ECON-613 HW #4

Peter Kim

2019-04-03

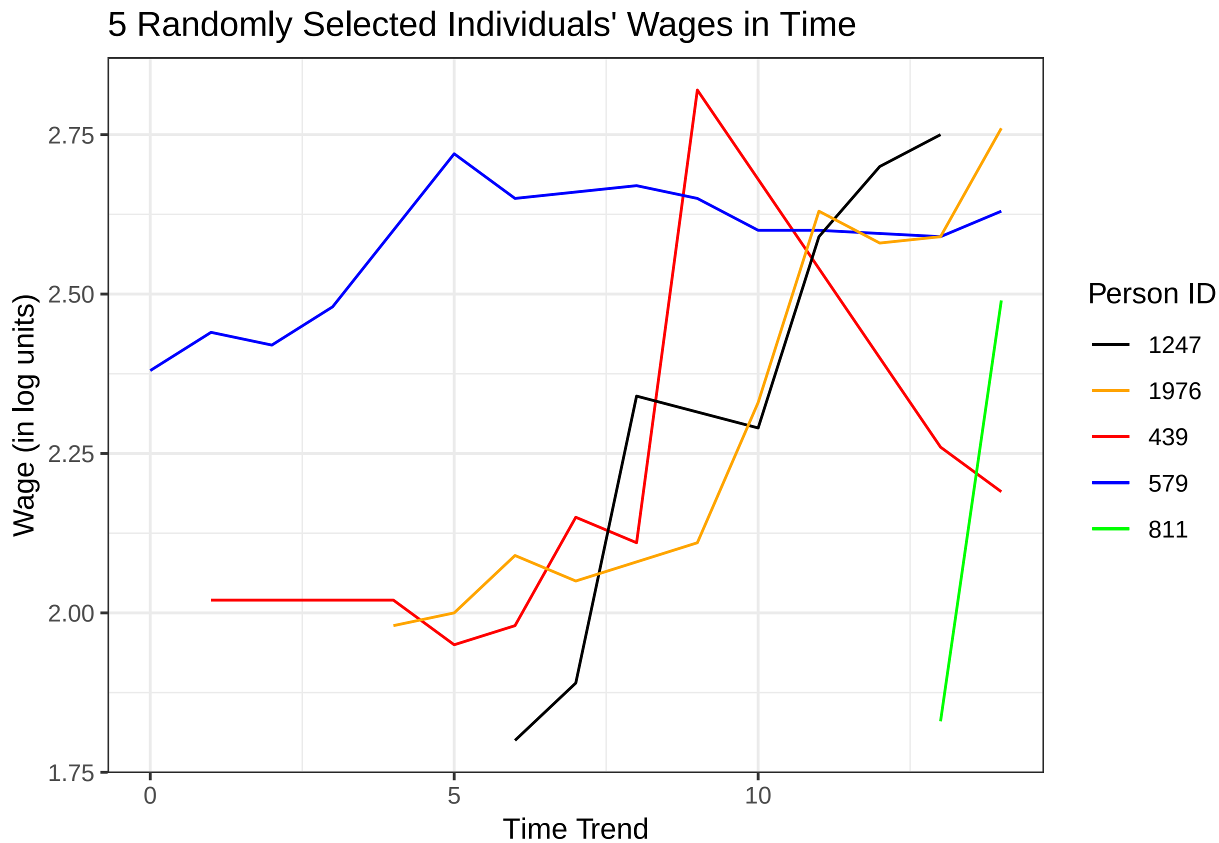
Exercise 1 Data

```
#Set Seed
set.seed(1)

#Read Data
data = read_csv("~/ECON-613/Assignment #4/HW #4/Koop-Tobias.csv")
```

```
## Parsed with column specification:
## cols(
##   PERSONID = col_double(),
##   EDUC = col_double(),
##   LOGWAGE = col_double(),
##   POTEXPER = col_double(),
##   TIMETRND = col_double(),
##   ABILITY = col_double(),
##   MOTHERED = col_double(),
##   FATHERED = col_double(),
##   BRKNHOME = col_double(),
##   SIBLINGS = col_double()
## )
```

```
#Identifying Unique Person ID
uniq_ID = unique(data$PERSONID) %>% length()

#Generating 5 Random Indices
ind = sample(x = 1:uniq_ID, size = 5, replace = FALSE)

#Representing the Panel Dimension of Wages for 5 Randomly Selected Individuals
rand5 = data %>%
  filter(PERSONID == ind[1] | PERSONID == ind[2] | PERSONID == ind[3] |
         PERSONID == ind[4] | PERSONID == ind[5]) %>%
  select(PERSONID, TIMETRND, LOGWAGE)

rand5.1 = rand5 %>% filter(PERSONID == unique(rand5$PERSONID)[1])
rand5.2 = rand5 %>% filter(PERSONID == unique(rand5$PERSONID)[2])
rand5.3 = rand5 %>% filter(PERSONID == unique(rand5$PERSONID)[3])
rand5.4 = rand5 %>% filter(PERSONID == unique(rand5$PERSONID)[4])
rand5.5 = rand5 %>% filter(PERSONID == unique(rand5$PERSONID)[5])

ggplot() +
  geom_line(data = rand5.1,
            mapping = aes(x = TIMETRND, y = LOGWAGE, color = "439")) +
  geom_line(data = rand5.2,
            mapping =  aes(x = TIMETRND, y = LOGWAGE, color = "579")) +
  geom_line(data = rand5.3,
```

```
            mapping =  aes(x = TIMETRND, y = LOGWAGE, color = "811")) +
  geom_line(data = rand5.4,
            mapping =  aes(x = TIMETRND, y = LOGWAGE, color = "1247")) +
  geom_line(data = rand5.5,
            mapping =  aes(x = TIMETRND, y = LOGWAGE, color = "1976")) +
  scale_color_manual(
    name = "Person ID",
    values = c("439" = "red",
               "579" = "blue",
               "811" = "green",
               "1247" = "black",
               "1976" = "orange")
  ) +
  labs(
    x = "Time Trend",
    y = "Wage (in log units)",
    title = "5 Randomly Selected Individuals' Wages in Time"
  ) +
  theme_bw()
```



