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

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

MOTIVATION: THE DISCRETIZATION PROBLEM

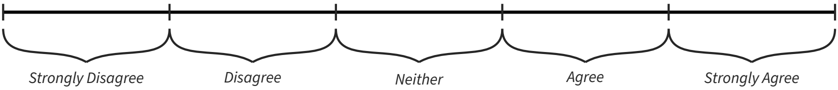
- Experiments to identify causal effects of treatments
- Outcomes are usually believed to be continuous in unidimensional space
  - Approval ratings (Canes-Wrone and De Marchi, 2002; Kriner and Schwartz, 2009)
  - Policy preferences (Scheve and Slaughter, 2001; Mayda and Rodrik, 2005; Wu, 2022)
- The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

The True Outcome (Y^*)

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

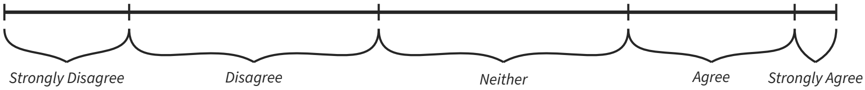
The True Outcome (Y^*)



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 (Y^*)



THE GOAL

- *Identification*
- *Estimation*

THE GOAL

- *Identification*

  Naive causal identification with these "index" may fail

  $\Longrightarrow$ Alternative estimand: Normalized Latent Treatment Effect

- *Estimation*

THE GOAL

- *Identification*

  Naive causal identification with these "index" may fail

  $\Longrightarrow$ Alternative estimand: Normalized Latent Treatment Effect

- *Estimation*

  Parametric ordered probit and logit require strong distributional assumptions

  $\Longrightarrow$ Flexible, density estimation based estimators

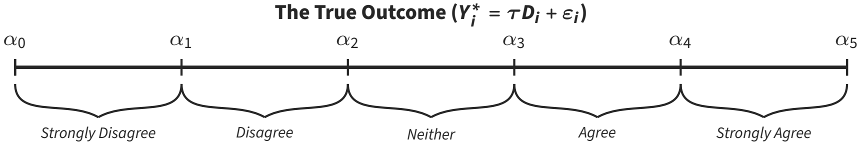
WHY NAIVE IDENTIFICATION FAILS

The True Outcome ($Y_i^* = \tau D_i + \varepsilon_i$)

	$Y^*(0)$	$Y^*(1)$
A	0.8	1.2
B	0.4	0.6
C	1.9	2.2
D	1.5	1.7
E	2.5	2.4

- $\tau = \mathbb{E}[Y^*(1)] - \mathbb{E}[Y^*(0)] \approx$ **+0.2**

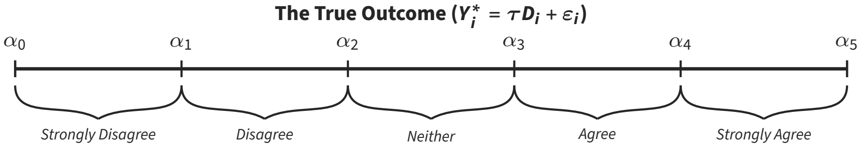
WHY NAIVE IDENTIFICATION FAILS



	$Y^*(0)$	$Y^*(1)$	$Y_a(0)$	$Y_a(1)$
A	0.8	1.2	Disagree	Neither
B	0.4	0.6	Disagree	Disagree
C	1.9	2.2	Neither	Agree
D	1.5	1.7	Neither	Neither
E	2.5	2.4	Agree	Agree

- $\tau = \mathbb{E}[Y^*(1)] - \mathbb{E}[Y^*(0)] \approx$ **+0.2**

WHY NAIVE IDENTIFICATION FAILS

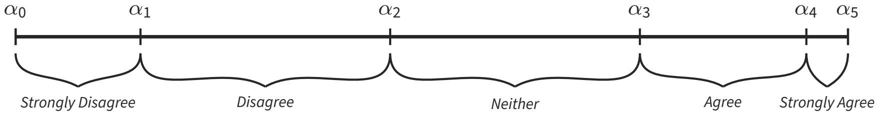


	$Y^*(0)$	$Y^*(1)$	$Y_a(0)$	$Y_a(1)$
A	0.8	1.2	2	3
B	0.4	0.6	2	2
C	1.9	2.2	3	4
D	1.5	1.7	3	3
E	2.5	2.4	4	4

- $\tau = \mathbb{E}[Y^*(1)] - \mathbb{E}[Y^*(0)] \approx$ **+0.2**
- $\tau_a = \mathbb{E}[Y_a(1)] - \mathbb{E}[Y_a(0)] \approx$ **+0.4**

