

Causal Inference in Transportation Demand Models

Hassan Obeid, Mohamed Amine Bouzaghrane, Timothy Brathwaite,
Akshay Vij, Joan Walker

April 23rd, 2020
Berkeley, California



Outline

- What is causal inference?
- Key takeaways
- How important is the data generating process?
- How does a latent confounder affect the workflow?



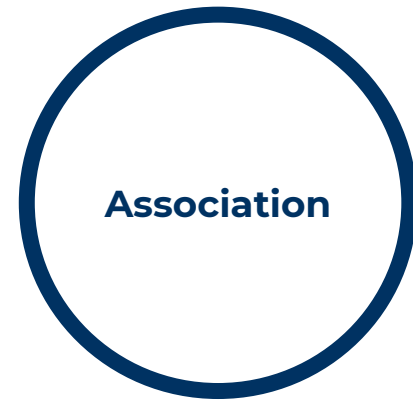
What is Causal Inference?

Use of data and assumptions to draw conclusions about causal relations between **external interventions** and outcomes.



Predict Y after **do** $X = x$

VS



Predict Y after **given** $X = x$

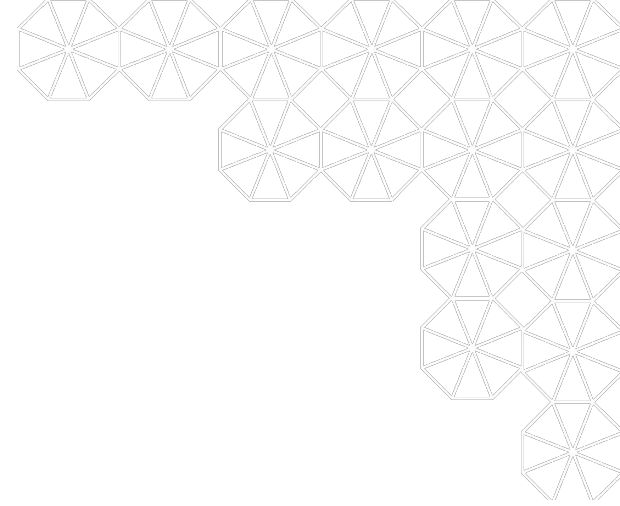
No causal inference without assumptions (even if not stated)

Why is it important to us?

“Travel demand models are used to aid in the evaluation of alternative policies. The purpose of the models is to predict the consequences of alternative policies or plans. [...] A model can duplicate the data perfectly, but may serve no useful purpose for prediction if it represents erroneous behavioral assumptions”. - **Ben Akiva**, '73

Key Takeaways

- Outcome model is not enough
 - Pay attention to the data generation process
- Dealing with latent confounding is an open problem



Importance of Causal Structure

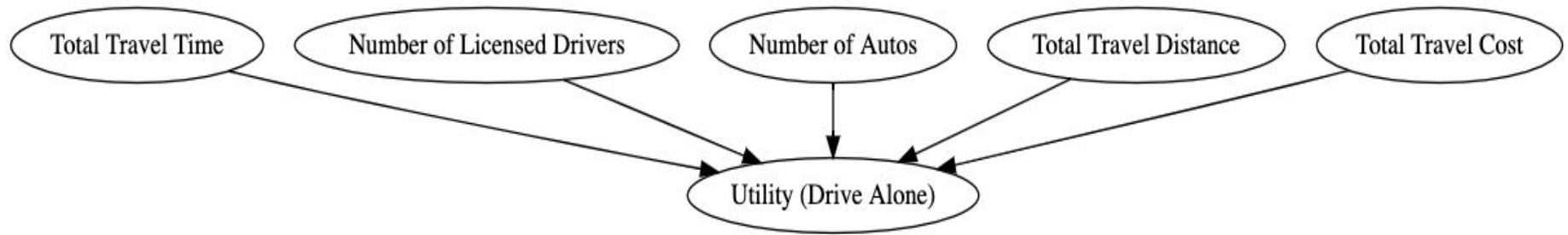
Causal Inference Workflow



- Always draw your assumptions first
- Test your assumptions
- Estimate your models
- Draw causal conclusions

