CAUSAL INFERENCE WITH ORDINAL OUTCOMES

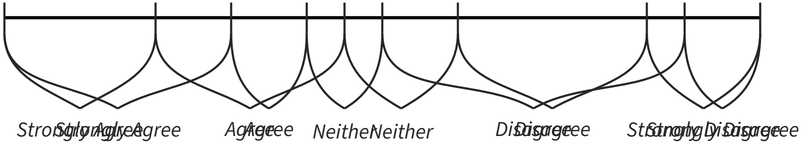
Density Estimation Based Approach

Chanhyuk Park
Washington University in St. Louis

MOTIVATION: THE DISCRETIZATION PROBLEM

- Experiments to identify causal effects of treatments
- Outcomes are usually believed to be continuous in unidimensional space
  - Approval ratings (Canes-Wrone and De Marchi, 2002; Kriner and Schwartz, 2009)
  - Policy preferences (Scheve and Slaughter, 2001; Mayda and Rodrik, 2005; Wu, 2022)
- The Problem - How we measure it

The True Outcome



THE GOAL

- *Identification*

  Naive causal identification with these "index" may fail

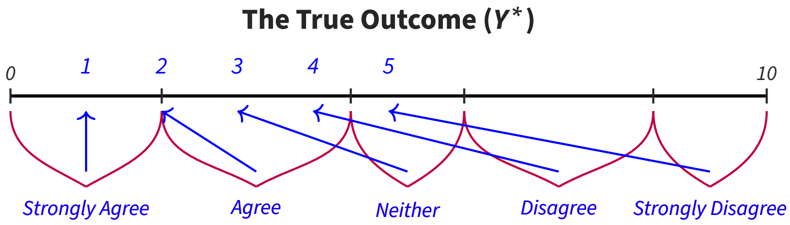
  $\Longrightarrow$ Normalized Latent Treatment Effect

- *Estimation*

  Standard parametric regression is rigid

  $\Longrightarrow$ Flexible, density estimation based estimators

WHY NAIVE IDENTIFICATION FAILS



- The treatment effect we want: $\mathbb{E}[Y^*(1)] - \mathbb{E}[Y^*(0)]$
- *Unknown Reporting Function g*
- *Arbitrary index f*
- The treatment effect we get: $\mathbb{E}[f(g(Y^*(1)))] - \mathbb{E}[f(g(Y^*(0)))]$

NORMALIZATION

- Unobserved Y^* and Unknown g

  $\Longrightarrow$ identify LTE up to scale

  $\Longrightarrow$ treatment effect τ^* cannot be distinguished form $C\tau^*$

- Normalization to cancel out C
  - Scale to fix $\sigma_{\varepsilon_i} = 1$

    $\Longrightarrow$ $\frac{\tau^*}{\sigma_{\varepsilon_i}}$ is identified
  - If there are multiple treatments, fixing $\tau_a = 1$
  - Pure probability scale

HOW CAN WE ESTIMATE?

- Ordered Probit, Ordered Logit, and most IRT models

  $\Longrightarrow$ Assume that ε_i follows specific distributions

- Distributional Assumptions can be violated
  - Pure misspecification
  - Unobserved confounder

- Become *inconsistent*

- Two Flexible Estimators without rigid Distributional Assumptions

ALTERNATIVE: ESTIMATE THE DISTRIBUTION

- Nonparametric method: Kernel Density Estimation
  - Smooth each observation using a kernel (usually Gaussian)

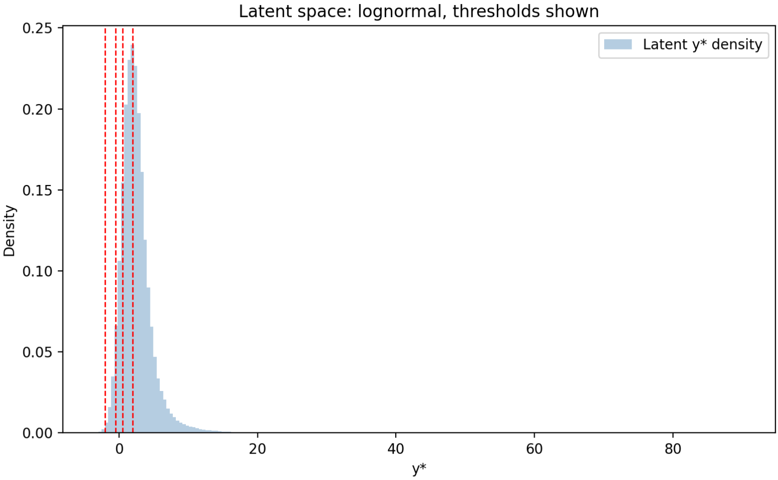
$$\hat{f}(x) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right)$$

- Parametric Generative Model: Normalizing Flows
  - Use the change-of-variable formula

$$f_\theta(h) = f_Z\left(T_\theta^{-1}(h)\right) \left|\det \mathrm{d}T_\theta^{-1}(h)\right|,$$

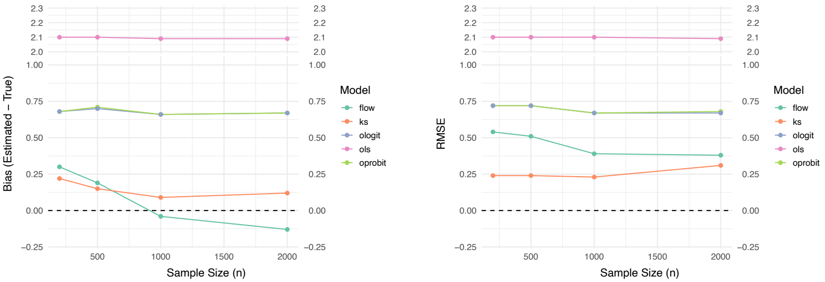
  - T_θ is a set of *invertible* transformation

SIMULATION



Latent space: lognormal, thresholds shown
Latent y* density
Density
y*
0.25
0.20
0.15
0.10
0.05
0.00
0
20
40
60
80

SIMULATION – LOGNORMAL CASE



True treatment size: 0.46

replication – Mattingly et al. (2025)

Outcome: Preference over Political System					
	ATE	NLTE			
	Original – OLS	Ordered Logit	Ordered Probit	KDE-based	NF-based
China	1.04***	0.73***	0.76***	0.36***	0.37***
	(0.05)	(0.04)	(0.04)	(0.07)	(0.04)
USA	−0.43***	−0.35***	−0.38***	−0.18***	−0.17***
	(0.04)	(0.04)	(0.04)	(0.05)	(0.03)
Competition	0.36***	0.21***	0.25***	0.13*	0.11**
	(0.05)	(0.04)	(0.04)	(0.06)	(0.04)

CONCLUSION

- *Identification*

  Naive use of ordinal "index" may mislead

  $\Longrightarrow$ Normalized Latent Treatment Effect

- *Estimation*

  Standard Ordinal regression approaches risk inconsistent

  $\Longrightarrow$ KDE or NF based estimators