WHY NAIVE IDENTIFICATION FAILS

The True Outcome ($Y_i^* = \tau D_i + \varepsilon_i$)

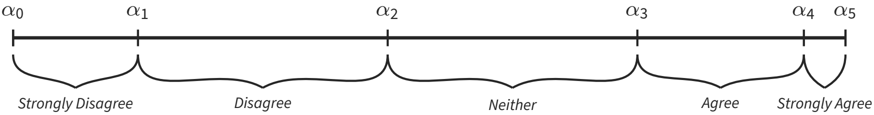


	$Y^*(0)$	$Y^*(1)$	$Y_b(0)$	$Y_b(1)$
A	0.8	1.2	Disagree	Disagree
B	0.4	0.6	Disagree	Disagree
C	1.9	2.2	Neither	Neither
D	1.5	1.7	Neither	Neither
E	2.5	2.4	Agree	Neither

- $\tau = \mathbb{E}[Y^*(1)] - \mathbb{E}[Y^*(0)] \approx$ **+0.2**

WHY NAIVE IDENTIFICATION FAILS

The True Outcome ($Y_i^* = \tau D_i + \varepsilon_i$)



	$Y^*(0)$	$Y^*(1)$	$Y_b(0)$	$Y_b(1)$
A	0.8	1.2	2	2
B	0.4	0.6	2	2
C	1.9	2.2	3	3
D	1.5	1.7	3	3
E	2.5	2.4	4	3

- $\tau = \mathbb{E}[Y^*(1)] - \mathbb{E}[Y^*(0)] \approx$ **+0.2**
- $\tau_b = \mathbb{E}[Y_b(1)] - \mathbb{E}[Y_b(0)] \approx$ **-0.2**

▸ Binarization

IDENTIFICATION THROUGH LINK FUNCTION

- $Y_i^* = \tau D_i + \varepsilon_i$
- Cumulative probabilities such as $\mathbb{P}(Y \leq \text{Agree})$
- If we can *map* this probability to the Y^* space → Identify τ!
- Link Function = CDF of ε_i (F_{ε_i}) does this:

$$F_{\varepsilon_i}^{-1}\big(\mathbb{P}(Y_i \leq \text{Agree})\big) = \underbrace{\alpha_4 - \tau D_i}_{\text{Defined in } Y^* \text{ space}}$$

- Then we can identify

$$\underbrace{F_{\varepsilon_i}^{-1}(\mathbb{P}(Y_i(0) \leq \text{Agree}))}_{\text{Control}} - \underbrace{F_{\varepsilon_i}^{-1}(\mathbb{P}(Y_i(1) \leq \text{Agree}))}_{\text{Treated}} = \tau$$

IDENTIFICATION THROUGH LINK FUNCTION

- $Y_i^* = \tau D_i + \varepsilon_i$
- Cumulative probabilities such as $\mathbb{P}(Y \leq \text{Agree})$
- If we can *map* this probability to the Y^* space → Identify τ!
- Link Function = CDF of ε_i (F_{ε_i}) does this:

$$F_{\varepsilon_i}^{-1}\big(\mathbb{P}(Y_i \leq \text{Agree})\big) = \underbrace{\alpha_4 - \tau D_i}_{\text{Defined in } Y^* \text{ space}}$$

- Then we can identify

$$\underbrace{F_{\varepsilon_i}^{-1}(\mathbb{P}(Y_i(0) \leq \text{Agree}))}_{\text{Control}} - \underbrace{F_{\varepsilon_i}^{-1}(\mathbb{P}(Y_i(1) \leq \text{Agree}))}_{\text{Treated}} = \tau$$

IDENTIFICATION THROUGH LINK FUNCTION

- $Y_i^* = \tau D_i + \varepsilon_i$
- Cumulative probabilities such as $\mathbb{P}(Y \leq \text{Agree})$
- If we can *map* this probability to the Y^* space → Identify τ!
- Link Function = CDF of ε_i (F_{ε_i}) does this:

$$F_{\varepsilon_i}^{-1}\Big(\mathbb{P}(Y_i \leq \text{Agree})\Big) = \underbrace{\alpha_4 - \tau D_i}_{\text{Defined in } Y^* \text{ space}}$$

- Then we can identify

$$\underbrace{F_{\varepsilon_i}^{-1}(\mathbb{P}(Y_i(0) \leq \text{Agree}))}_{\text{Control}} - \underbrace{F_{\varepsilon_i}^{-1}(\mathbb{P}(Y_i(1) \leq \text{Agree}))}_{\text{Treated}} = \tau$$

