PS 2

Hyoungchul Kim

2025-04-06

Part 1

First, read in the data

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

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

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

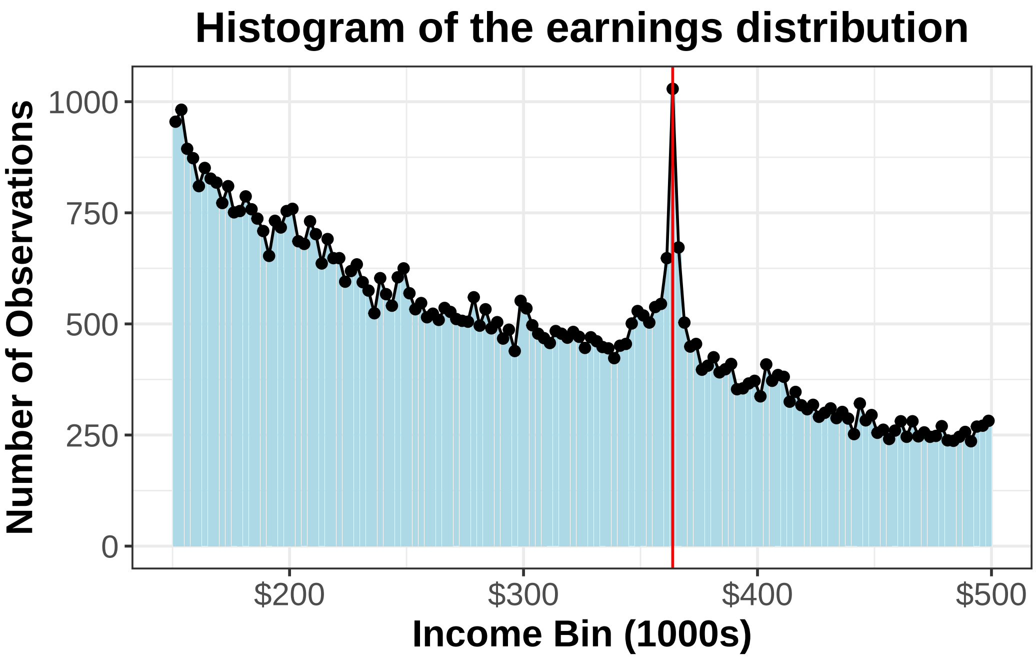
```
# A tibble: 6 x 2
  income_bin     n
       <dbl> <dbl>
1     151250   955
2     153750   982
3     156250   894
4     158750   873
5     161250   810
6     163750   851
```

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
```



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:

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

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 $\delta = 8$ (Note that since out data is in income bin of width 2500, this is equivalent to 20,000 difference). Numerically, it will be calculated as follows (this is just following the equation (6) in the paper):

```
B1  <- data %>%
  filter(income_bin >= 342750 & income_bin <= 383650) %>%  # 20,000
  ↪ differences
  count(wt=n) %>%
  pull() #8,574
```

```r
B2 <- data %>%
  filter(income_bin >= 323750 & income_bin <= 342750) %>%
  count(wt=n) %>%
  pull() # 3615

B3 <- data %>%
  filter(income_bin >= 383750 & income_bin <= 402750) %>%
  count(wt=n) %>%
  pull() # 2982

B = B1 - B2 - B3

B #1977
```

```
[1] 1977
```

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

```r
h_min = B2 / 20000
h_plus = B3 / 20000

h_min
```

```
[1] 0.18075
```

```r
h_plus
```

```
[1] 0.1491
```

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:

```r
# Define the function whose root we want to find
f <- function(e) {
  363750 * (((1-0.21) / (1-0.28))^e - 1) * ( ((h_min + h_plus) / ((1-0.21) /
   ↪  (1-0.28))^e) / 2) - B
}
```

```
# Use uniroot to solve f(e) = 0 in a reasonable range for e
result <- uniroot(f, lower = -10, upper = 10)

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

```
[1] 0.361171
```

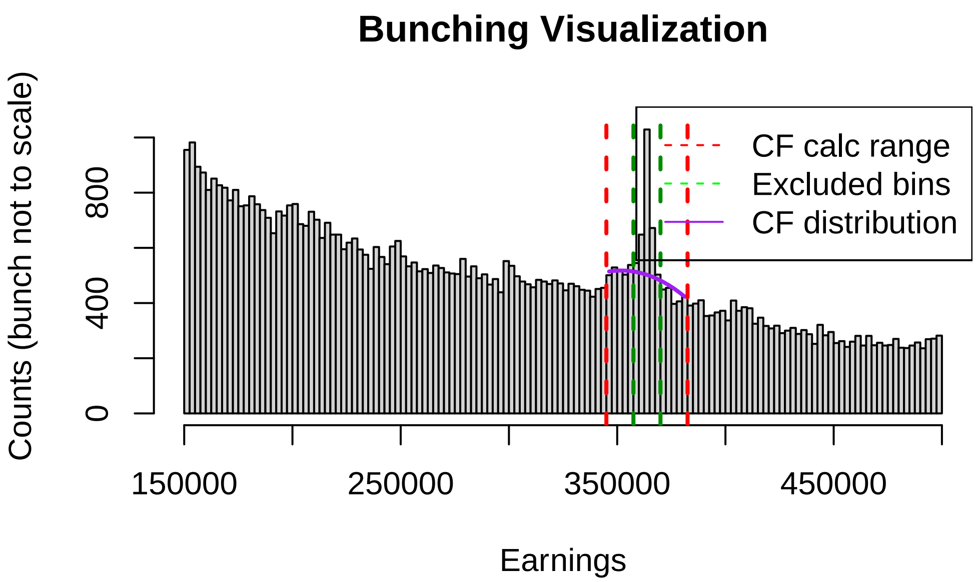
c.

```
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 = 363750,
  binw = 2500,
  t1 = 0.21,
  t2 = 0.28,
  cf_start = 7,
  cf_end = 7,
  exclude_before = 2,
  exclude_after = 2,
```

```r
  poly_size = 2,
  draw = TRUE
)
```




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

```
[1] 0.1318855
```

d.

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

Part 2

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.