In-person session 9

March 7, 2024

PMAP 8521: Program evaluation
Andrew Young School of Policy Studies

Plan for today

Quick questions

LLMs

Matching and IPW

Two-way fixed effects

Quick questions

Can you combine design-based and model-based identification?

Like diff-in-diff with a DAG?

DIDID(IDIDID)?

The effect of mandatory maternity benefits on wages

New Jersey implements policy; Pennsylvania doesn't

Only applies to married women who have kids

Married women 20–40 -
single men/unmarried women/older women
in NJ and PA

TABLE 3—DDD ESTIMATES OF THE IMPACT OF STATE MANDATES ON HOURLY WAGES

Location/year	Before law change	After law change	Time difference for location
A. *Treatment Individuals: Married Women, 20–40 Years Old:*			
Experimental states	1.547 (0.012) [1,400]	1.513 (0.012) [1,496]	−0.034 (0.017)
Nonexperimental states	1.369 (0.010) [1,480]	1.397 (0.010) [1,640]	0.028 (0.014)
Location difference at a point in time:	0.178 (0.016)	0.116 (0.015)	
Difference-in-difference:	−0.062 (0.022)		
B. *Control Group: Over 40 and Single Males 20–40:*			
Experimental states	1.759 (0.007) [5,624]	1.748 (0.007) [5,407]	−0.011 (0.010)
Nonexperimental states	1.630 (0.007) [4,959]	1.627 (0.007) [4,928]	−0.003 (0.010)
Location difference at a point in time:	0.129 (0.010)	0.121 (0.010)	
Difference-in-difference:	−0.008: (0.014)		
DDD:	**−0.054** **(0.026)**		

The DAG from the test

LLMs

How have you used LLMs like ChatGPT?

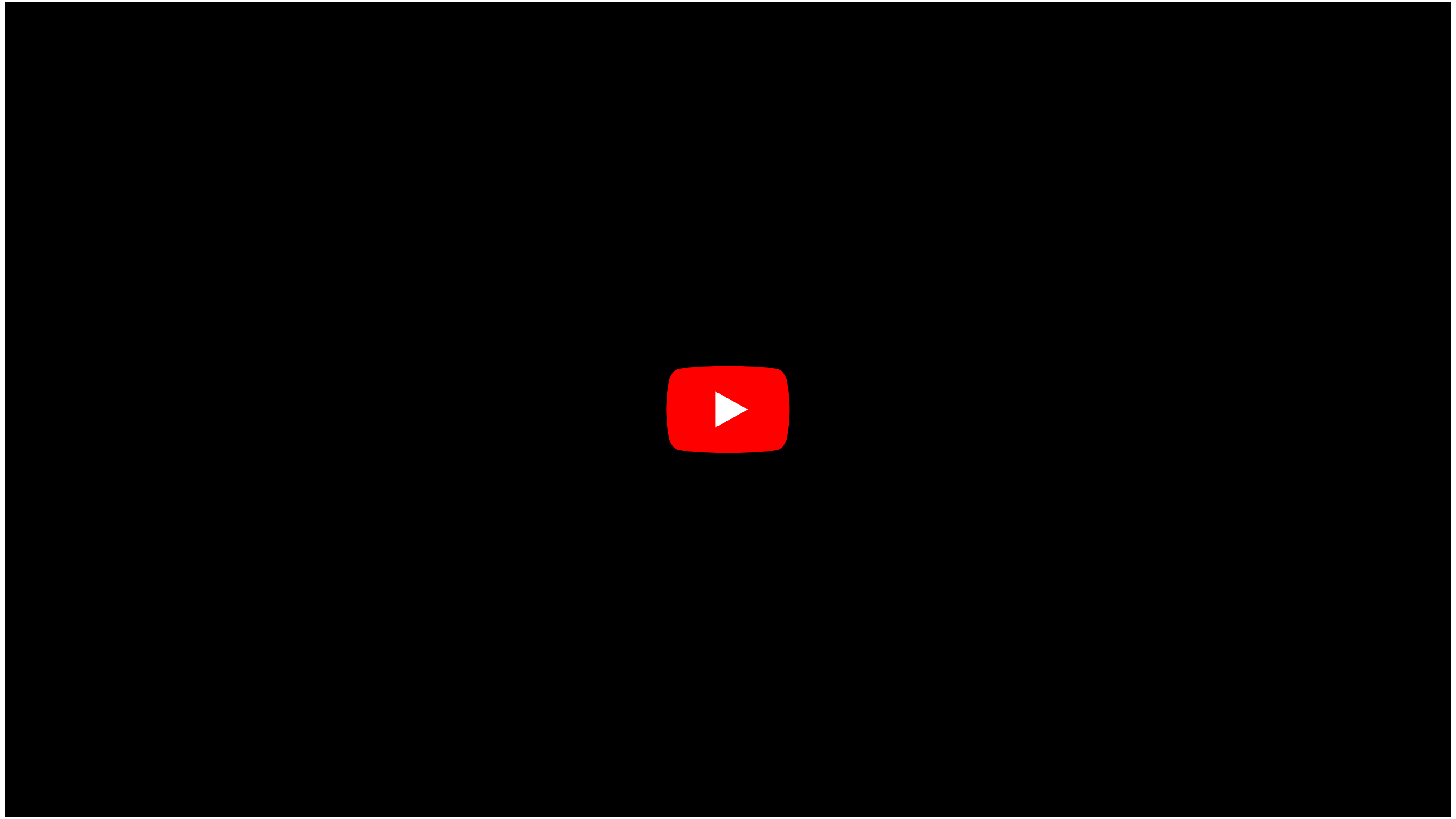
What worries have you had?