IDENTIFICATION THROUGH LINK FUNCTION

- $Y_i^* = \tau D_i + \varepsilon_i$
- Cumulative probabilities such as $\mathbb{P}(Y \leq \text{Agree})$
- If we can *map* this probability to the Y^* space → Identify τ!
- Link Function = CDF of ε_i (F_{ε_i}) does this:

$$F_{\varepsilon_i}^{-1}\left(\mathbb{P}(Y_i \leq \text{Agree})\right) = \underbrace{\alpha_4 - \tau D_i}_{\text{Defined in } Y^* \text{ space}}$$

- Then we can identify

$$\underbrace{F_{\varepsilon_i}^{-1}(\mathbb{P}(Y_i(0) \leq \text{Agree}))}_{\text{Control}} - \underbrace{F_{\varepsilon_i}^{-1}(\mathbb{P}(Y_i(1) \leq \text{Agree}))}_{\text{Treated}} = \tau$$

IDENTIFICATION THROUGH LINK FUNCTION

- $Y_i^* = \tau D_i + \varepsilon_i$
- Cumulative probabilities such as $\mathbb{P}(Y \leq \text{Agree})$
- If we can *map* this probability to the Y^* space → Identify τ!
- Link Function = CDF of ε_i (F_{ε_i}) does this:

$$F_{\varepsilon_i}^{-1}\left(\mathbb{P}(Y_i \leq \text{Agree})\right) = \underbrace{\alpha_4 - \tau D_i}_{\text{Defined in } Y^* \text{ space}}$$

- Then we can identify

$$\underbrace{F_{\varepsilon_i}^{-1}(\mathbb{P}(Y_i(0) \leq \text{Agree}))}_{\text{Control}} - \underbrace{F_{\varepsilon_i}^{-1}(\mathbb{P}(Y_i(1) \leq \text{Agree}))}_{\text{Treated}} = \tau$$

More on Identification

NEW ESTIMANDS: NORMALIZED TREATMENT EFFECT

- Problem is, for a positive constant C,

$$\mathbb{P}\left(Y_i^* \leq \alpha_4\right) = \mathbb{P}\left(C \cdot Y_i^* \leq C \cdot \alpha_4\right)$$

- Thus τ becomes indistinguishable from $C \cdot \tau$
- Normalization to cancel out C
  1. Scale to fix $\sigma_{\varepsilon_i} = 1$
  2. If there are multiple treatments, fixing $\tau_a = 1$

NEW ESTIMANDS: NORMALIZED TREATMENT EFFECT

- Problem is, for a positive constant C,

$$\mathbb{P}\left(Y_i^* \leq \alpha_4\right) = \mathbb{P}\left(C \cdot Y_i^* \leq C \cdot \alpha_4\right)$$

- Thus τ becomes indistinguishable from $C \cdot \tau$
- Normalization to cancel out C
  1. Scale to fix $\sigma_{\varepsilon_i} = 1$
  2. If there are multiple treatments, fixing $\tau_d = 1$

NEW ESTIMANDS: NORMALIZED TREATMENT EFFECT

- Problem is, for a positive constant C,

$$\mathbb{P}\left(Y_i^* \leq \alpha_4\right) = \mathbb{P}\left(C \cdot Y_i^* \leq C \cdot \alpha_4\right)$$

- Thus τ becomes indistinguishable from $C \cdot \tau$
- Normalization to cancel out C
  1. Scale to fix $\sigma_{\varepsilon_i} = 1$
  2. If there are multiple treatments, fixing $\tau_a = 1$

NEW ESTIMANDS: NORMALIZED TREATMENT EFFECT

- Problem is, for a positive constant C,

$$\mathbb{P}\left(Y_i^* \leq \alpha_4\right) = \mathbb{P}\left(C \cdot Y_i^* \leq C \cdot \alpha_4\right)$$

- Thus τ becomes indistinguishable from $C \cdot \tau$
- Normalization to cancel out C
  1. Scale to fix $\sigma_{\varepsilon_i} = 1$
  2. If there are multiple treatments, fixing $\tau_a = 1$

NEW ESTIMANDS: NORMALIZED TREATMENT EFFECT

- Problem is, for a positive constant C,

$$\mathbb{P}\left(Y_i^* \leq \alpha_4\right) = \mathbb{P}\left(C \cdot Y_i^* \leq C \cdot \alpha_4\right)$$

- Thus τ becomes indistinguishable from $C \cdot \tau$
- Normalization to cancel out C
  1. Scale to fix $\sigma_{\varepsilon_i} = 1$
  2. If there are multiple treatments, fixing $\tau_a = 1$