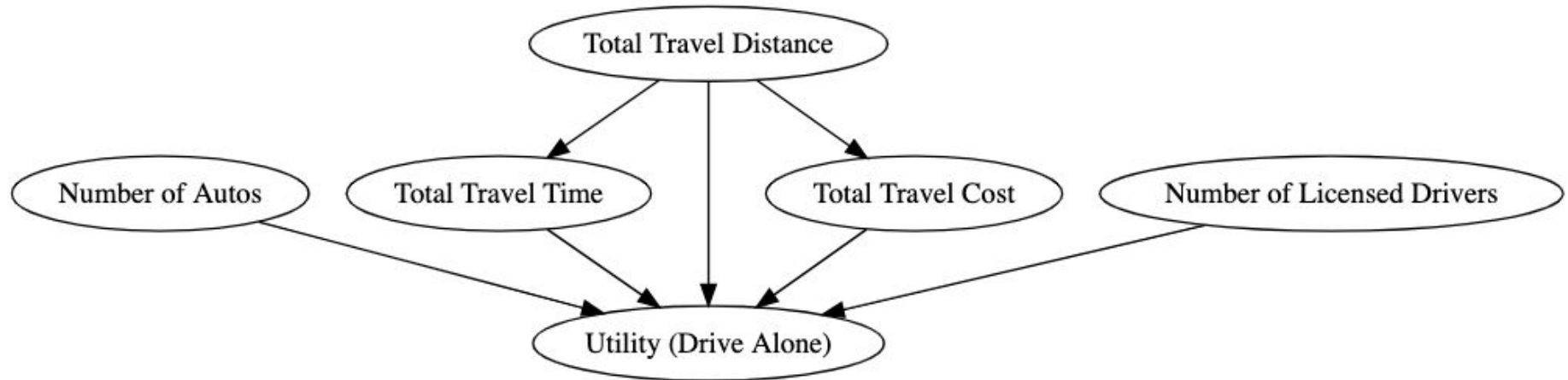
Importance of Causal Structure

1.



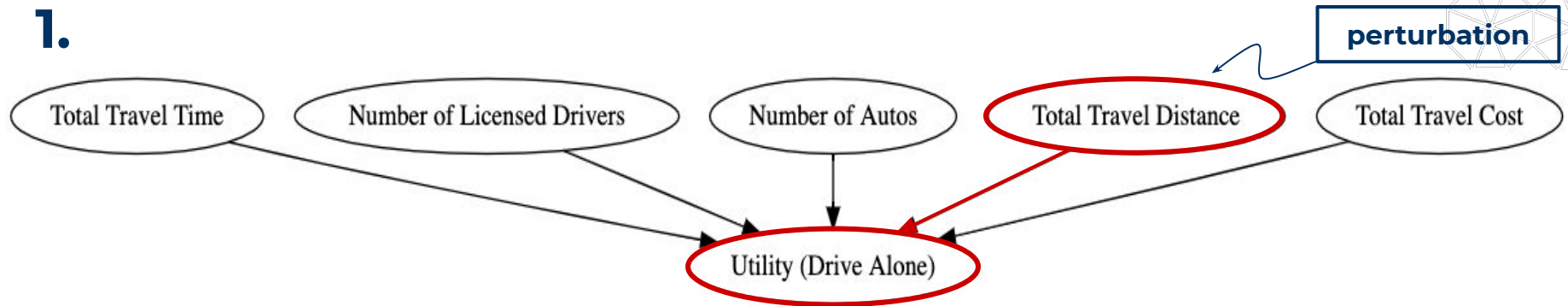
VS

2.



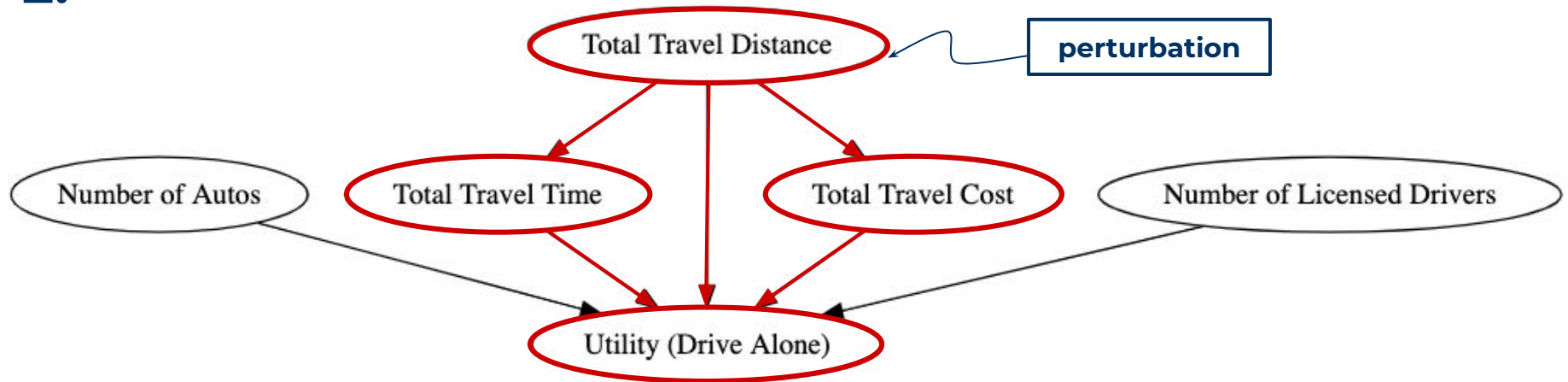
Importance of Causal Structure

1.

