Syllabus

Department of Economics and Business
CENTRAL EUROPEAN UNIVERSITY

Course title	**DATA ANALYSIS WITH AI – PHD TRACK**
Instructor	Gábor Békés
Email	bekesg@ceu.edu
Office	by appointment
Credits	1 US credits (2 ECTS credits)
Module	**(***in which the course is offered)*
Term	Winter 2024-2025
Course level	Phd
Prerequisites equivalent.	none
Course drop	

1. COURSE DESCRIPTION

Content. This course will equip students, who are already versed in core data analysis methods, with experience to harness AI technologies to improve productivity. We will focus on using large language models (LLMs) such as OpenAI's ChatGPT, and Google Gemini) to carry out tasks in data analysis. The course will focus on data extraction and wrangling, data exploration and descriptive statistics, selection of estimation methods, and creating reports. The course involves a great deal of hands-on practice with weekly case studies and assignments.

Relevance. AI is everywhere and has become essential, most analytical work will be using it.

Expected background.

- Data Analysis 1 and 2, Introductory Econometrics or equivalent (descriptive statistics, regressions)
- Students are expected to have (a) low intermediate knowledge of Python or R or (b) a basic knowledge of Python and intermediate knowledge of Stata
- Students will benefit from having familiarity with VScode or Rstudio or similar IDE
- They shall have a subscription of the professional version of an AI service (such as ChatGPT 4).

2. LEARNING OUTCOMES

Key outcomes. By the end of the course, students will be able to

- Gain experience using at least ChatGPT 4.0. level AI for their work
- Build AI in coding practice including data wrangling, description and reporting
- Have a basic idea of use cases when AI assistance is (1) greatly useful, (2) helpful, (3) *currently* problematic
- Have an understanding of resources to follow for updates

Other outcomes. The course will also help develop skills in the following areas.

Learning Area	Learning Outcome
Critical Thinking	Access replies from AI
Quantitative Reasoning	Have the ability to present empirical analysis and write short reports with data.
Technology Skills	Introduction to using ChatGPT 4
Interpersonal Communication Skills	-
Management Knowledge and Skills	Managing a workflow with AI
Cultural Sensitivity and Diversity	-
Ethics and Social Responsibility	-

3. READING LIST

- [BK] Békés, Gábor and Gábor Kézdi, "Data Analysis for Business, Economics, and Policy". Cambridge University Press, 2021: Hardcover, paperback and e-book. [Textbook website, Cambridge UP],
- [Mollick] Mollick, Ethan "Co-Intelligence: Living and Working with AI", Pengium Random House, 2024
- [Korinek] Korinek, Anton: "Generative AI for Economic Research: Use Cases and Implications for Economists", JOURNAL OF ECONOMIC LITERATURE, VOL. 61, NO. 4, DECEMBER 2023
- Additional material on staggered difference in differences, such as https://github.com/Mixtape-Sessions/Advanced-DID/blob/main/Slides/02-staggered.pdf

4. TEACHING METHOD AND LEARNING ACTIVITIES

The course will involve 1 lecture and 5 lab sessions.

- This is the intermediate data course of the program. It will be built on introductory data analysis courses.
- Learning objectives will be achieved by students experimenting with prompting in class and in assignments.
- Assignments will aim at consolidation of the experience

Class 5 and 6 will focus on testing benefits of AI use

5. ASSESSMENT

Grading will be based on the total score out of 100, in line with CEU Department of Economics and Business grading guidelines. In particular,

a. The median student can expect to get a B+
b. Probably not more than 1/3 of the students can expect to get an A or A-
c. The passing grade is 50%.

There will be no exam, but several assignments. Overall, students will be expected to read materials, and experiment with AI for project. Expected extra classroom workload is 3-8 hours per week (depending on background).

The final grade is based on:

- Assignments to be submitted in between classes. (60%)
- A final report, an essay summarizing experience with using AI for data analysis (30%)
- In class participation (10%)

6. TECHNICAL REQUIREMENTS

There is a technical requirement. You are expected to come prepared. Please consult with your more experienced colleagues re setup if necessary. One good starting point is https://paulgp.substack.com/p/setting-up-github-copilot-and-vscode

- Python / R users --- Python or R environments set up. IDE that includes copilot (such as RStudio, VScode, Anaconda).
- Stata users – Have Stata 16+ on your laptop. Have VScode as well setup for Stata

ChatGPT 4o or Claude subscription

8. TOPIC OUTLINE AND SCHEDULE

WEEK BY WEEK

Session	Topics	Readings /expected background
1	**Intro to LLMs,** Part 1: the "stochastic parrot" concept, overview of use cases, types of LLMs, "cyborgs" vs "centaurs", idea generation Part 2: How coding assistance help. Repetitive tasks. Writing functions. Importance of names. Case study: Survey design	Mollick BK Chapter 1
2	**Understanding datasets** Loading in data, describing variables, dealing with data problems, devising cleaning strategies. Case study: Football managers and team performance; swimming laps	BK Chapter 2-3 Korinek
3	**Exploratory data analysis** Descriptive statistics, graphs. Correlations and tests. Writing reports Case study: Earnings and gender gap, football managers and team performance	BK Chapter 3
4	**Understanding context, creating multivariate regression** Case study: Earnings and gender gap, world value survey	BK Chapter 10
5	**Calculations and simulations** Finding appropriate statistics, calculations. Doing simulations such as bootstrap. Case study: measuring and benchmarking diversity in football teams.	BK chapter 5
6	**Designing event studies** Case study: TBD	BK chapter 22,24 Handouts

9. SHORT BIO OF THE INSTRUCTOR

I am an associate professor at the Department of Economics and Business of Central European University. I'm also a senior research fellow at the KRTK Institute of Economics and a research associate at the London-based Center for Economic Policy Research (CEPR). I received my Ph.D. degree in 2007 from the Central European University and an MSc in Economics from the London School of Economics. My research is focused on international economics; economic geography and applied IO. My research has been published among others by *Management Science, Global Strategy Journal, Journal of International Economics, Regional Science and Urban Economics, and Economic Policy* and have authored blogs and commentary on VOXEU, de facto and other outlets. I have worked on projects for the European Commission and the ECB.

Social media: Twitter: @GaborBekes

My textbook "Data Analysis for Business, Economics and Policy " with Gábor Kézdi (U Michigan) was published by the Cambridge University Press in 2021. You may learn more by following the textbook on Facebook or Twitter.