NEW ESTIMANDS: NORMALIZED TREATMENT EFFECT

- Problem is, for a positive constant C,

$$\mathbb{P}\left(Y_i^* \leq \alpha_4\right) = \mathbb{P}\left(C \cdot Y_i^* \leq C \cdot \alpha_4\right)$$

- Thus τ becomes indistinguishable from $C \cdot \tau$
- Normalization to cancel out C
  1. Scale to fix $\sigma_{\varepsilon_i} = 1$

     $\implies \frac{C \cdot \tau}{C \cdot \sigma_{\varepsilon_i}} = \frac{\tau}{\sigma_{\varepsilon_i}}$ is identified
  2. If there are multiple treatments, fixing $\tau_a = 1$

HOW CAN WE ESTIMATE?

- The key is Link Function (F_{ϵ_i})
- Ordered Probit, Ordered Logit, and most IRT models
- Assume F_{ε_i} is a specific distribution → MLE
- Distributional Assumptions can be violated
  - Pure misspecification
  - Unaccounted covariate

$\Longrightarrow$ Become *inconsistent*

- Two Flexible Estimators without rigid Distributional Assumptions

HOW CAN WE ESTIMATE?

- The key is Link Function (F_{ϵ_i})
- Ordered Probit, Ordered Logit, and most IRT models
- Assume F_{ε_i} is a specific distribution $\rightarrow$ MLE
- Distributional Assumptions can be violated
  - Pure misspecification
  - Unaccounted covariate

$\Longrightarrow$ Become *inconsistent*

- Two Flexible Estimators without rigid Distributional Assumptions

HOW CAN WE ESTIMATE?

- The key is Link Function (F_{ϵ_i})
- Ordered Probit, Ordered Logit, and most IRT models
- Assume F_{ε_i} is a specific distribution → MLE
- Distributional Assumptions can be violated
  - Pure misspecification
  - Unaccounted covariate

$\Longrightarrow$ Become *inconsistent*

- Two Flexible Estimators without rigid Distributional Assumptions

HOW CAN WE ESTIMATE?

- The key is Link Function (F_{ϵ_i})
- Ordered Probit, Ordered Logit, and most IRT models
- Assume F_{ε_i} is a specific distribution $\rightarrow$ MLE
- Distributional Assumptions can be violated
  - Pure misspecification
  - Unaccounted covariate

$\Longrightarrow$ Become *inconsistent*

- Two Flexible Estimators without rigid Distributional Assumptions

ALTERNATIVE: ESTIMATE THE DISTRIBUTION

- Estimated $\hat{F}_{\varepsilon_i}$ from the Data
- Nonparametric method: Kernel Density Estimation
  - Use Kernels to smooth each observations
- Parametric Generative Model: Normalizing Flows
  - Transform complex distributions to simple one
- Same MLE

ALTERNATIVE: ESTIMATE THE DISTRIBUTION

- Estimated $\hat{F}_{\varepsilon_i}$ from the Data
- Nonparametric method: Kernel Density Estimation
  - Use Kernels to smooth each observations
- Parametric Generative Model: Normalizing Flows
  - Transform complex distributions to simple one
- Same MLE

ALTERNATIVE: ESTIMATE THE DISTRIBUTION

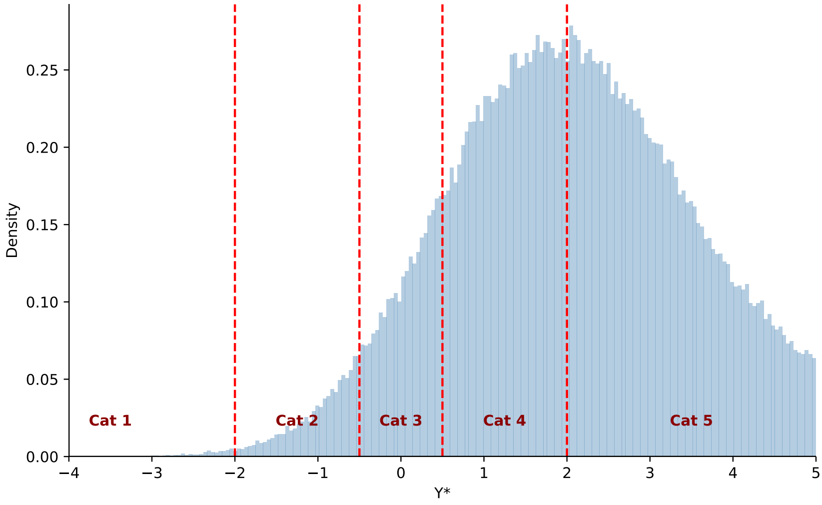
- Estimated $\hat{F}_{\varepsilon_i}$ from the Data
- Nonparametric method: Kernel Density Estimation
  - Use Kernels to smooth each observations
- Parametric Generative Model: Normalizing Flows
  - Transform complex distributions to simple one
- Same MLE

