PS 2

Hyoungchul Kim

2025-04-14

Part 1

First, read in the data

```
library(tidyverse)
library(data.table)

data <- read_csv("data/middle_kink.csv")

# view the data
data %>% head()
```

```
# A tibble: 6 x 2
  income_bin     n
       <dbl> <dbl>
1     151250   955
2     153750   982
3     156250   894
4     158750   873
5     161250   810
6     163750   851
```

a.

Now we will plot the publication-quality histogram of the earnings distribution:

```r
library(ggtext)

earning_dist <- data %>%
  ggplot(aes(x=income_bin, y = n)) +
  geom_col(fill = "lightblue") +
  geom_point() +
  # geom_smooth(method = "lm", se = FALSE, color = "black") +
  geom_line() +
  scale_x_continuous(labels = scales::label_number(scale = 0.001, prefix =
  ↪  "$")) +
  geom_vline(xintercept = 363750, color="red", shape="solid") +
  # annotate("text", label = "kink", x = 380000, y = 800, size = 5, colour =
  ↪  "black") +
  labs(title = "**Histogram of the earnings distribution**",
       x = "**Income Bin (1000s)**",
       y = "**Number of Observations**") +
  theme_bw() +
  theme(
    plot.title = element_markdown(size = 16, hjust = 0.5),
    axis.title.x = element_markdown(size = 14),
    axis.title.y = element_markdown(size = 14),
    axis.text.x = element_text(size = 12),
    axis.text.y = element_text(size = 12)
  )
earning_dist
```

Histogram of the earnings distribution
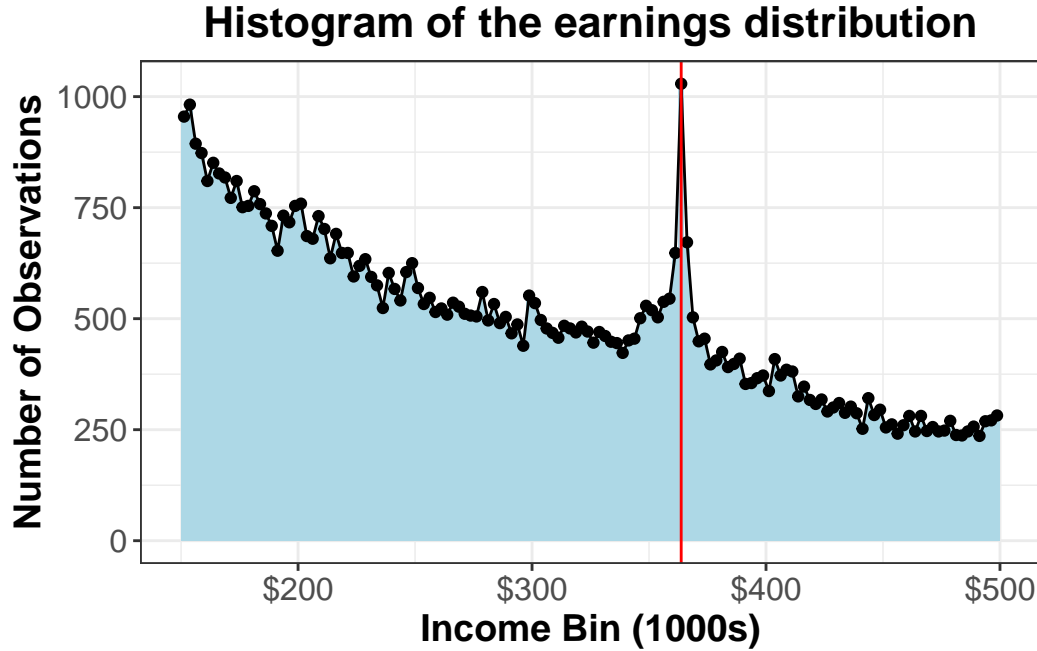
b.