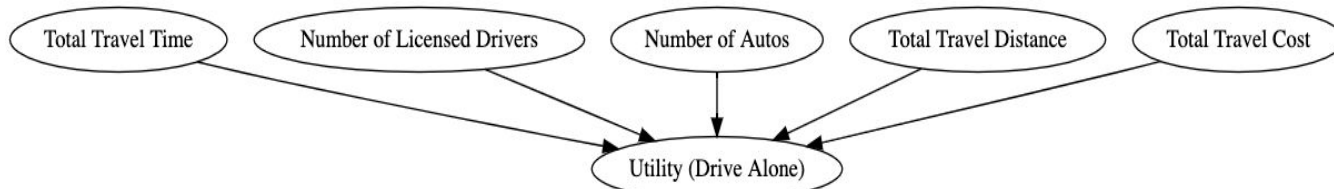


VS

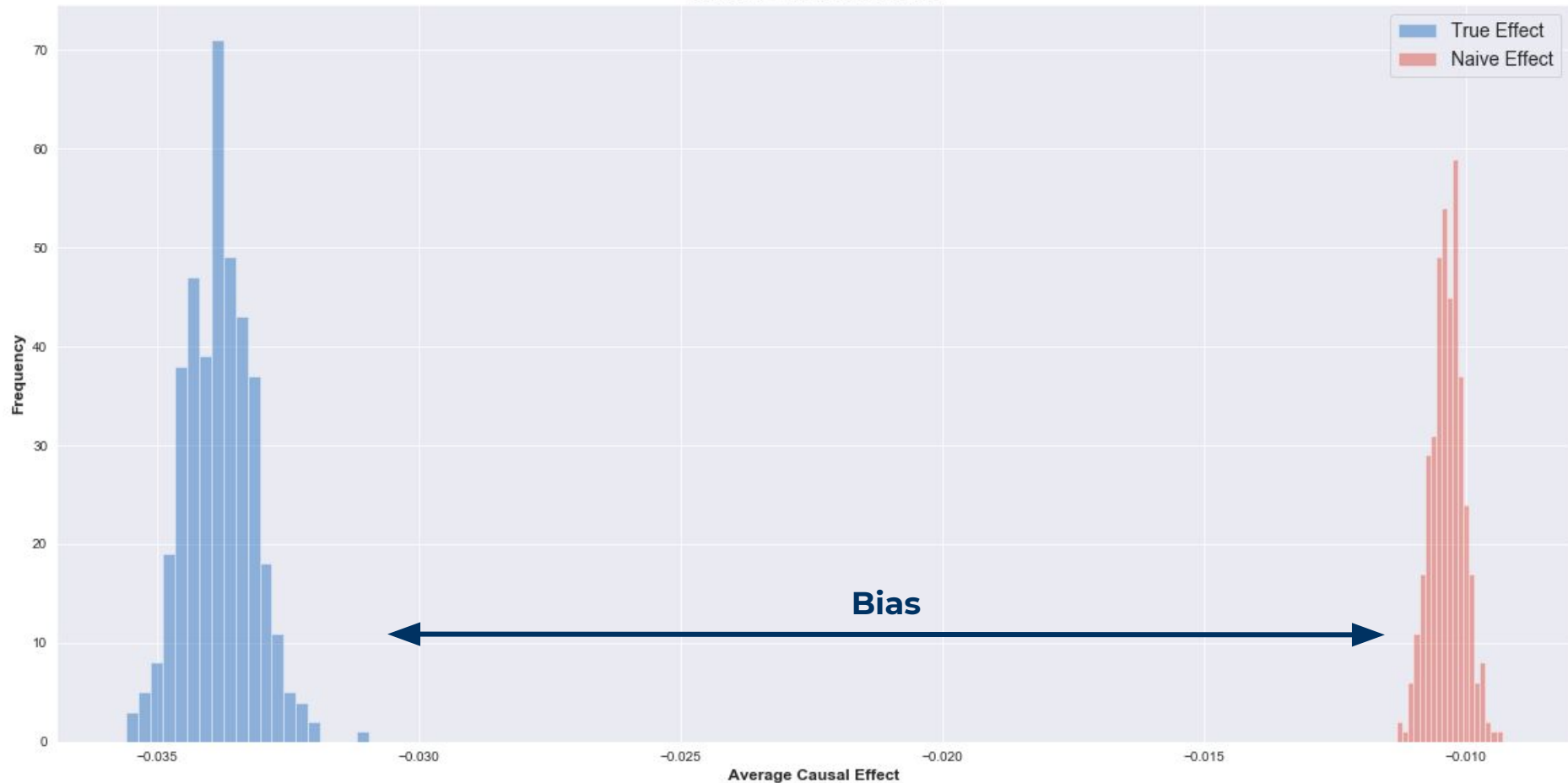
2.