ALTERNATIVE: ESTIMATE THE DISTRIBUTION

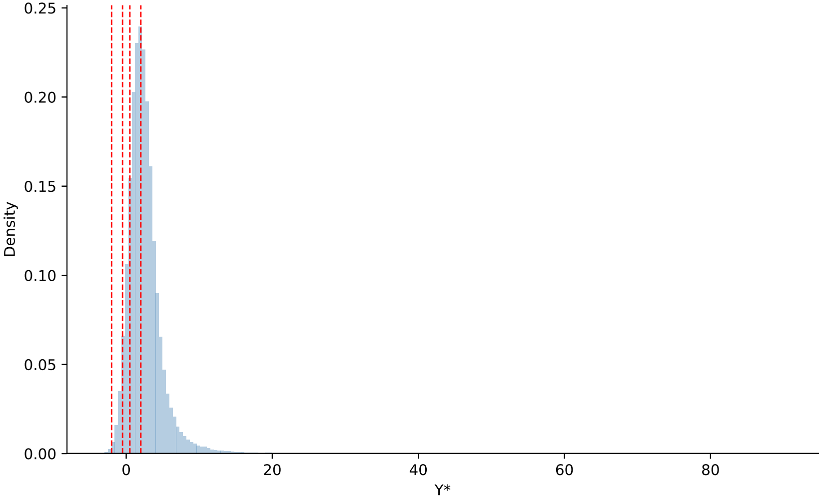
- Estimated $\hat{F}_{\varepsilon_i}$ from the Data
- Nonparametric method: Kernel Density Estimation
  - Use Kernels to smooth each observations
- Parametric Generative Model: Normalizing Flows
  - Transform complex distributions to simple one
- Same MLE

