Caltech

BEM 106: Data Science in Economics and Finance

Course Syllabus – Spring Quarter / 2025
Division of the Humanities and Social Sciences
Caltech

Course Instructor

Prof. Pawel Janas (he/him)
Office: Baxter 135
Email: pawel@caltech.edu
TA: Georgia Malueg
TA Office Hours: Fridays 12:30 – 1:30 in Baxter B133
Prof. Office Hours: Tuesdays, 4:00 – 5:00 PM in Baxter 135

Course Description

This course revolves around qualitative topics and quantitative methods in data science for finance and economics. We discuss linear models and inference, sources of bias, databases and data validation, web scraping, and introductory machine learning methods for clustering and text analysis. A special emphasis is placed on developing a robust understanding of the strengths and weaknesses of social science research methods as they apply to real-world problems. The class also introduces students to the Python programming language and its ecosystem of packages.

Course Goal

By the end of the quarter, students will develop critical thinking and data science skills to be able to reason like an empirical economist.

Learning Outcomes

The successful student will be able to do the following after finishing the course:

1. Distinguish between statistical correlation and causal inference.
2. Interpret the results of common empirical analyses used in social science research.
3. Understand the various strengths and weaknesses of empirical research that uses observational economic and financial data from the real world.
4. Develop code in Python to apply the methods learned.

Prerequisites

There are no prereqs for this class. Students should have some familiarity with statistical programming, mathematical statistics, finance and economics. Experience with **econometrics** and Python programming is **strongly recommended** but not required.

Required Text

There are no required texts for this course. In fact, I encourage you to become adept at quickly finding reputable solutions to programming questions online. When they arise, I will provide you

with all the reading assignments on Canvas. I will also provide you with all lecture slides, but they will not contain all the details of what is discussed during the lecture - it is your responsibility to take notes in class.

Reference Text

The first third of the class is introductory econometrics and linear models. I recommend *Mostly Harmless Econometrics* by Angrist and Pischke and *Introductory Econometrics: A Modern Approach* by Wooldridge as reference books. For the other topics, I will post articles on Canvas for further reading. My slides, the notes that you take during lecture, quizzes, and the homework assignments are sufficient materials to prepare you for the exams.

Course Website
https://caltech.instructure.com/courses/8395

Assessment

Homework Assignments (7 scheduled)	40%
Attendance quizzes	20%
In-Class Midterm	20%
In-Class Final	20%

A : >93%   A- : 93-90%   B+ :90 – 87%   B :83-87%   B- :80-83%   C+ : <80%

Pass/Fail threshold: 60%

Attendance and Participation

Class attendance and participation are core tenets of this class. As Caltech students, I believe you can learn most, if not all, of the technical skills in this course by yourself. However, technical skills are only useful if you know when and how to apply them. Therefore, class time will be dedicated to deepening your understanding of how and why data science is used in economics and finance. I will periodically give unannounced quizzes in class to check for attendance. Please communicate any absences related to family emergencies, religious holidays, conferences, job interviews, or other events that prevent you from attending class as soon as possible.

Wellness Policy

If you find yourself overwhelmed or encountering other personal challenges during the term, please reach out to me so we can develop a plan for you to pursue success in this course in a healthy way. In addition, I encourage you to utilize Caltech's resources (see Canvas).

Students with Documented Disabilities

Students who may need an academic accommodation based on the impact of a disability must initiate the request with Caltech Accessibility Services for Students (CASS). Professional staff will evaluate the request with required documentation, recommend reasonable accommodations, and prepare an Accommodation Letter for faculty dated in the current quarter in which the

request is being made. Students should contact CASS as soon as possible, since timely notice is needed to coordinate accommodations. For more information: *https://cass.caltech.edu/*, *cass@caltech.edu*. If you are having difficulties with access or other challenges in the class you think might be related to a disability, but do not yet have a diagnosis, please feel free to reach out to CASS to learn more about resources.

Academic Integrity, Collaboration Policy, and ChatGPT

Caltech's Honor Code: "No member of the Caltech community shall take unfair advantage of any other member of the Caltech community." All instances of plagiarism or other academic misconduct will be referred to the *Board of Control* for undergraduates.