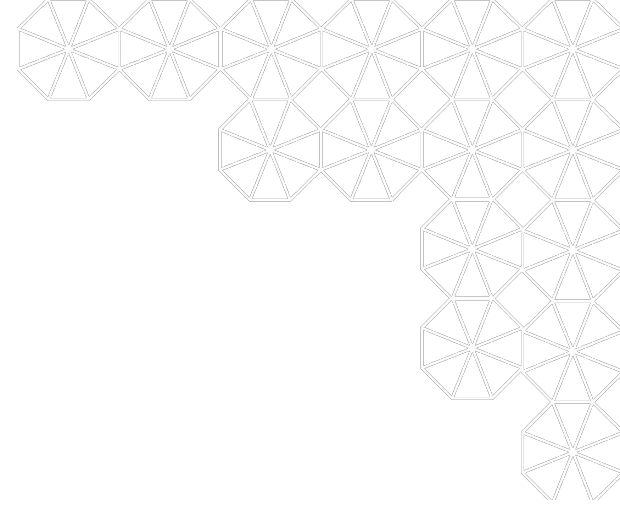


Importance of Causal Structure



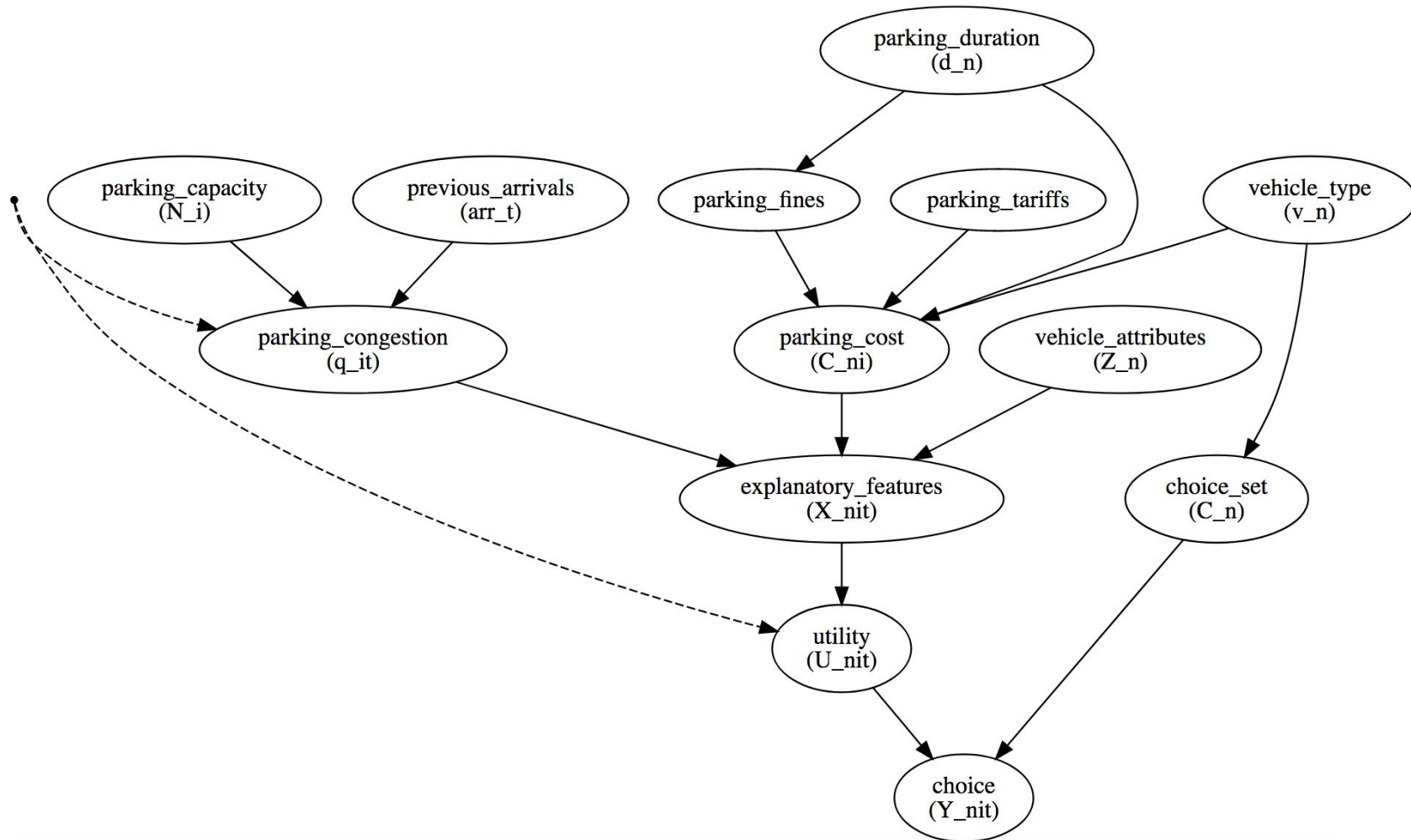
True Effect vs. Naive Effect