▸ Little More on Density Estimation

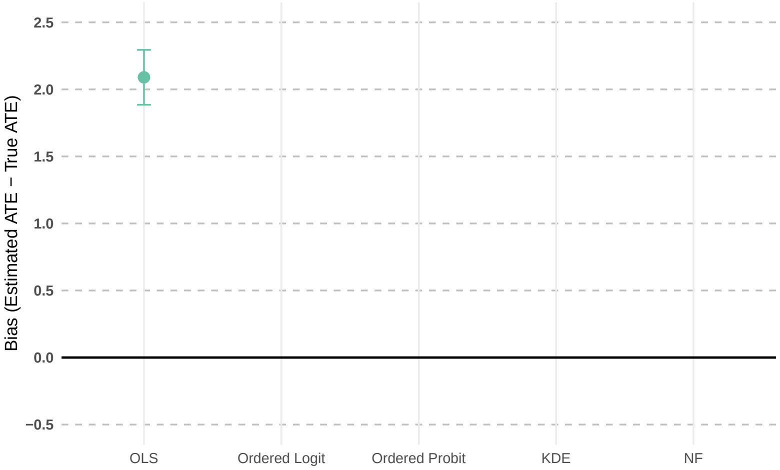
SIMULATION SETTING



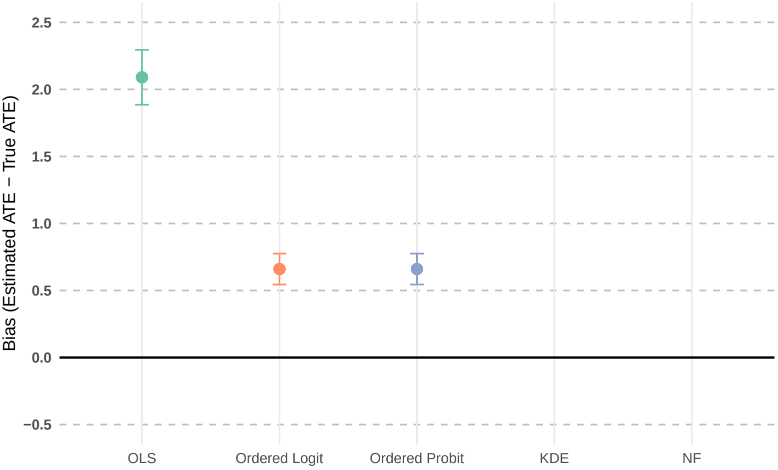
SIMULATION SETTING



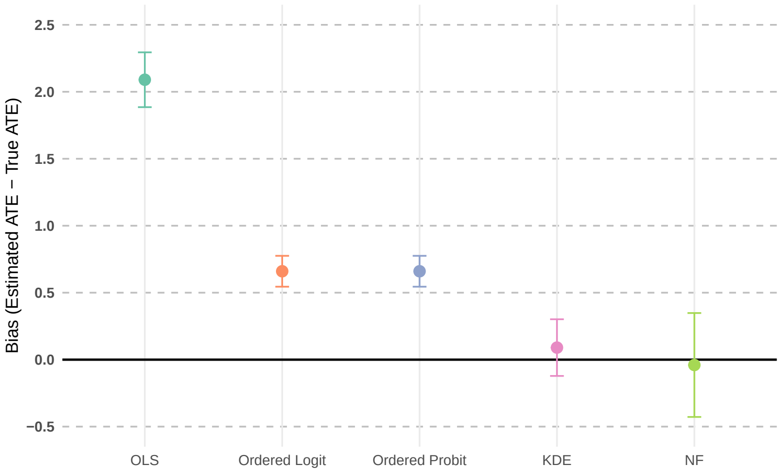
SIMULATION RESULTS: BIAS AND MC SD



SIMULATION RESULTS: BIAS AND MC SD



SIMULATION RESULTS: BIAS AND MC SD



replication – Mattingly et al. (2025)

- State-produced promotional media → preference on political system
- 19 different countries with $n = 6,000$
- Three treatments:
  1. China: Two CCP produced videos
  2. USA: Two US produced videos
  3. Competition: One from CCP and the other from US
- Outcome: Preference on Political System

  Strongly Prefer the US to *Strongly Prefer China*

replication – Mattingly et al. (2025)

- State-produced promotional media → preference on political system
- 19 different countries with $n = 6,000$
- Three treatments:
  1. China: Two CCP produced videos
  2. USA: Two US produced videos
  3. Competition: One from CCP and the other from US
- Outcome: Preference on Political System

Strongly Prefer the US to *Strongly Prefer China*

REPLICATION – MATTINGLY ET AL. (2025)

- State-produced promotional media → preference on political system
- 19 different countries with $n = 6,000$
- Three treatments:
  1. China: Two CCP produced videos
  2. USA: Two US produced videos
  3. Competition: One from CCP and the other from US
- Outcome: Preference on Political System

  Strongly Prefer the US to *Strongly Prefer China*

Replication – Mattingly et al. (2025)

- State-produced promotional media → preference on political system
- 19 different countries with $n = 6,000$
- Three treatments:
  1. China: Two CCP produced videos
  2. USA: Two US produced videos
  3. Competition: One from CCP and the other from US
- Outcome: Preference on Political System

  Strongly Prefer the US to *Strongly Prefer China*

replication – Mattingly et al. (2025)

- State-produced promotional media → preference on political system
- 19 different countries with $n = 6,000$
- Three treatments:
  1. China: Two CCP produced videos
  2. USA: Two US produced videos
  3. Competition: One from CCP and the other from US
- Outcome: Preference on Political System

  Strongly Prefer the US to *Strongly Prefer China*

replication – Mattingly et al. (2025)

- State-produced promotional media → preference on political system
- 19 different countries with $n = 6,000$
- Three treatments:
  1. China: Two CCP produced videos
  2. USA: Two US produced videos
  3. Competition: One from CCP and the other from US
- Outcome: Preference on Political System

  Strongly Prefer the US to *Strongly Prefer China*

replication – Mattingly et al. (2025)

- State-produced promotional media → preference on political system
- 19 different countries with $n = 6,000$
- Three treatments:
  1. China: Two CCP produced videos
  2. USA: Two US produced videos
  3. Competition: One from CCP and the other from US
- Outcome: Preference on Political System

  Strongly Prefer the US to *Strongly Prefer China*

REPLICATION RESULTS

	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)

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

REPLICATION RESULTS

	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)

***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

REPLICATION RESULTS

	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)