We will be following Saez (2010) to construct the equation to retrieve the elasticity e. Note for our case, the kink happens as $z^* = 363750$ and marginal tax rate changes from 0.21 to 0.28. We need to use equation (5) in the paper to get the elasticity. The equation is as follows:

$$B = z^* \left[\left(\frac{1-t_0}{1-t_1} \right)^e - 1 \right] \frac{h(z^*)_- + h(z^*)_+ \left/ \left(\frac{1-t_0}{1-t_1} \right)^e \right.}{2}.$$

In order to compute B, we need to decide δ to calculate the width we will use to calculate excess bunching. We will use the "simplest method" mentioned in the paper which is to select δ graphically such that the full excess bunching is included in the band $(z^* - \delta + z^* + \delta)$. In our case, it seems to be about $\delta = 10$ (Note that since out data is in income bin of width 2,500, this is equivalent to 25,000 difference). Numerically, it will be calculated as follows (this is just following the equation (6) in the paper):

```
A <- data %>% count(wt=n) %>% pull()
A
```

```
[1] 69656
```

```r
b <- function(delta) {

B1  <- data %>%
  filter(income_bin >= 365000 - 2500*delta & income_bin <= 365000 +
  ↪  2500*delta) %>%  # 25,000 differences
  count(wt=n) %>%
  pull() #8,574

B2 <- data %>%
  filter(income_bin >= 365000 - 2500*2*delta & income_bin <= 365000 -
  ↪  2500*delta) %>%
  count(wt=n) %>%
  pull() # 3615

B3 <- data %>%
  filter(income_bin >= 365000 + 2500*delta & income_bin <= 365000 +
  ↪  2500*2*delta) %>%
  count(wt=n) %>%
  pull() # 2982

B = (B1 - B2 - B3) / A

}

b1 <- function(delta) {

B1  <- data %>%
  filter(income_bin >= 365000 - 2500*delta & income_bin <= 365000 +
  ↪  2500*delta) %>%  # 25,000 differences
  count(wt=n) %>%
  pull() #8,574

}

b2 <- function(delta) {

B2 <- data %>%
  filter(income_bin >= 365000 - 2500*2*delta & income_bin <= 365000 -
  ↪  2500*delta) %>%
  count(wt=n) %>%
  pull() # 3615
```

```
}

b3 <- function(delta) {

B3 <- data %>%
  filter(income_bin >= 365000 + 2500*delta & income_bin <= 365000 +
  ↪  2500*2*delta) %>%
  count(wt=n) %>%
  pull() # 2982
}


B <- b(2)
B1 <- b1(2)
B2 <- b2(2)
B3 <- b3(2)

B #1977
```

```
[1] 0.01241817
```

```
B1
```

```
[1] 2852
```

```
B2
```

```
[1] 1083
```

```
B3
```

```
[1] 904
```

Now we also need to compute two h in the main equation. Empirically we can calculate this by dividing B2, B3 by δ respectively.

```
h_min = (B2 / 5000) / A
h_plus = (B3 / 5000) / A

h_min
```

```
[1] 3.109567e-06
```

```
h_plus
```

```
[1] 2.595613e-06
```

Finally, we can plug in the values we got from the data and get the elasticity e. Here, we are just basically getting the solution by plugging in the empirical numbers we computed from the data into the main equation:

```
# Define the function whose root we want to find
f <- function(e) {

  z   <- 365000
  ratio <- (1 - 0.07) / (1 - 0.21)

  z * (ratio^e - 1) * ( (h_min + (h_plus / ratio^e)) / 2) - B
}

# Use uniroot to solve f(e) = 0 in a reasonable range for e
result <- uniroot(f, lower = -1, upper = 1)

# Extract the solution
e_solution <- result$root
print(e_solution)
```

```
[1] 0.07306182
```