Can we use LLMs like ChatGPT?

How do we use them?

Is it okay to use them?

LLMs are not magical. They're stats.



So many ethical issues!

Environmental concerns

"Environmental Impact of Large Language Models" and "The mounting human and environmental costs of generative AI" and "AI water footprint suggests that large language models are thirsty"

Racial and gender bias

"AI chatbots use racist stereotypes even after anti-racism training"

Human toll

"OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic"

Stolen training data

"ChatGPT Stole Your Work. So What Are You Going to Do?" and "Congress Wants Tech Companies to Pay Up for AI Training Data" and "ChatGPT can leak training data, violate privacy"

Ouroboros effect

"Meet the Serbian DJ Running an AI Clickbait Business" and "The Perfect Webpage"

Making stuff up

"AI models make stuff up. How can hallucinations be controlled?"

You need to figure out your own ethics.

LLMs and programming

GitHub Copilot specifically trained on code

Works *surprisingly* well

But it's dangerous if you don't know what you're doing!

Copilot skills

Talking to Copilot requires special skills and practice!

***Rep*roducible *ex*amples!**

Reprex slides

GitHub Gists

Things Copilot is good at

Explaining and annotating code

Translating between languages

Generating boilerplate/starter code

Cleaning and rewriting code

Demonstration!

Matching and IPW

Two-way fixed effects (TWFE)

Two states: Alabama vs. Arkansas

$$\begin{aligned}\text{Mortality} = \beta_0 + \beta_1 \text{ Alabama} + \beta_2 \text{ After 1975} + \\ \beta_3 \text{ (Alabama} \times \text{After 1975)}\end{aligned}$$

All states: `Treatment == 1`
if legal for 18-20-year-olds to drink

$$\text{Mortality} \ = \ \beta_0 + \beta_1 \ \text{Treatment} + \beta_2 \ \text{State} + \beta_3 \ \text{Year}$$

$$\text{Mortality} = \beta_0 + \beta_1 \text{ Alabama} + \beta_2 \text{ After 1975} + \beta_3 \text{ (Alabama} \times \text{After 1975)}$$

vs.

$$\text{Mortality} = \beta_0 + \beta_1 \text{ Treatment} + \beta_2 \text{ State} + \beta_3 \text{ Year}$$

$$\text{Mortality} = \beta_0 + \beta_1 \text{ Alabama} + \beta_2 \text{ After 1975} + \beta_3 \text{ (Alabama} \times \text{After 1975)}$$

vs.

$$\text{Mortality} = \beta_0 + \beta_1 \text{ Treatment} + \beta_2 \text{ State} + \beta_3 \text{ Year}$$

vs.

$$\text{Mortality} = \beta_0 + \beta_1 \text{ Treatment} + \beta_2 \text{ State} + \beta_3 \text{ Year} + \beta_4 \text{ (State} \times \text{Year)}$$

TABLE 5.2
Regression DD estimates of MLDA effects on death rates

Dependent variable	(1)	(2)	(3)	(4)
All deaths	10.80	8.47	12.41	9.65
	(4.59)	(5.10)	(4.60)	(4.64)
Motor vehicle accidents	7.59	6.64	7.50	6.46
	(2.50)	(2.66)	(2.27)	(2.24)
Suicide	.59	.47	1.49	1.26
	(.59)	(.79)	(.88)	(.89)
All internal causes	1.33	.08	1.89	1.28
	(1.59)	(1.93)	(1.78)	(1.45)
State trends	No	Yes	No	Yes
Weights	No	No	Yes	Yes

Notes: This table reports regression DD estimates of minimum legal drinking age (MLDA) effects on the death rates (per 100,000) of 18–20-year-olds. The table shows coefficients on the proportion of legal drinkers by state and year from models controlling for state and year effects. The models used to construct the estimates in columns (2) and (4) include state-specific linear time trends. Columns (3) and (4) show weighted least squares estimates, weighting by state population. The sample size is 714. Standard errors are reported in parentheses.

$$\begin{aligned}\text{Donation rate} = \beta_0 + \beta_1 \text{ California} + \beta_2 \text{ After Q22011} +\\ \beta_3 \,(\text{California} \times \text{After Q22011})\end{aligned}$$

vs.

$$\begin{aligned}\text{Donation rate} = \ & \beta_0 + {\color{red}\beta_1} \text{ Treatment} +\\ & \beta_2 \text{ State} + \beta_3 \text{ Quarter}\end{aligned}$$

What about this staggered treatment stuff?

See this