因果推論 HW2

新庄紘己

2023-10-28

1

　局所線形推定量は、端点を滑らかに推定できるので、ランニング変数の閾値近傍における処置効果の推定を行う RDD と相性が良い。局所定数推定量はバンド幅にあるデータの重み付け平均を取るので、端点を十分に近似できず、逆に次数が高いと、最適な次数選択が難しく、また誤った信頼区間を推定する可能性がある。

2

a

$$ATE = E[Y(1) - Y(0)] = E[1 + u] = E[1] = 1$$

b

$e = \Phi(v) \in [0,1]$ より、確率変数 e の分布関数は

$$F(e) = Pr(e \leq x) = Pr(\Phi(v) \leq x) = Pr(v \leq \Phi^{-1}(x)) = \Phi(\Phi^{-1}(x)) = x$$

である。従って、確率変数 e の確率密度関数は

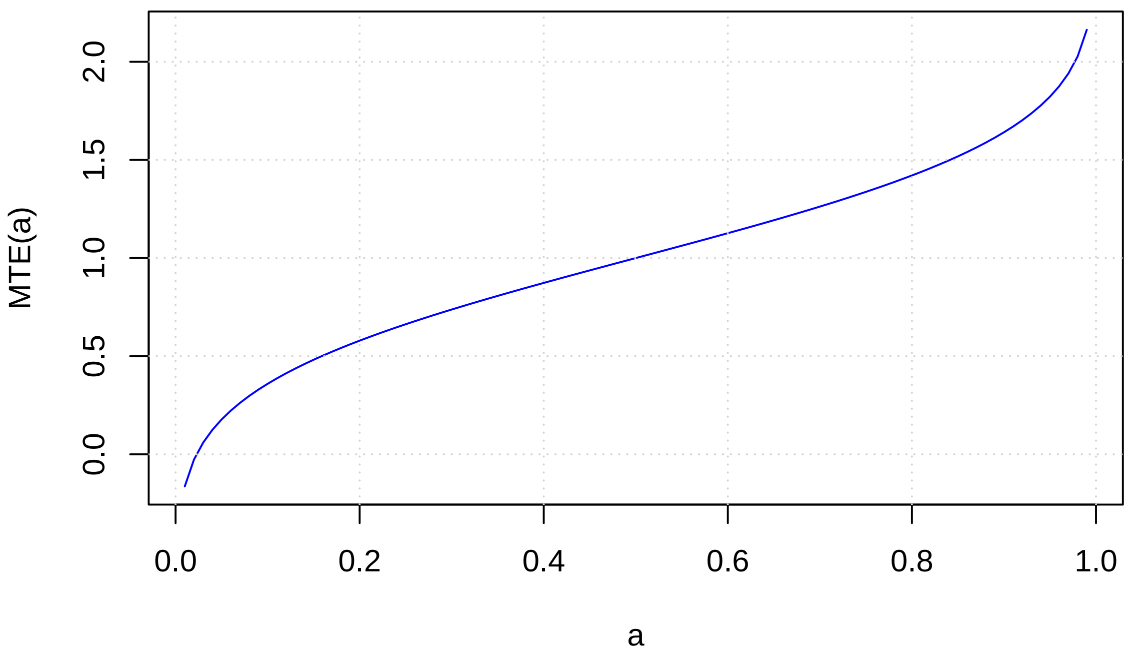
$$f(e) = \begin{cases} 1 \text{ if } x \in [0,1] \\ 0 \text{ if } x \notin [0,1] \end{cases}$$

と計算できる。これは、確率変数 e が、閉区間 $[0,1]$ 上の連続一様分布に従う定義に他ならない。

c

$$
\begin{aligned}
MTE(a) &= E[Y(1) - Y(0) \mid e = a] = E[1 + u \mid e = a] \\
&= E[1 + u \mid \Phi(v) = a] \\
&= E[1 + u \mid v = \Phi^{-1}(a)] \\
&= 1 + \frac{1}{2}\Phi^{-1}(a) \because 2\text{ 変量正規分布に従う確率変数の条件付分布の公式}
\end{aligned}
$$

```r
# Define MTE
a <- seq(0.01, 0.99, by = 0.01)
MTE <- 1 + 0.5 * qnorm(a)

# Plot MTE
plot(a, MTE, type = "l", col = "blue", xlab = "a", ylab = "MTE(a)")
grid()
```



