Using AI for Data Analysis

ECON 490

Taylor Mackay || Email: tmackay@fullerton.edu

Slides Overview

In these slides, we'll discuss:

- General suggestions for using AI effectively
- Specific suggestions for using AI to help with data analysis

Setting the Stage

I teach ECON 100 – every semester, someone submits a copy-pasted ChatGPT essay

- This is ***not*** what I mean by using AI in this course
- AI is ***not*** a contractor that you can “hire” to independently complete tasks for you

Instead, think about AI as an ***assistant*** to help complete tasks

- As an assistant, AI can make some tasks much faster…
- But it requires your supervision!

General Tips for Using AI Effectively

Think carefully about what your desired output is ***before*** asking AI for help

Whatever level of detail you're providing in a prompt, add more!

- **Bad:** *"Give me R code to run this regression…"*
- **Good:** *"I am economics student working on a data analysis project using R and tidyverse. Can you give me code to estimate a regression model where…"*

Work iteratively – provide a prompt, ask for updates, provide another prompt, …

Using AI for Feedback

One place where AI can be helpful is evaluating your work. For example:

I am economics student working on a research project. The first step of the project is completing a proposal worksheet laying out my planned project. I have included the worksheet below with my answers. Can you provide 5 bullet pointed suggestions for ways to improve my proposal?

With prompts like this, judgement is important – think of feedback as a ***starting point***

- From feedback above, each point was superficially relevant but 2-3 were “fluffy”
- 2 points were very helpful (thinking about omitted variables and data cleaning)

General Note on Feedback and Evaluating your Work

In college, once you finish a project or class, you're done!

- If a task didn't go well, or needed changes, you don't have to worry about it
- This isn't how things work in most jobs

This difference can make the adjustment to working rocky for some people

- You're often doing similar tasks repeatedly…
- And co-workers and managers expect you to improve!

Learning to seek out and use feedback proactively is one of the most useful skills you can acquire in any job

AI for Studying – Examples

OVB Examples: *I am an economics student studying omitted variables bias in OLS models. Can you provide me 5 real-world examples of regressions using economics topics where OVB might be an important consideration?*

Interpreting Logs in Regressions: *I am an economics student studying OLS regression. I am considering a regression with Y as my outcome variable and X as my explanatory variable. Can you explain to me how using either the log or the level of Y and X changes the interpretation of my estimated regression coefficient on X?*

Interpreting DiD Models: *I am an economics student trying to understand a paper that uses a difference-in-differences model. The outcome variable in the primary regression from the paper is employment and the explanatory variable is minimum wages. The regression includes state and year fixed effects. Can you tell me how I should interpret the estimated coefficient on minimum wages?*

Using AI for Data Analysis

Data analysis in R (or any language) can be time intensive – using AI can improve your productivity on two main margins:

1. Helping you with errors & sticking points – *How does this function work?*
2. Providing you with “canned” code for clearly defined tasks (more on next slide)

There’s a reason we’re just talking about this now (in Week 7)

- Using AI for data analysis requires you to be clear about conceptual goals
- What is the relationship you want to estimate? What approach makes sense?

Specific Suggestions

AI (specifically ChatGPT) is great at adding options to existing R code

- We'll see this in Coding Activity 4 and creating graphs in R
- There's tons of options for editing graphs in R – let AI sort through them for you

Given ***narrowly defined*** data cleaning tasks, AI can provide R code

- Again, the more detail, the better → best option is uploading data
- E.g., "I'd like to filter out variables meeting X condition using tidyverse, and then calculate the state-by-year averages of variables A and B."

Don't just copy paste code – run things line-by-line and make sure you "own" it

When NOT to Use AI

In general, your background knowledge with a topic acts as “guardrails” for AI

- The more you know about a topic, the better you can filter AI feedback
- Doesn’t mean you can’t use AI to learn about new topics… but be cautious

Specific things AI (currently) is not consistently good at:

- Doing literature reviews
- Writing in a non-annoying “one on hand, on the other…” voice
- Doing anything in a fully self-directed or unsupervised fashion

What Tools to Use?

Several options here, including:

- ChatGPT Pro for $20 (worth it) or ChatGPT (free with an account)
- Perplexity lets you try lots of different tools
- Google Gemini, Claude, etc.

Which one to choose?

- Relative trade-offs change frequently, so there’s not a single “best” answer
- Free ChatGPT is likely good enough for most cases, however…
- Free version of Perplexity is great at finding links, Claude is better at writing

General AI Guidelines for this Class

Be clear about when and where you're using AI

- Some assignments might explicitly tell you to use ChatGPT (e.g., Coding Activity 4)
- For your capstone data analysis code, you're welcome to use AI to help you write your code – just make sure you know exactly what your code does

For any writing assignment (remote homework, capstone paper, etc.), all writing that isn't an explicit citation from another source in quotes should be in your own words

- Copying and pasting AI output and claiming it as your own writing is plagiarism
- You are welcome to use AI as an assistant while writing (e.g., outlining, etc.)

General rule – if you're not sure if an application is appropriate, just ask me!