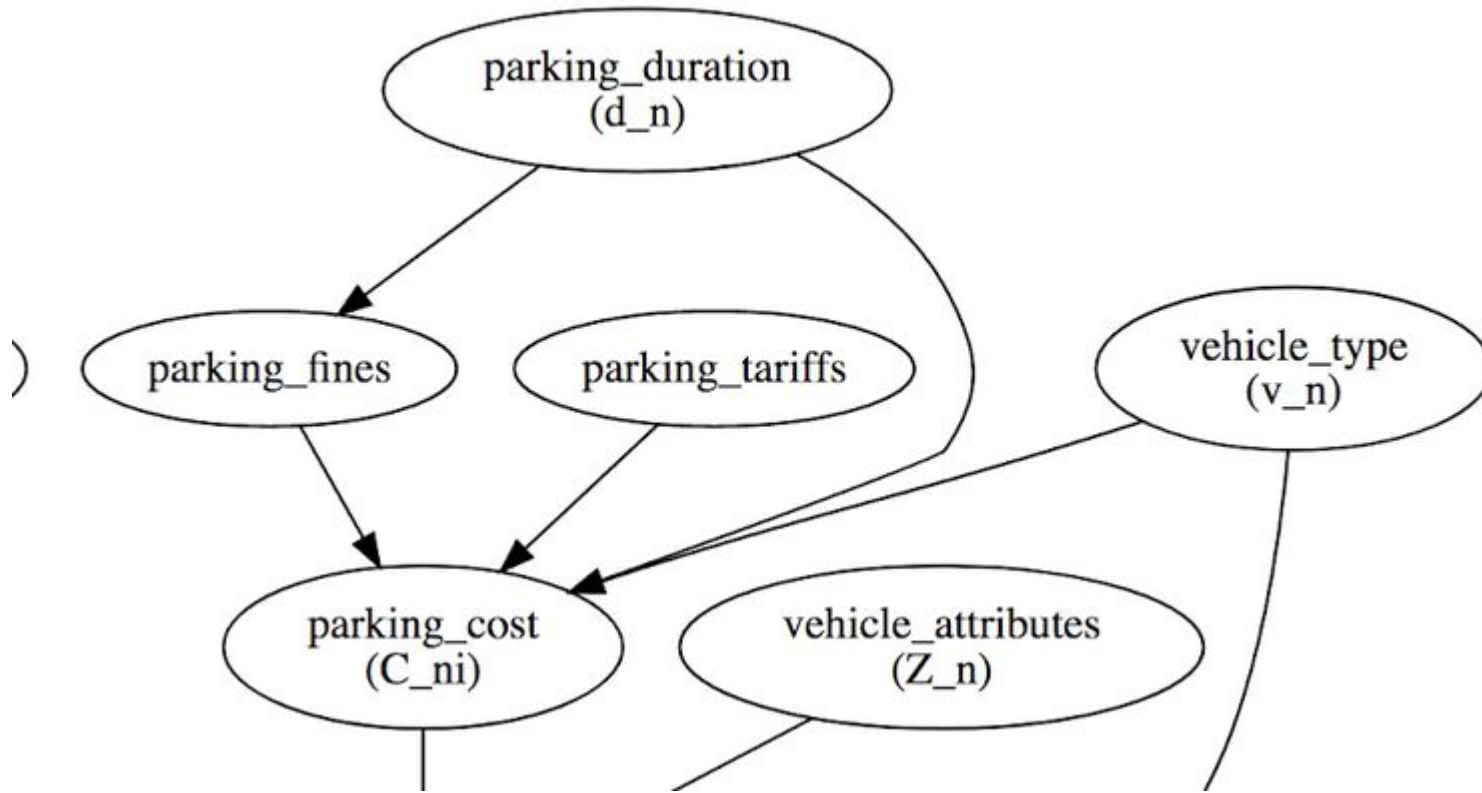


**What data generating
process do people
assume?**

Chiara et al. (2020)



Chiara et al. (2020) - Problems?



