CAUSAL INFERENCE WITH ORDINAL OUTCOMES

A Semiparametric Approach

Chanhyuk Park
Washington University in St. Louis

MOTIVATION

- There has been growing emphasis on causal identification in political science
- In the mean time, many research measures their outcome with ordinal scales
  - Approval ratings (Canes-Wrone and De Marchi, 2002; Kriner and Schwartz, 2009)
  - Trade policy (Scheve and Slaughter, 2001; Mayda and Rodrik, 2005; Wu, 2022)
- The Problem:
  a. The usual causal inference tools are designed to serve cardinal or at least interval outcomes
  b. Standard ordinal regression based on strong distributional assumption
- This paper to suggest an alternative estimand and estimation process

PROBLEM SETTING

- Suppose a DGP

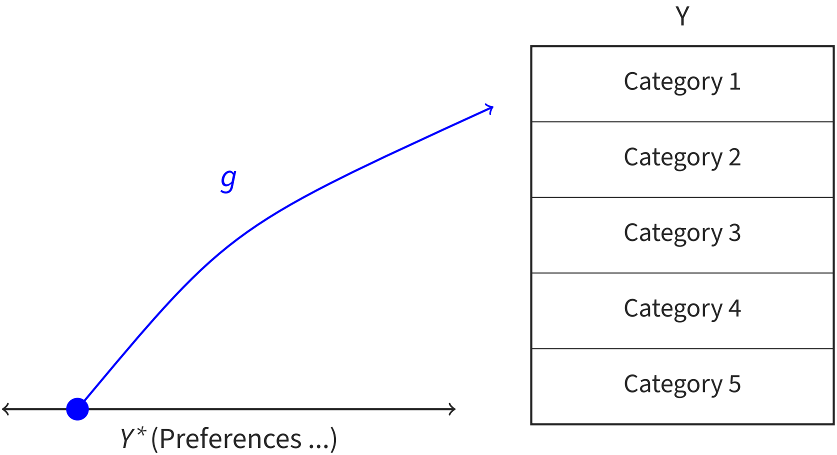
$$Y^* = \tau D + \beta X + \epsilon$$

- Y^* is outcome in unidimensional space that we cannot observe
- D denotes a binary treatment
- We only can observe the transformed version of it, $Y \in \{1, \ldots, j, \ldots, J\}$

$$Y = \begin{cases} 0 & \text{if } \alpha_{-1} < Y^* \leq \alpha_0 \\ \vdots & \vdots \\ J & \text{if } \alpha_{J-1} < Y^* \leq \alpha_J \end{cases}$$

, where α_k denotes the threshold points for each ordinal category

PROBLEM SETTING



POTENTIAL OUTCOME FRAMEWORK

- Denote the binary treatment status of individual $d_i \in \{0, 1\}$
- Denote the potential outcomes in latent preference scale as $Y_i^*(d_i)$
  - $Y_i^*(1) = \tau + \beta X_i + \epsilon_i$
  - $Y_i^*(0) = \beta X_i + \epsilon_i$
- Denote the potential outcomes in observed ordinal scale (POO) as $Y_i(d_i) = g(Y_i^*(d_i)) \in \{1, \ldots, j, \ldots J\}$
- We are interested in the average treatment effect in the latent space (LTE)

$$\begin{aligned} LTE &= \mathbb{E}\left[Y^*(1)\right] - \mathbb{E}\left[Y^*(0)\right] \\ &= \tau \end{aligned}$$

LTE WITH $Y(d_i)$

- If we knew Y^*, there is not problem
- If we knew g and it maps one value of Y^* to Y, then we may get LTE, but this not true in most settings

$$\mathbb{E}\left[g^{-1}(Y_i(1))\right] - \mathbb{E}\left[g^{-1}(Y_i(0))\right] \neq LTE$$

- Instead, we may identify it up to scale

ALTE

- We can identify the LTE up to scale of a constant if we know or can estimate the distribution of the ϵ_i
- Suppose the error term has a cdf of F
- From the model,

$$\mathbb{P}\left(Y(D) \le j\right) = \mathbb{P}\left(Y^* \le \alpha_j\right)$$

- However, if we scale Y^* side by a constant, c, we still get the same probability

$$\begin{aligned}\mathbb{P}\left(Y(D) \le j\right) &= \mathbb{P}\left(Y^* \le \alpha_j\right) \\ &= \mathbb{P}\left(cY^* \le c\alpha_j\right)\end{aligned}$$

ALTE

- Taking account the constant c, we can get

$$
\begin{aligned}
F^{-1}(\mathbb{P}(Y(0) &\leq j)) - F^{-1}(\mathbb{P}(Y(1) \leq j)) \\
&= F^{-1}\left(\mathbb{P}\left(cY^*(0) \leq c\alpha_j\right)\right) - F^{-1}(\mathbb{P}\left(cY^*(1) \leq c\alpha_j\right)) \\
&= F^{-1}\left(F(c\alpha_j - c\beta X)\right) - F_{c\epsilon}^{-1}(F(c\alpha_j - c\beta X - c\tau)) \\
&= c\tau
\end{aligned}
$$

- Good news is similarly we can identify β up to the same constant scale too
- This enable us to identify LTE anchored on β (ALTE)

$$
ALTE = \frac{\tau}{\beta}
$$

COMMON APPROACH

- Cardinalization of the ordinal outcome
- Ordered Logit and Probit

CARDINALIZATION

- Assume that each responses came from certain numeric value
- In other words, the transformation function g will look just like:

$$g = \begin{cases} \text{Strongly Diagree} & \text{if } Y^* = 1 \\ \text{Disagree Somewhat} & \text{if } Y^* = 2 \\ \text{Neither} & \text{if } Y^* = 3 \\ \text{Agree Somewhat} & \text{if } Y^* = 4 \\ \text{Strongly Agree} & \text{if } Y^* = 5 \end{cases}$$