MTE(a) は a に対して増加関数なので、処置の恩恵を受けない人ほど、処置を受け取りやすいと言える。

d

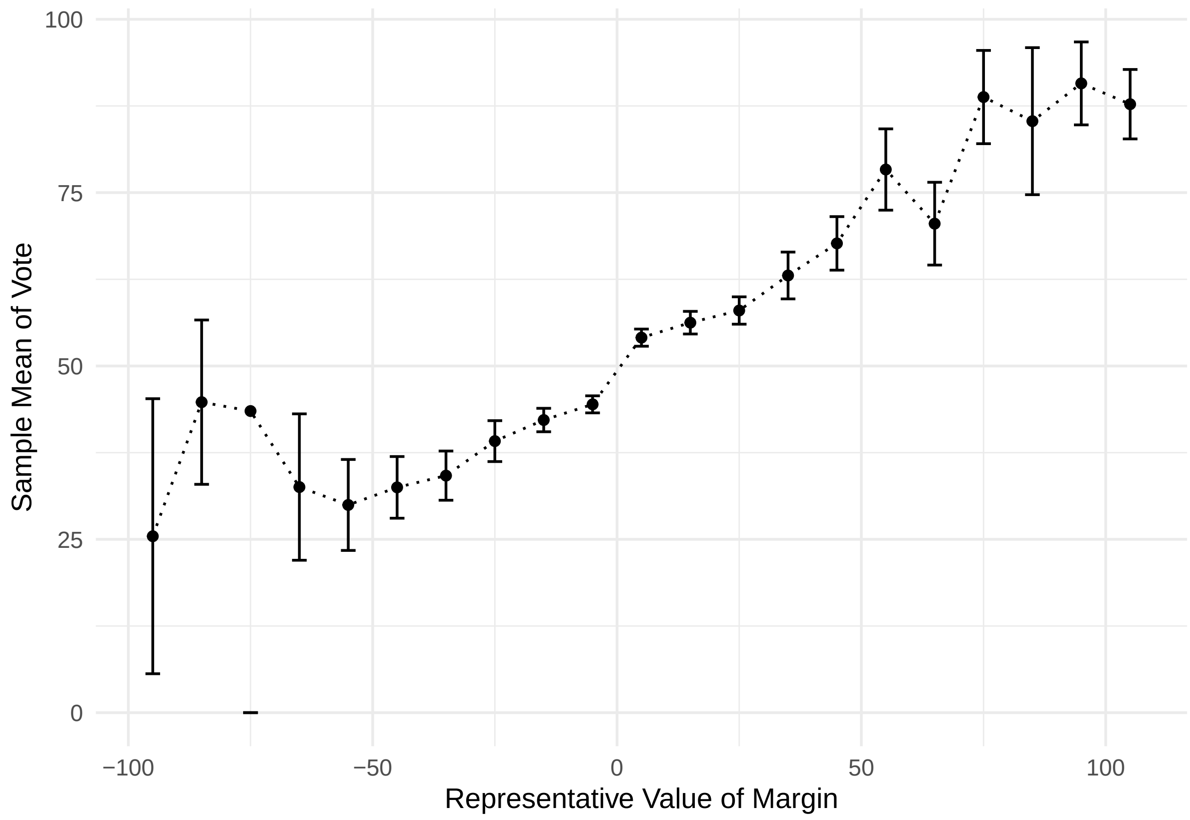
$$
\begin{aligned}
\int MTE(a)da &= \int_0^1 (1 + \frac{1}{2}\Phi^{-1}(a))da \\
&= 1 + \frac{1}{2}\int_{-\infty}^{\infty} \Phi^{-1}(\Phi(v))\phi(v)dv \\
&= 1 + \frac{1}{2}\int_{-\infty}^{\infty} v\phi(v)dv = 1 + \frac{1}{2}E[v] \\
&= 1 = ATE
\end{aligned}
$$

3.