Testing

Assumptions + **R**epresentation = Causal **G**raph

↑
Testing

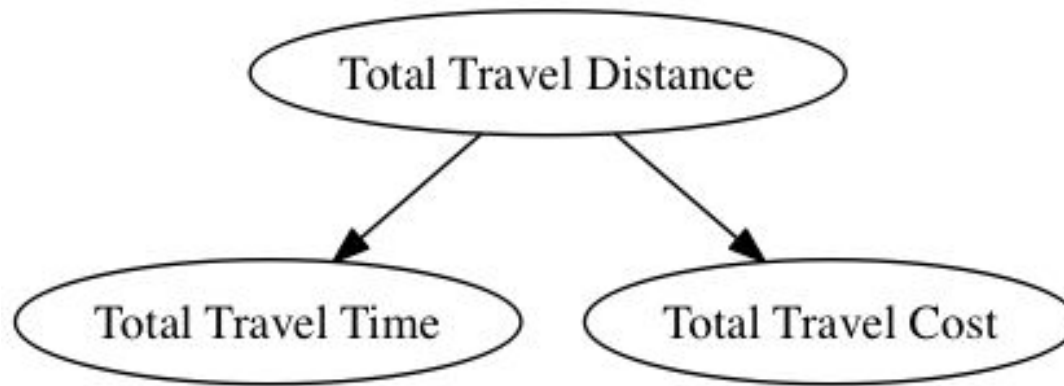
=

↑
Testing

Testing

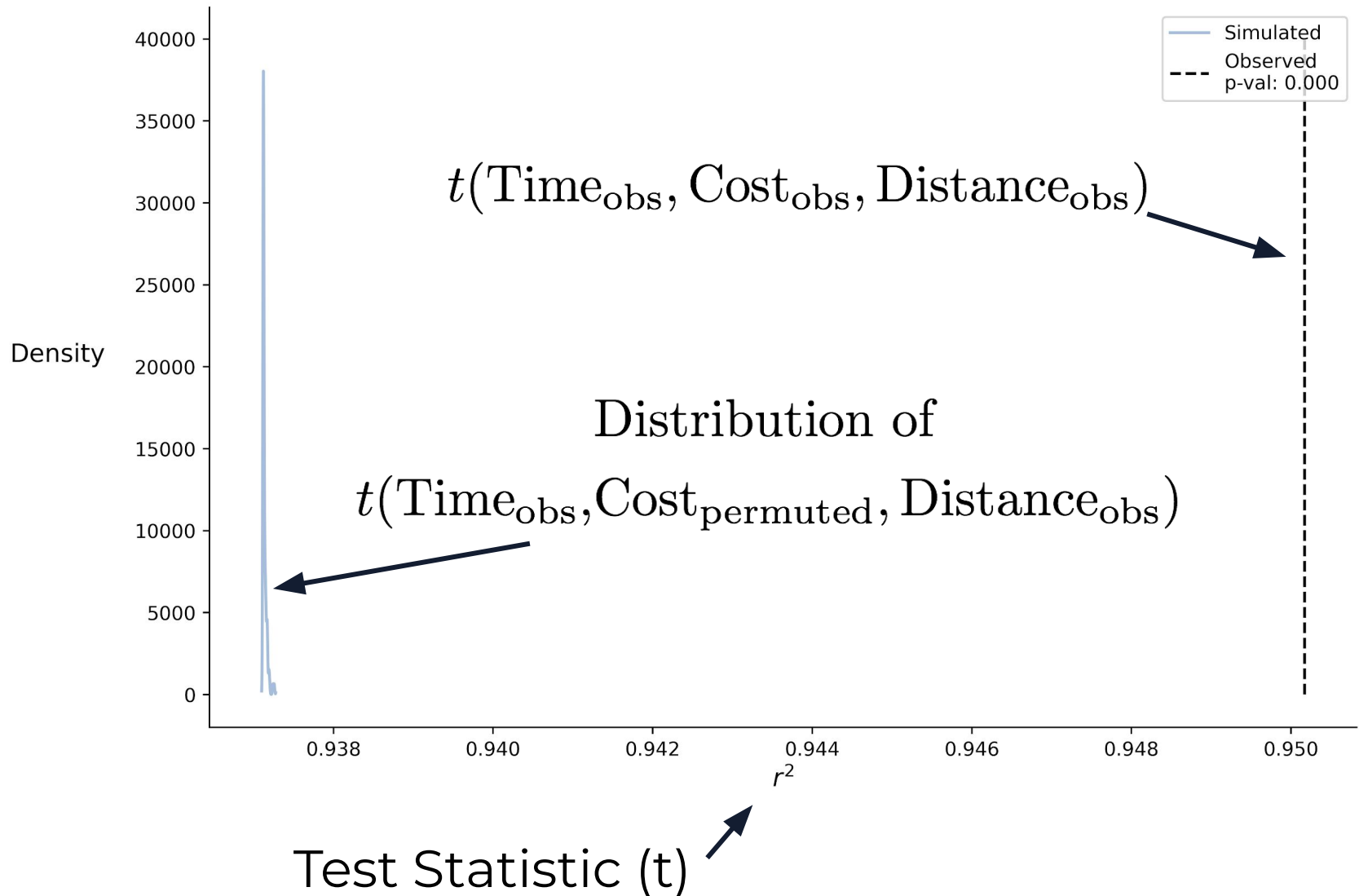
- Two main types of test needed to validate/falsify a hypothesized causal graph
 - Marginal independence testing
 - Conditional independence testing

Testing (Example)



implies
Time \perp Cost | Distance

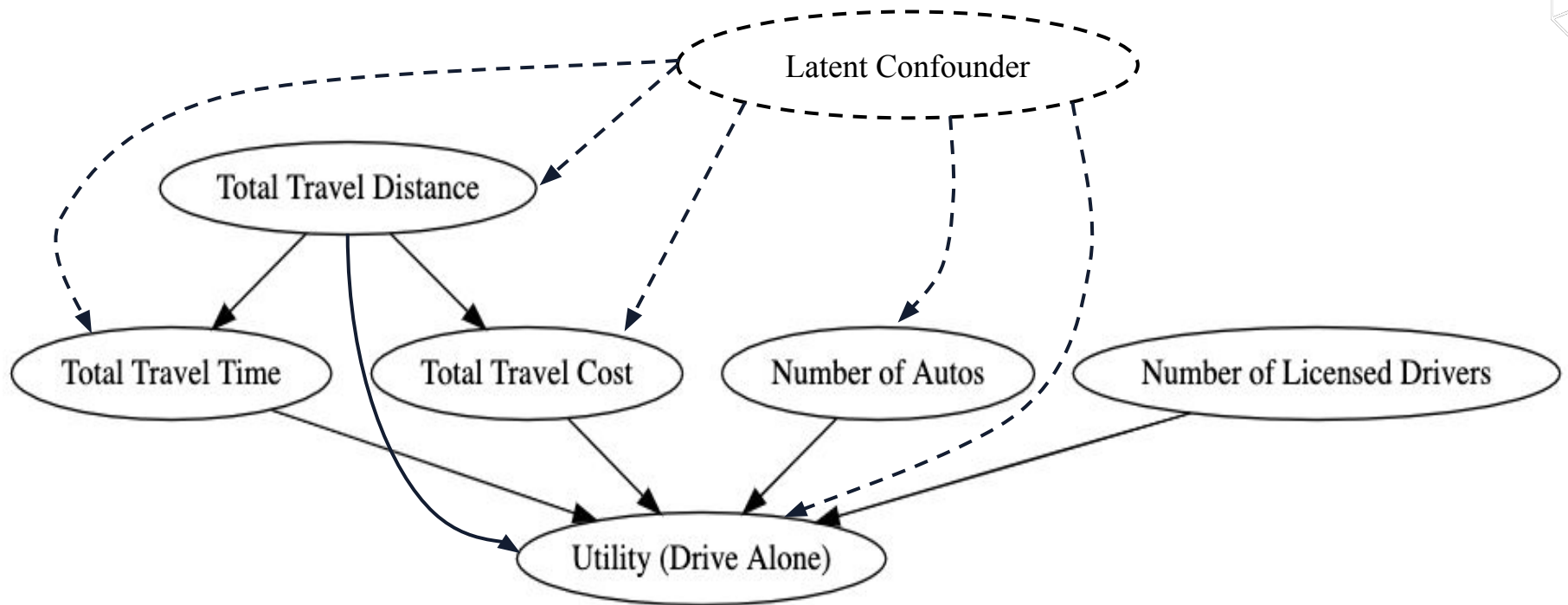
Example: Permutation tests





Latent Confounding

What if you have latent confounders?