- If this is true, we can calculate the average treatment effect by taking the inverse of g.

$$\begin{aligned} LTE &= \mathbb{E}\left[g^{-1}(Y(1))\right] - \mathbb{E}\left[g^{-1}(Y(0))\right] \\ &= \mathbb{E}\left[Y^*(1)\right] - \mathbb{E}\left[Y^*(0)\right] \end{aligned}$$

- The problem is, the assumption is unlikely and arbitrary

SMALL EXAMPLE

	$Y^*(0)$	$Y^*(1)$	$Y(0)$	$Y(1)$
A	1.37	0.09	Agree Somewhat	Disagree Somewhat
B	-0.56	1.71	Disagree Somewhat	Agree Somewhat
C	0.36	0.11	Neither	Disagree Somewhat
D	0.63	2.22	Neither	Agree Somewhat
E	0.4	0.14	Neither	Disagree Somewhat

- The LTE in the latent preference space is: 0.41

SMALL EXAMPLE

- Average Joe may assign 1 to 5 to each responses

	$Y^*(0)$	$Y^*(1)$	$Y_{Joe}(0)$	$Y_{Joe}(1)$
A	1.37	0.09	4	2
B	-0.56	1.71	2	4
C	0.36	0.11	3	2
D	0.63	2.22	3	4
E	0.4	0.14	3	2

- The LTE based on the Joe's cardinalization: $-0.2 \neq 0.41$

SMALL EXAMPLE

- Bold Soyeon comes in and argue that it should be -5, -2, 0, 8, 10 to each responses

	$Y^*(0)$	$Y^*(1)$	$Y_{Soyeon}(0)$	$Y_{Soyeon}(1)$
1	1.37	0.09	8	-2
2	-0.56	1.71	-2	8
3	0.36	0.11	0	-2
4	0.63	2.22	0	8
5	0.4	0.14	0	-2

- The LTE based on the Soyeon's cardinalization: $0.8 \neq 0.41$

SMALL EXAMPLE

- Smart Jacob finally argue that it should be -5, -2, 0, 8, 10 to each responses

	$Y^*(0)$	$Y^*(1)$	$Y_{Jacob}(0)$	$Y_{Jacob}(1)$
1	1.37	0.09	6	-2
2	-0.56	1.71	-2	6
3	0.36	0.11	0	-2
4	0.63	2.22	0	6
5	0.4	0.14	0	-2

- The LTE based on the Jacob's cardinalization: $0.4 \approx 0.41$

ORDERED LOGIT AND PROBIT

- Ordered Logit and Probit assume specific F (Logistic and Standard Normal)
- And use maximum likelihood to estimate the coefficients
- As we discussed, even we assume the F, we only can identify coefficients up to scale
- The problem is if F is different from Logistic or Standard Normal, the MLE becomes inconsistent

ALTERNATIVE: SEMIPARAMETRIC APPROACH

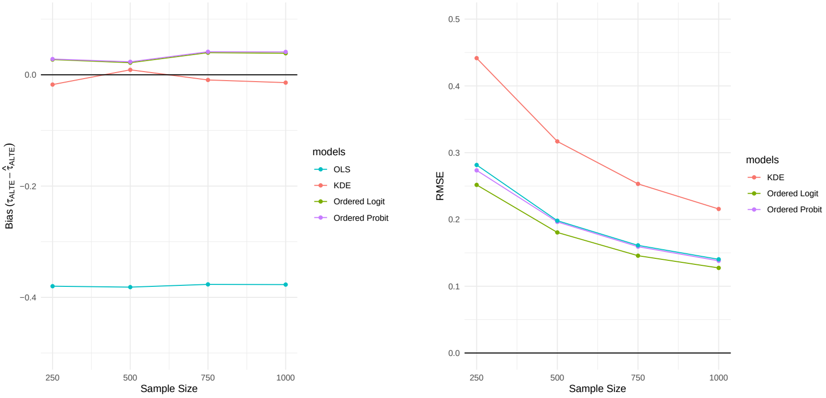
- Instead of assuming a specific F, we can estimate it from the data using kernel density estimation
- For notational ease, let's denote $f(X, \beta) + D^T\tau$ as V. By the Bayes' rule, $\hat{F}(Y \leq j \mid V))$ can be expressed as:

$$\hat{F}(Y \leq j \mid V) = \frac{\mathbb{P}(Y \leq j) \times g_1(V \mid Y \leq j)}{\mathbb{P}(Y \leq j) \times g_1(V \mid Y \leq j) + \mathbb{P}(Y > j) \times g_0(V \mid Y > j)}$$

  , where $g_0(\cdot)$ and $g_0(\cdot)$ denote conditional density of V given $(Y > j)$ or $(Y \leq j)$ respectively.
- Then we can construct a quasi log likelihood function based on $\hat{F}$

$$\frac{1}{n}\sum_{i=1}^{n}\hat{m}(X_i)\sum_{j=0}^{J}1_{\{Y=j\}}\log\left[\hat{\mathbb{P}}(Y \leq j \mid X_i, \beta, D_i, \tau) - \hat{\mathbb{P}}(Y \leq j-1 \mid X_i, \beta, D_i, \tau)\right) \tag{1}$$

SIMULATION



(a) (b)

Monte Carlo Simulation Results: t-distribution Error