(a)

```r
# Create the data set
library(readr)
df <- read_csv("rdrobust_senate.csv",show_col_types = FALSE)
df <- df[,c("margin","vote")]
df <- na.omit(df)

intervals <- seq(-100, 110, by = 10)
means <- c()
sds <- c()
ns <- c()

for (j in 1:21) {
  interval_mean <- mean(df$vote[df$margin>=intervals[j]&df$margin<intervals[j+1]])
  interval_sd <- sd(df$vote[df$margin>=intervals[j]&df$margin<intervals[j+1]])
  interval_n <- length(df$vote[df$margin>=intervals[j]&df$margin<intervals[j+1]])

  means <- c(means, interval_mean)
  sds <- c(sds, interval_sd)
  ns <- c(ns,interval_n)
}
midpoints <- seq(-95, 105, by = 10)
df_plot <- data.frame(midpoints,means,sds,ns)
```

```r
# Create the plot
library(ggplot2)
# Combine the data for plotting
plot_data <- data.frame(x = df_plot$midpoints, y = df_plot$means,
                        ymin = means - 1.96 * df_plot$sds / sqrt(df_plot$ns),
                        ymax = means + 1.96 * df_plot$sds / sqrt(df_plot$ns))
plot_data[is.na(plot_data)] = 0

ggplot(plot_data, aes(x = x, y = y)) +
  geom_point() +
  geom_errorbar(aes(ymin = ymin, ymax = ymax), width = 3) +
  geom_line(linetype = "dotted") +
  labs(x = "Representative Value of Margin", y = "Sample Mean of Vote") +
  theme_minimal()
```



	(1)
(Intercept)	40.921***
	(0.499)
W	21.701***
	(0.792)
Num.Obs.	1297
+ p < 0.1, * p < 0.05, ** p < 0.01, *** p < 0.001	

(b)

```
library("estimatr")
```

```
## Warning: パッケージ 'estimatr' はバージョン 4.2.3 の R の下で造られました
```

```
library("modelsummary")
```

```
## Warning: パッケージ 'modelsummary' はバージョン 4.2.3 の R の下で造られました
```

```
## Warning in !is.null(rmarkdown::metadata$output) && rmarkdown::metadata$output
## %in% : 'length(x) = 3 > 1' in coercion to 'logical(1)'
```

```
df$W <- ifelse(df$margin>=0,1,0)
b <- lm_robust(vote ~ W,data = df)
msummary(b,star = TRUE, gof_omit = "R2|R2 Adj.|AIC|BIC|RMSE")
```

この推定方法において、処置状態を表す 2 項変数の回帰係数は、ATE を表す。しかしながら、処置割当のメカニズムの情報が研究者に既知であるにも関わらず、我々はその情報を回帰式に含めていない。従って、手元にある情報を十分に活用できていない点で、この推定方法は問題がある。その解決策として。処置を受けるかどうか、ギリギリのランニング変数の人における、処置効果を識別する (RDD を用いる) ことで、より研究者の知りたい因果効果に近づける可能性がある。

(c)

```
df$WtimesMargin <- df$W*df$margin
df_c <- df[df$margin<=12 & df$margin>=-12,]
```

	(1)
(Intercept)	45.780***
	(1.261)
W	7.076***
	(1.608)
margin	0.287
	(0.196)
WtimesMargin	−0.075
	(0.246)
Num.Obs.	523

+ $p < 0.1$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

```
c <- lm_robust(vote ~ 1 + W + margin + WtimesMargin,data = df_c)
msummary(c,star = TRUE, gof_omit = "R2|R2 Adj.|AIC|BIC|RMSE")
```

W の係数の推定量は、$E[Y(1) - Y(0) \mid margin = 0]$ の推定量である。図より条件付き平均処置効果は、1% の有意水準で統計的に有意に正の影響を与えると言える。