$^{***}p < 0.001$; $^{**}p < 0.01$; $^{*}p < 0.05$

CONCLUSION

- *Identification*
- *Estimation*

CONCLUSION

- *Identification*

  Naive use of ordinal "index" is not what we want

  $\Longrightarrow$ Alternative estimand: Normalized Latent Treatment Effect

- *Estimation*

CONCLUSION

- *Identification*

  Naive use of ordinal "index" is not what we want

  $\Longrightarrow$ Alternative estimand: Normalized Latent Treatment Effect

- *Estimation*

  Parametric ordinal regressions require strong assumptions

  $\Longrightarrow$ KDE or Normalizing Flows based estimators

MORE ON IDENTIFICATION

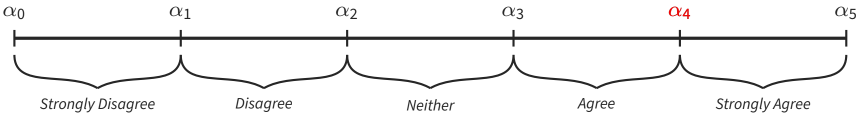
- Cumulative probabilities such as $\mathbb{P}(Y \leq \text{Agree})$
- Link function = CDF of the error term(F_{ε_i})
- Then we can identify

$$\underbrace{F_{\varepsilon_i}^{-1}(\mathbb{P}(Y_i(0) \leq \text{Agree}))}_{\text{Control}} - \underbrace{F_{\varepsilon_i}^{-1}(\mathbb{P}(Y_i(1) \leq \text{Agree}))}_{\text{Treated}} = \tau$$

Back

MORE ON IDENTIFICATION

The True Outcome ($Y_i^* = \tau D_i + \varepsilon_i$)



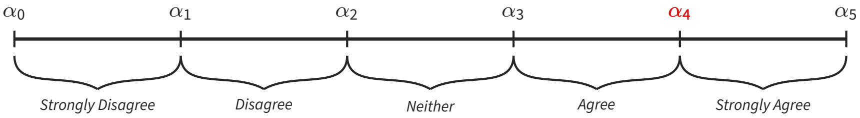
- Link function = CDF of the error term(F_{ε_i})

$$F_{\varepsilon_i}^{-1}\left(\mathbb{P}(Y_i \leq \text{Agree})\right) = F_{\varepsilon_i}^{-1}\left(\mathbb{P}\left(Y_i^* \leq \alpha_4\right)\right)$$

Back

MORE ON IDENTIFICATION

The True Outcome ($Y_i^* = \tau D_i + \varepsilon_i$)



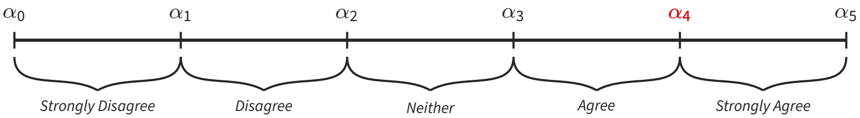
- Link function = CDF of the error term(F_{ε_i})

$$
\begin{aligned}
F_{\varepsilon_i}^{-1}\left(\mathbb{P}(Y_i \leq \text{Agree})\right) &= F_{\varepsilon_i}^{-1}\left(\mathbb{P}\left(Y_i^* \leq \alpha_4\right)\right) \\
&= F_{\varepsilon_i}^{-1}\left(\mathbb{P}\left(\tau D_i + \varepsilon_i \leq \alpha_4\right)\right)
\end{aligned}
$$

Back

MORE ON IDENTIFICATION

The True Outcome ($Y_i^* = \tau D_i + \varepsilon_i$)



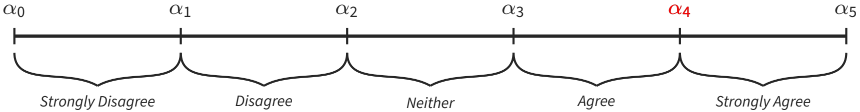
- Link function = CDF of the error term(F_{ε_i})

$$
\begin{aligned}
F_{\varepsilon_i}^{-1}\left(\mathbb{P}(Y_i \leq \text{Agree})\right) &= F_{\varepsilon_i}^{-1}\left(\mathbb{P}\left(Y_i^* \leq \alpha_4\right)\right) \\
&= F_{\varepsilon_i}^{-1}\left(\mathbb{P}\left(\tau D_i + \varepsilon_i \leq \alpha_4\right)\right) \\
&= F_{\varepsilon_i}^{-1}\left(\mathbb{P}\left(\varepsilon_i \leq \alpha_4 - \tau D_i\right)\right)
\end{aligned}
$$

Back

MORE ON IDENTIFICATION

The True Outcome ($Y_i^* = \tau D_i + \varepsilon_i$)



