

Big Data for Finance Business 9334 Spring 2024

Overview

Instructors

- Kirsten Burr
- Michael Ewens
- Yiming Ma
- Harry Mamaysky (coordinator)
- Ciamac Moallemi
- Tomek Piskorski
- Razvan Popescu

Room

TBD

Lecture hours

Fridays, around 9am-noon

TAs

Meha Sadasivam

Office hours

TBD

Course Description

This course provides an introduction to financial data, data analysis tools and approaches, and analyzing statistical models using output from these datasets. The course schedule below lists the data sets that will be covered.

Course Requirements

The course will have extensive weekly homework assignments and a take-home final exam. Course attendance is required. Participation is strongly encouraged.

Course Prerequisites

This course requires three PhD courses: B9320, Econometrics and Statistical Inference (our first term PhD course); B9122: Computing for Business Research; and B9302 Finance Theory I. We assume students are familiar with the finance theories and econometric techniques used in these courses.

In terms of programming background, we expect proficiency in Python as well as knowledge of SQL (both covered in Fall Term “Computing”).

Assignment and Grading

Problem Sets: *(Note the exact method of handing in homework may change.)*

- There are weekly problem sets during the semester, each will be graded on a scale of 0-10 points. The lowest one will be dropped from your grade.
- Students are allowed/encouraged to work in groups, but each student must do their own analysis and submit their own original work, which includes computer code. Every student is required to do their own coding (you won't learn to code unless you actually code). It is, however, useful to discuss issues with others so we encourage students to work in groups with each other. Turning in another student's code (even if you change all the variable and function names) is a violation of the honor code.
- Each homework assignment will indicate the day on which it is due. Late homework is not accepted under any circumstances. Since your lowest homework is dropped, you have a cushion to work with.
- Code guidelines for homework. It is important for your own work/research to write clear, well-documented code. We will enforce the following requirements for all homework assignments, except those requiring you to hand in shell/bash scripts. We will deduct up to 2 out of 10 points for every assignment you hand in that does not follow these guidelines:
 - Driver Script: You should submit a main file that contains the answers to all questions. We recommend using a Jupyter notebook but a script file also works. If you are using a Jupyter notebook, please also attach a .html or .pdf file. Please make sure that we can run all the codes/scripts – a common mistake is to include absolute paths. If you are using functions from other files, these functions should be properly imported.
 - Functions: Please write and use functions in your driver script. Do not repeat the same code in multiple places. All the functions should be in a separate module/file.
 - It is up to you how you break down the logic (you are allowed to have more than one function for each part of the assignment), but the naming must be clear and indicative of what the function is doing.
 - Documentation: Your code should be self-explanatory. This means that you must add comments wherever relevant to explain what you are doing in that line/section of code, and what it is used for. Use your best judgment when adding comments – “over commenting” is okay, but “under commenting” is not. Commenting your code will make it understandable to *you* when you come back to it months or years later.
 - Submission: Zip the files before uploading to Canvas, and name the zip `UNI_firstname_lastname.zip`

Exams: There is a take home final, details TBA.

Course Grade: The grade is 40% final and 50% homework and 10% participation. The latter will be enforced via attendance sign-in sheets at the start of each class.

Course Schedule for spring 2024

Date	Class	Instructor	Topic
1/26	1	Razvan Popescu	Grid and parallel computing
2/02	2		
2/09	3	Harry Mamaysky	Thomson-Reuters and natural language processing
2/16	4		
2/23	5	Kirsten Burr	CRSP, Compustat, and Fama-French
3/01	6		
3/08		No class	Exam period (this class has no midterm)
3/15			Spring break
3/22	7	Yiming Ma	Mutual funds and ETFs
3/29	8		
4/05	9	Michael Ewens	Analysis of VC/PE space using SEC filings and Form ADVs
4/12	10	Tomek Piskorski	Real estate data: Residential real estate
4/19	11		Real estate data: Commercial real estate
4/26	12	Ciamac Moallemi	Cryptocurrency data