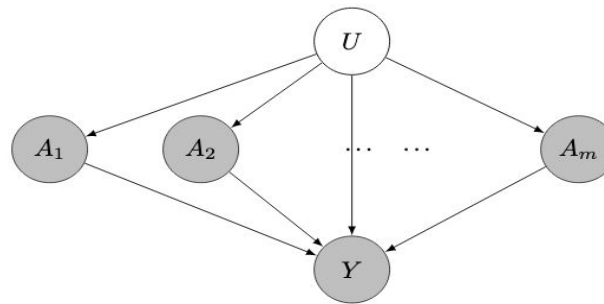
What can we do?

- Need to model the latent confounder
- Existing methods typically require more data
 - What if we cannot collect more data?



Deconfounder Algorithm

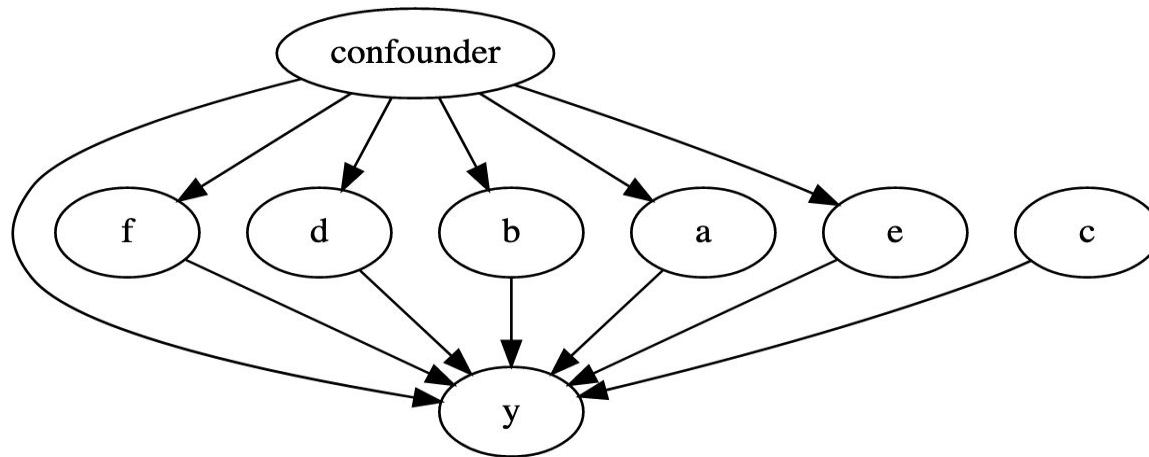
- Construct a substitute confounder
 - Factor models, mixture models, topic models, etc...
- Test your factor model (important)
- Fit an outcome model controlling for the substitute confounder
- Estimate distribution of causal effect of interest, $P(y|\text{do}(.))$



Does it work?

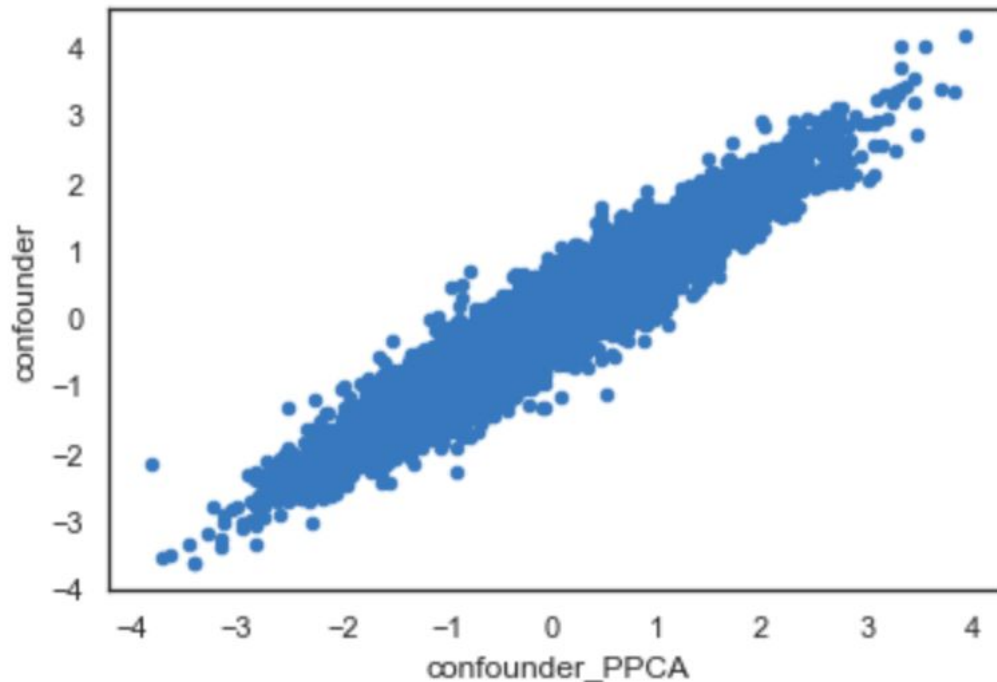
- Tested the deconfounder on a simple linear regression simulation exercise:

$$y = 5 + 2a + 2b - 5c + 3d + 2e - 5f + 7\text{confounder}$$



Estimating the substitute confounder

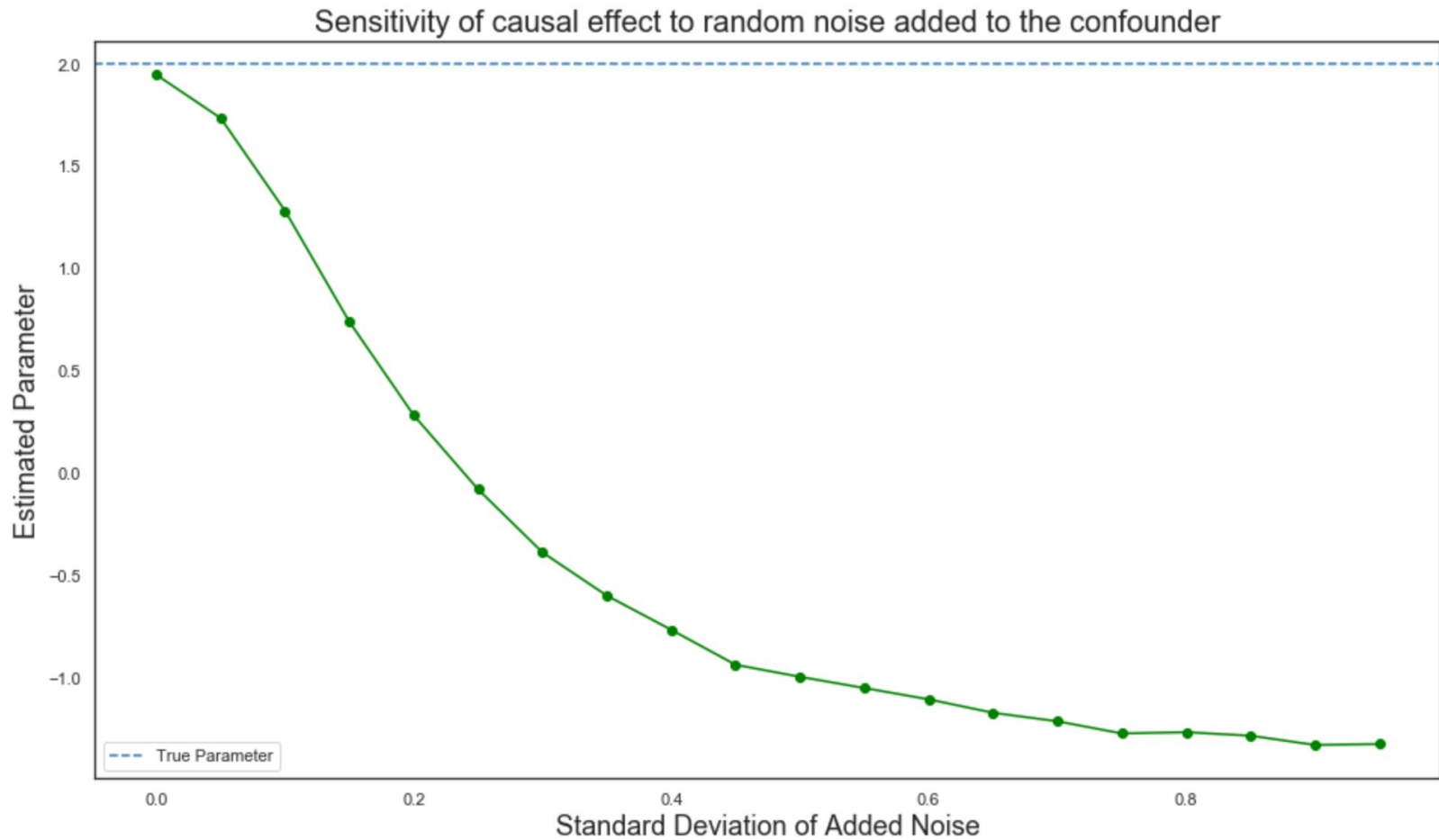
- Ran Probabilistic Principal Components Analysis (PPCA) on explanatory variables



So, does it work?

	estimated_params	std_errors	true_params
Intercept	5.004327	0.017393	5
a	2.420599	0.035443	2
b	-1.504753	0.052781	2
c	-4.982046	0.017408	-5
d	5.160087	0.051200	3
e	2.975473	0.042617	2
f	-5.471528	0.034864	-5
confounder_PPCA	-0.130928	0.121611	7

What's going on?



What does this mean?

- Eliminating bias in the presence of latent confounding is hard
- It is crucial to recover an accurate substitute confounder
- Even then, best case scenario is we're reducing the bias, but not eliminating it



Thanks!

**Questions? Please get in touch
directly by submitting a github
issue**