c.

```
library(bunchr)
# # analyzing a kink
# ability_vec <- 4000 * rbeta(100000, 2, 5)
# earning_vec <- sapply(ability_vec, earning_fun, 0.2, 0, 0.2, 0, 1000)
# earning_vec
# # bunch_viewer(earning_vec, 1000, 20, 20, 1, 1, binw = 20)
# estim <- bunch(earning_vec, 1000, 0, 0.2, Tax = 0, 20, 20, 1, 1,
# binw = 20, draw=TRUE, nboots = 0, seed = 16)
# estim$e
```
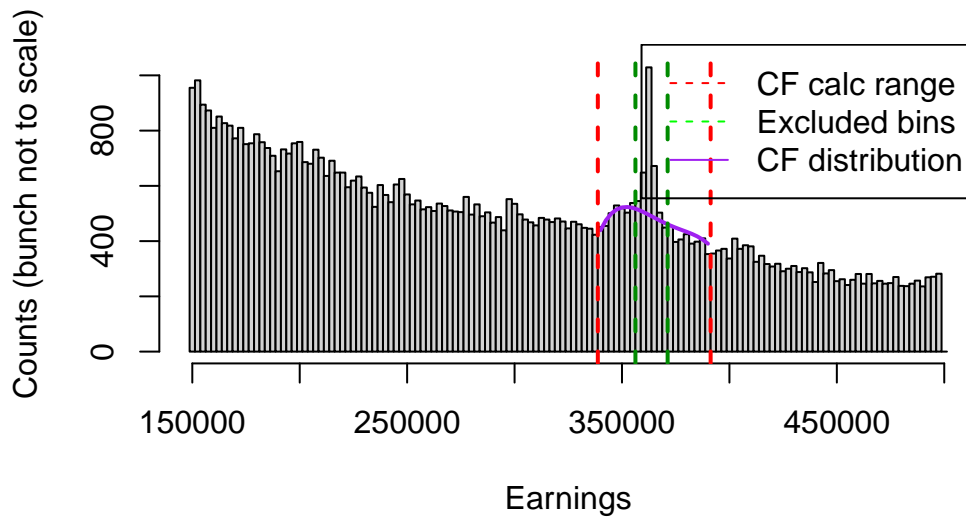
```
# Step 1: Expand binned data into a raw vector
z_vector <- data %>%
  rowwise() %>%
  summarise(vec = list(rep(income_bin, n))) %>%
  pull(vec) %>%
  unlist()

# Step 2: Estimate bunching
estim <- bunch(
  earnings = z_vector,
  zstar = 365000,
  binw = 2500,
  t1 = 0.07,
  t2 = 0.21,
  cf_start = 10,
  cf_end = 10,
  exclude_before = 3,
  exclude_after = 2,
  # poly_size = 2,
  draw = TRUE,
  nboots = 100
)
```

```
# Estimate of the elasticity from the package
estim$e
```

```
[1] 0.077713
```

d.

- Different: (1) setting the bunching width, (2) excluding bins, (3) polynomial order.

Part 2

There is no straightforward answer to this question. Before we started to consider the dynamic nature of the taxation policy, the consensus was that we should not tax the capital. The reason was similar to our consensus on not wanting to tax the commodity. By Atkinson-? result, we should only be taxing through income as taxing through commodity or capital will further distort the behavior of people.

But our consensus changes if we move onto a dynamic setting. In fact, some papers have shown that the capital taxation should be positive in the dynamic setting. The intuition for this is because in the dynamic setting people try to save more for the future to account for any possible risk in the future such as negative shock on their ability.Hence it leads to a negative fiscal externality for the government revenue. In order to alleviate this, the government should try to tax the capital. In this case, we could say it should tax the capital to internalize this negative externality.

However, setting positive capital taxation might not be that important in terms of policy perspective. This is because some papers have shown that while it is true that capital tax is positive in dynamic setting, its actual magnitude seems to be small (close to zero). Thus in real life policy, it might not make much of a difference whether you tax capital (very slightly) or not.

In the end, I think we can summarize it as follows:

1. In static model, it was not recommended to tax the capital.

2. In dynamic setup, the results from static case changed and the optiaml capital taxation should be positive.

3. However, even accounting for this the real life optimal tax for the capital should be near zero.

References

Saez, Emmanuel. 2010. "Do Taxpayers Bunch at Kink Points?" *American Economic Journal: Economic Policy* 2 (3): 180–212. https://doi.org/10.1257/pol.2.3.180.