Exercise 2 Random Effects

```
#Implementing Linear Regresson
randlm = lm(LOGWAGE ~ EDUC + POTEXPER, data = data)
```

```r
randlm_coef = coefficients(randlm) %>% as.data.frame()
colnames(randlm_coef) = c("OLS Coefficients")

kable(randlm_coef, digits = 4,
      caption = "Table of OLS Coefficients for Random Effects Model")
```

Table 1: Table of OLS Coefficients for Random Effects Model

	OLS Coefficients
(Intercept)	0.7942
EDUC	0.0939
POTEXPER	0.0374

Comment

$\hat{\beta}_{intercept} = 0.7942$: When education and potential experience are 0, log wage will be 0.7942. This quantity, however, is meaningless since no individuals are legally permitted to have 0 units of education.

$\hat{\beta}_{EDUC} = 0.0939$: Ceteris paribus, a one unit increase in education will increase log wage by 0.0939 on average.

$\hat{\beta}_{POTEXPER} = 0.0374$: Ceteris paribus, a one unit increase in potential experience will increase log wage by 0.0374 on average.

Exercise 3

Between Estimator

```r
between_data = data %>%
  select(PERSONID, EDUC, LOGWAGE, POTEXPER, TIMETRND) %>%
  group_by(PERSONID) %>%
  summarise(mean_LOGWAGE = mean(LOGWAGE),
            mean_EDUC = mean(EDUC),
            mean_POTEXPER = mean(POTEXPER))

between_coef =
  lm(mean_LOGWAGE ~ mean_EDUC + mean_POTEXPER, data = between_data) %>%
  coefficients() %>%
  as.data.frame()
```

Within Estimator

```r
#Creating Initial Within Data
within_data = left_join(data, between_data, by = "PERSONID")

#Updating Within Data - Adding Columns of Y - Y_bar and X - X_bar
within_data2 = within_data %>%
```

```r
  mutate(bet_resp = LOGWAGE - mean_LOGWAGE,
         bet_EDUC = EDUC - mean_EDUC,
         bet_POTEXPER = POTEXPER - mean_POTEXPER) %>%
  select(PERSONID, bet_resp, bet_EDUC, bet_POTEXPER)

#Calculating Within Coefficients
within_coef =
  lm(bet_resp ~ bet_EDUC + bet_POTEXPER - 1, data = within_data2) %>%
  coefficients() %>%
  as.data.frame()
```

First Time Difference Estimator

```r
#Creating First Difference Data
first_data =
  data %>% group_by(PERSONID) %>%
  mutate(LOGWAGE_Diff = LOGWAGE - lag(LOGWAGE),
           EDUC_Diff = EDUC - lag(EDUC),
           POTEXPER_Diff = POTEXPER - lag(POTEXPER)) %>%
  select(PERSONID, LOGWAGE_Diff, EDUC_Diff, POTEXPER_Diff) %>%
  na.omit()

first_coef =
  lm(LOGWAGE_Diff ~ EDUC_Diff + POTEXPER_Diff - 1, data =  first_data) %>%
  coefficients() %>%
  as.data.frame()
```

Here, we create a table of $\hat{\beta}_{EDUC}$ and $\hat{\beta}_{POTEXPER}$

```r
coef_data = data.frame(
  c(between_coef %>% unlist() %>% as.numeric()),
  c(NA, within_coef %>% unlist() %>% as.numeric()),
  c(NA, first_coef %>% unlist() %>% as.numeric)
)
colnames(coef_data) = c("Between", "Within", "First")
rownames(coef_data) = c("Intercept", "Education", "POTEXPER")

kable(coef_data, digits = 4,
      caption = "Coefficients Under Different Models")
```

Table 2: Coefficients Under Different Models

	Between	Within	First
Intercept	0.8456	NA	NA
Education	0.0931	0.1237	0.0479
POTEXPER	0.0260	0.0386	0.0329

Comparison of $\hat{\beta}_{education}$'s:

We observe that $\hat{\beta}_{between}$, $\hat{\beta}_{within}$, and $\hat{\beta}_{first}$ are all positive. Each model believes that a unit increase in education will increase wage (or log wage). The "within" model has the largest magnitude, while the "first-difference" model has the smallest magnitude.

Comparison of $\hat{\beta}_{POTEXPER}$'s:

Similar to education, we observe that $\hat{\beta}_{between}$, $\hat{\beta}_{within}$, and $\hat{\beta}_{first}$ are all positive. Each model believes that a unit increase in potential experience will increase wage (or log wage).. The "within" model has the largest magnitude, while the "between" model has the smallest magnitude.

Exercise 4