Integrity and programming: in this course, it **IS** acceptable to copy code from generative AI such as ChatGPT/Copilot– in fact, **it is the preferred way**. This is a somewhat controversial stance, and I will discuss my reasoning during Lecture 1. Chat is best used iteratively through the "trust but verify" method – you don't know how to solve a coding problem, generative AI will spit something out, you verify it's 90% correct, repeat. Very rarely will you come across a bug or issue that has not been solved by someone else – instead of re-inventing the wheel, you must learn how to adapt solutions to your own problems, somewhat quickly (as you will be asked to do in industry). I will ask you (pseudo)coding questions on the in-class midterm and final, so if you blindly turn in ChatGPT generated homeworks, you will probably do well on the homework portion of the grade, but almost surely fail the in-class exams.

My Status as a "Responsible Employee"

As a faculty member, I am required to notify the Institute's Equity and Title IX Office when I become aware of discrimination, sexual harassment, or sex- or gender-based misconduct involving our community members. If one of my students shares such an experience with me, I can help connect them to support resources but will not be able to keep that information confidential as part of fulfilling my responsibility to make sure my students are offered the opportunity to access information and support by the Institute. For more information, you can email *equity@caltech.edu*, go to *equity.caltech.edu*, or review the Institute's *Sex- and Gender-Based Misconduct Policy*.

If you have experienced such prohibited conduct and want to report it or speak to a confidential resource, consult the *Equity and Title IX Office's webpage on reporting* for guidance.

Caltech

Course Schedule
***No class on 4/1/2025

Week	Dates	Topics	HW Due
1	4/3	Correlation and causation, experimental framework	none
2	4/8 & 4/10	Interpreting linear models and inference	4/13
3	4/15 & 4/17	Statistical and non-statistical sources of bias	4/20
4	4/22 & 4/24	Scraping techniques, data sources and validation	4/27
5	4/29 & 5/1	Week 4 continuation + in-class Midterm (5/1)	Midterm
6	5/6 & 5/8	Clustering/unsupervised learning	5/11
7	5/13 & 5/15	Clustering/unsupervised learning	5/18
8	5/20 & 5/22	Text as data, natural language processing	5/25
9	5/27 & 5/29	Text as data, natural language processing	6/1
10	6/3 & 6/5	Catch-up and in-class Final (6/5)	Final

Academic Resources for Students

- **Tutoring:** The undergraduate dean's office provides a free peer tutoring service; If the course isn't listed, students can talk with the dean's office to arrange for a tutor; *https://deans.caltech.edu/*
- **Writing:** The Hixon Writing Center provides professional writing tutors as well as peer tutors, individual and group writing space, and additional resources; *https://writing.caltech.edu*
- **Registrar & FERPA:** The registrar can answer questions about degree progress, privacy of student records, and course enrollment procedures; *https://registrar.caltech.edu*. The website also lists *Option Representatives* for option-specific advising, policies, and information.
- **Library:** Borrow books, retrieve journal articles, receive guidance about research; *https://library.caltech.edu/*
- **Dean of Undergraduate Students**: Wide-ranging assistance addressing issues (academic and other) for undergraduates; *https://deans.caltech.edu*

Additional Resources for Students

- **Student Wellness Center:** Wide variety of health and wellbeing services; *https://wellness.caltech.edu/*
- **Counseling Services:** Free for all students, regardless of insurance plan; *https://counseling.caltech.edu*
- **Occupational Therapy:** Individual sessions and consultations on building healthy habits and routines, time management, planning and organization, and more. Free for all students; *https://ot.caltech.edu*
- **Center for Inclusion and Diversity**: Resources concerning navigating diversity and inclusion, including staff who can speak with students about challenges of harassment and discrimination; *https://diversity.caltech.edu/*
- **Title IX**: Caltech's Title IX Coordinator (*titleix@caltech.edu*) works with students on issues related to sexual harassment, sexual misconduct, and sex discrimination; *https://titleix.caltech.edu/*

- **Caltech Accessibility Services for Students**: The Accessibility Services Specialist works with students with temporary medical conditions, or mental, physical or learning disabilities on accommodation requests and services; *https://cass.caltech.edu*
- **Residential Support**: Resident Associates (RAs) and Residential Life Coordinators (RLCs) are also resources for TAs and students; *https://residentialexperience.caltech.edu/*
- **Career Advising and Experiential Learning:** Provides resources to help students make career decisions and implement career plans; *https://career.caltech.edu/*