: Setting up your environment

1. Introduction
2. WRDS basics
3. How to download data into Python

Lecture 9: CRSP

1. Homework I due.
2. CRSP
   (a) Securities File Monthly
   (b) Securities File Daily
   (c) Events Table
   (d) Stock Header Info

Lecture 9: Compustat

1. Compustat:
   (a) Fundamentals Annual
   (b) Fundamentals Quarterly
   (c) Pension Annual
   (d) Names Table
2. Characteristics Construction: Fama and French (2015)
   (a) Book to Market (Compustat)
   (b) Profitability (Compustat)
   (c) Investment (Compustat)

Lecture 10: Factor Investing Part I

1. Homework II due.
2. CRSP and Compustat merge
3. Characteristics Construction: Fama and French (2015) + Momentum
   - Size (CRSP)
   - Momentum (CRSP)
4. Portfolio sorts

Lecture 11: Factor Investing Part II

1. Homework III Due.
2. Replicate Fama and French (2015) five factors and momentum factor.
3. Alpha evaluation.
4. Signal evaluation: Fama-MacBeth.