Making Do-Calculus Accessible in Econometrics

DAGs as Complementary Tools for Causal Estimation

Patrick Beal

Brigham Young University

December 2025

The Problem: Causal Inference Without Guidance

Question: How should we choose which variables to control for?

- Traditional econometrics: Controlling for all omitted variables
- Control for everything (Kitchen-sink): Can increase bias

DAGs provide a formal framework for encoding causal assumptions and deriving identification conditions.

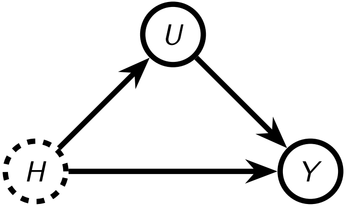
What are DAGs? (Directed Acyclic Graphs)

Formal Definition:

- Nodes = Variables
- Arrows = Direct causal effects
- Dashed nodes = Unobserved variables
- Acyclic = No feedback loops

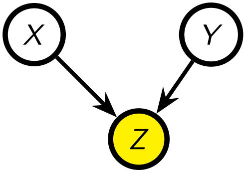
Key Concept: D-Separation

- Tests conditional independence
- Determines which variables to control
- Automatically validated in data



Example 1: Collider Bias (Synthetic)

DAG: Collider Structure



Fact 1: $X \perp\!\!\!\perp Y$
Problem: $X \not\perp\!\!\!\perp Y \mid Z$

Data Generating Process:

$$X \sim N(0, 1)$$
$$Y \sim N(0, 1)$$
$$Z = 0.5X + 0.5Y + \epsilon$$

Key Insight:

Controlling for a collider creates spurious correlation.

Simulation Result:

Method	Bias	Coverage
Naive (no control)	0.000	95%
Control Z	-0.248	0%

True effect $= 0$. Controlling for Z produces bias

Example 2: Frontdoor Criterion (Synthetic Bellemare)

Research Question

Does allowing riders to authorize shared rides affect tipping?

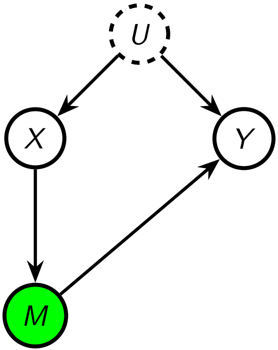
The Challenge:

- Generosity affects both treatment and outcome
- Generosity is unobserved

Data Generating Process:

$$\begin{aligned} U &\sim N(50, 15^2) \\ P(X=1) &= \text{logit}(-2 + 0.05U) \\ M &= 60 - 12X + \epsilon_M \\ Y &= 5 + 0.1U - 0.08M + \epsilon_Y \end{aligned}$$

DAG: Frontdoor Structure



X = Share authorized, M = Actually shared
Y = Tip, U = Unobserved generosity

Example 2: Frontdoor Criterion (Synthetic Bellemare)

How Frontdoor Works

- Stage 1: Estimate $X \rightarrow M$ effect
- Stage 2: Estimate $M \rightarrow Y$ effect
- Combine: Effect $= \beta_1 \times \beta_2$

Why It Works

- No confounding of $X \rightarrow M$ path
- M blocks all backdoor paths
- Unobserved U becomes irrelevant

Example 2 Results: Frontdoor vs. Naive

Simulation Results (Synthetic Bellemare, n=100,000)

Estimator	Effect	Interpretation
Naive OLS	+0.03pp	Biased
Standard Frontdoor	-0.95pp	Accurate
True Effect	-0.96pp	—

Example 2 (Continued): Comparing Estimation Approaches

Do Both Approaches Give Similar Results?

Estimator	Effect	Error from True
Parametric Frontdoor	-0.95pp	1.0%
Nonparametric Frontdoor	-0.84pp	12.3%
True Effect	-0.96pp	—

Example 2 (Continued): Comparing Estimation Approaches

Parametric

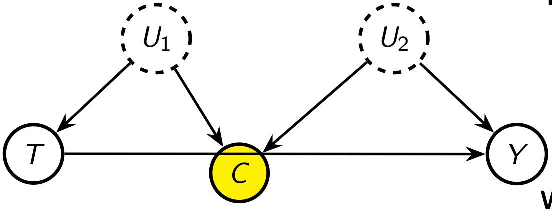
- Stage 1: Linear regression
- Stage 2: Linear regression
- Delta method SEs
- More efficient

Nonparametric

- Bin mediator values
- Estimate densities
- Pearl's formula directly
- No linearity assumption

Example 3: M-Bias (Synthetic)

The M-Bias Structure



C is a **pre-treatment collider**

$$\begin{aligned} U_1 &\sim N(0,1) \\ U_2 &\sim N(0,1) \\ C &= 0.6U_1 + 0.4U_2 + \epsilon_C \\ T &= 0.5U_1 + 0.3C + \epsilon_T \\ Y &= 0.4T + 0.5U_2 + \epsilon_Y \end{aligned}$$

The Problem

- Traditional wisdom: "Control for C"
- DAG analysis: Controlling for C induces bias

Why?

- Controlling for C opens blocked path

Insight:

Traditional guidance contradicts the DAG, leads to bias

Example 3 Results: Why DAG Guidance Matters

Simulation Results: Bias for Different Strategies

Estimator	Bias	vs. Naive
Naive (no controls)	0.087	Baseline
Traditional (control C)	0.215	+148%
DAG-Guided (no control)	0.087	Minimal Bias
True effect	0.000	—

Key Takeaway:

Controlling for a pre-treatment collider can **double the bias**!

Making Assumptions Testable

Strategy: Use data to validate DAG assumptions

D-Separation Tests

- Test conditional independence
- Expected to hold if DAG is correct
- Apply to synthetic and real data

Placebo Tests

- Test effects that should be zero
- Test relationships that shouldn't exist
- Validates causal structure

When to Use DAGs in Econometrics

DAGs and Traditional Methods are Complementary

DAGs Provide	Traditional Provides	Together:
Systematic identification	Efficient estimation	Robust inference
Assumption validation	Hypothesis testing	Credible results
Covariate selection	Standard errors	Publishable

Best Practice:

1. Draw the DAG (encode causal assumptions)
2. Use d-separation (determine controls)
3. Run regression (estimate effect)

Key Takeaways

1. DAGs provide **systematic rules** for covariate selection
2. Traditional guidance can increase bias (colliders, M-bias)
3. D-separation enables identification in new contexts (frontdoor)
4. Assumptions are testable with falsification tests
5. Combine DAGs with traditional econometrics for best results

Questions?