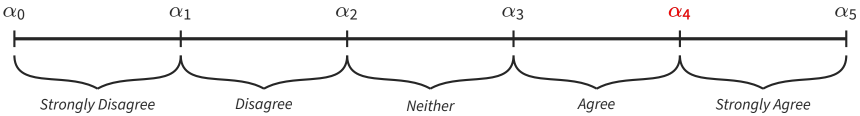
- Link function = CDF of the error term(F_{ε_i})

$$
\begin{aligned}
F_{\varepsilon_i}^{-1}\left(\mathbb{P}(Y_i \leq \text{Agree})\right) &= F_{\varepsilon_i}^{-1}\left(\mathbb{P}\left(Y_i^* \leq \alpha_4\right)\right) \\
&= F_{\varepsilon_i}^{-1}\left(\mathbb{P}\left(\tau D_i + \varepsilon_i \leq \alpha_4\right)\right) \\
&= F_{\varepsilon_i}^{-1}\left(\mathbb{P}\left(\varepsilon_i \leq \alpha_4 - \tau D_i\right)\right) \\
&= F_{\varepsilon_i}^{-1}\left(F_{\varepsilon_i}\left(\alpha_4 - \tau D_i\right)\right)
\end{aligned}
$$

Back

MORE ON IDENTIFICATION

The True Outcome ($Y_i^* = \tau D_i + \varepsilon_i$)



- Link function = CDF of the error term(F_{ε_i})

$$
\begin{aligned}
F_{\varepsilon_i}^{-1}\left(\mathbb{P}(Y_i \leq \text{Agree})\right) &= F_{\varepsilon_i}^{-1}\left(\mathbb{P}\left(Y_i^* \leq \alpha_4\right)\right) \\
&= F_{\varepsilon_i}^{-1}\left(\mathbb{P}(\tau D_i + \varepsilon_i \leq \alpha_4)\right) \\
&= F_{\varepsilon_i}^{-1}\left(\mathbb{P}(\varepsilon_i \leq \alpha_4 - \tau D_i)\right) \\
&= F_{\varepsilon_i}^{-1}\left(F_{\varepsilon_i}(\alpha_4 - \tau D_i)\right) \\
&= \alpha_4 - \tau D_i
\end{aligned}
$$

Back

MORE ON IDENTIFICATION

- Cumulative probabilities such as $\mathbb{P}(Y \leq \text{Agree})$
- Link function = CDF of the error term(F_{ε_i})

$$F_{\varepsilon_i}^{-1}\left(\mathbb{P}(Y_i \leq \text{Agree})\right) = \alpha_4 - \tau D_i$$

- Then we can identify

$$\underbrace{F_{\varepsilon_i}^{-1}(\mathbb{P}(Y_i(0) \leq \text{Agree}))}_{\text{Control}} - \underbrace{F_{\varepsilon_i}^{-1}(\mathbb{P}(Y_i(1) \leq \text{Agree}))}_{\text{Treated}} = \tau$$

Back

LITTLE MORE ON DENSITY ESTIMATION

- Kernel Density Estimation

  Let $V_i(\beta, \tau) = f(X_i, \beta) + D_i^\top \tau$

  By the Bayes' rule,

$$\mathbb{P}(Y_i \leq j \mid V_i = v) = \frac{\mathbb{P}(Y_i \leq j) g_1(v \mid Y \leq j)}{\mathbb{P}(Y_i \leq j) g_1(v \mid Y \leq j) + \mathbb{P}(Y_i > j) g_0(v \mid Y > j)}$$

  Estimate g_i using KDE

$$\hat{g}_1(v \mid Y \leq j) = \frac{1}{n_1(j) h_{1j}} \sum_{i:Y_i \leq j} K\left(\frac{v - V_i}{h_{1j}}\right)$$

$$\hat{g}_0(v \mid Y > j) = \frac{1}{n_0(j) h_{0j}} \sum_{i:Y_i > j} K\left(\frac{v - V_i}{h_{0j}}\right)$$

Back

LITTLE MORE ON DENSITY ESTIMATION

- Normalizing Flows
  - Use the change-of-variable formula

$$f_\theta(\varepsilon) = f_Z\left(T_\theta^{-1}(\varepsilon)\right)\left|\det\left(\frac{\partial T_\theta^{-1}(\varepsilon)}{\partial \varepsilon}\right)\right|,$$

  - $\varepsilon_i = T_\theta(Z_i)$
  - T_θ is a set of *invertible* transformation
  - Maps ε_i to simple Z (e.g. Standard Normal)
  - Estimate based on Z and then translate it back to ε_i

Back

WHY NOT JUST USE BINARY OUTCOMES?

- Loss of Power
  - Collapsing categories wastes information
  - Require larger samples ($\approx 5\times$)
- Aggregation Bias
  - Arbitrary grouping may fail to identify effects or *flip the sign* of the effect
  - Depends on shifts in middle categories
- Example:
  - Treatment moves "Strongly Disagree" → "Disagree"
  - Binary ("Positive" vs "Negative") sees *zero* effect
  - Ordinal model captures the improvement

◂ Back