A3

Yue Han

4/24/2021

```
library(tidyverse)
```

```
## -- Attaching packages ---------------------------------------------------------------- tidyverse 1.3
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.3     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.3.1     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------------------------------- tidyverse_conflicts
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(xts)
```

```
## Loading required package: zoo
```

```
##
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric
```

```
##
## Attaching package: 'xts'
```

```
## The following objects are masked from 'package:dplyr':
##
##     first, last
```

```
library(plm)
```

```
##
## Attaching package: 'plm'
```

```
## The following objects are masked from 'package:dplyr':
##
##     between, lag, lead
```

Exercise 1 Links to the datasets

```
population <- read.csv("https://www.dropbox.com/s/s38cde88670y5mw/population.csv?dl=1")
crime_long <- read.csv("https://www.dropbox.com/s/t3vushurhm3s5my/crime_long.csv?dl=1")
officers <- read.csv("https://www.dropbox.com/s/8q2fpdb7phy86m8/officers.csv?dl=1")
```

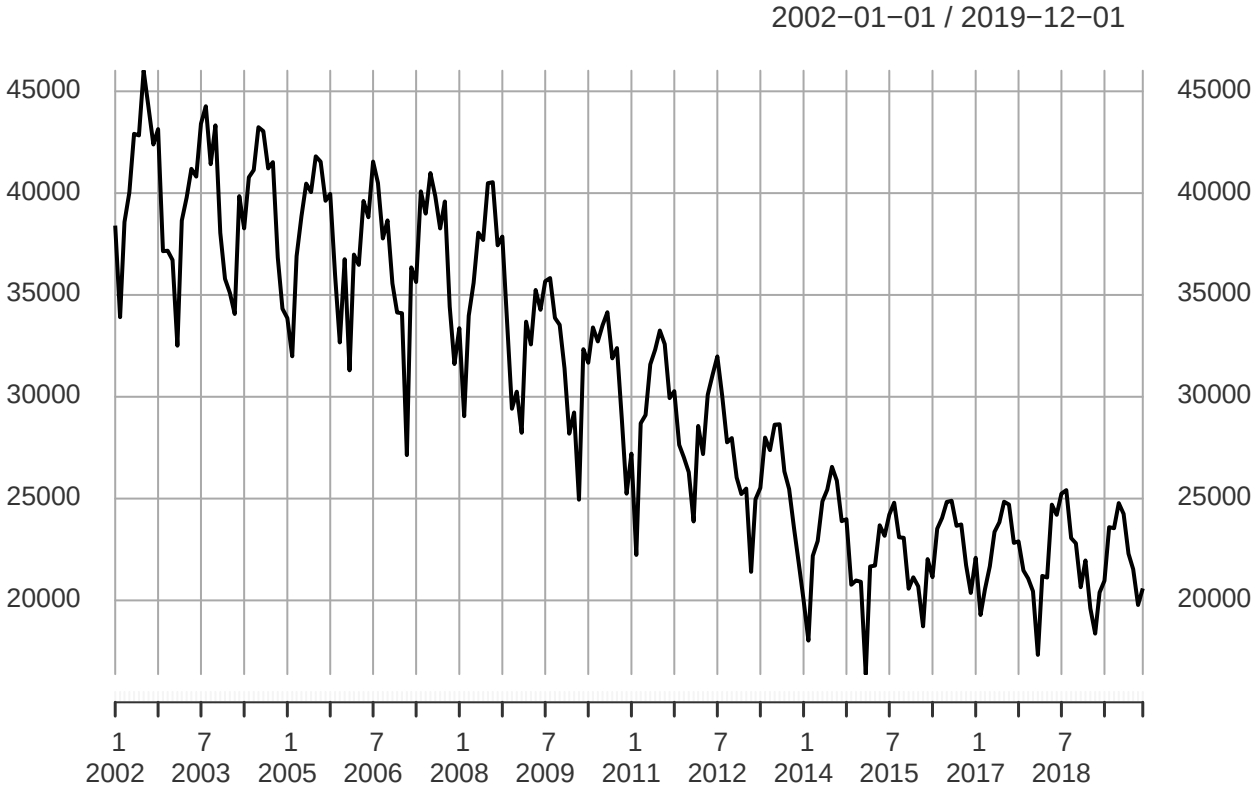
Exercise 2 Data Manipulation

Calculate total crime per month and plot the time series of crime

```
# total crime per month
crime_monthly <- crime_long %>%
  group_by(crime_month) %>%
  summarize(crime_by_month = sum(crimes))
```

```
## `summarise()` ungrouping output (override with `.groups` argument)
```

```
# plot time series of crime
crime_monthly_xts <- xts(crime_monthly$crime_by_month, as.Date(crime_monthly$crime_month, format='%Y-%m
plot(crime_monthly_xts,type = 'l',main=' ')
```



Merge the two datasets by districts-units and period

```
crime_population <- merge(crime_long, population,by.x = c("district","crime_month"), by.y = c("district
```

Construct a panel data of unit over time with the following variables

– Total crimes per resident – Violent crimes per resident – Property crimes per resident – Median income – Share of black, Hispanic, and white residents

```
panel_data <- crime_population %>%
  mutate(
    violent_crimes = case_when(
      crime_type == "violent" ~ crimes,
      TRUE ~ 0L
    ),
```

```
    property_crimes = case_when(
      crime_type == "property" ~ crimes,
      TRUE ~ 0L
    )
  ) %>%
  group_by(district,crime_month) %>%
  summarize(
    total_crimes = sum(crimes),
    violent_crimes = sum(violent_crimes),
    property_crimes = sum(property_crimes),
    median_income = p50_inc,
    share_of_black = tot_black/tot_pop,
    share_of_hisp = tot_hisp/tot_pop,
    share_of_white = tot_white/tot_pop
  ) %>%
  distinct()
```

```
## `summarise()` regrouping output by 'district', 'crime_month' (override with `.groups` argument)
```

Exercise 3 Panel Data: Introduction

```
df <- merge(officers, panel_data, by.x = c("month", "unit"), by.y=c("crime_month", "district"), all.x=TR
panel_df <- pdata.frame(df,index=c("NUID","month"))
```

```
# pooled model
pooled <- plm(
  formula = arrest ~ tenure + total_crimes + median_income + share_of_black + share_of_hisp + share_of_w
  data = panel_df,
  model = "pooling"
)
# estimator
# beta
pooled$coefficients[2]
```

```
##       tenure
## -4.16066e-06
```

```
# gamma
pooled$coefficients[3:7]
```

```
##   total_crimes  median_income share_of_black  share_of_hisp share_of_white
##   2.229548e-07   1.617511e-08  -8.101603e-03  -5.362641e-03  -1.207314e-02
```

```
# we can also use ols
lm_pooled <- lm(
  formula = arrest ~ tenure + total_crimes + median_income + share_of_black + share_of_hisp + share_of_w
  data = df
)
# estimators
# beta
lm_pooled$coefficients[2]
```

```
##       tenure
## -4.16066e-06
```

```r
# gamma
lm_pooled$coefficients[3:7]
```

```
##   total_crimes  median_income share_of_black  share_of_hisp share_of_white
##   2.229548e-07   1.617511e-08  -8.101603e-03  -5.362641e-03  -1.207314e-02
```

Exercise 4 Panel Data: More controls

```r
# fixed effect model
fe1 <- plm(
    formula = arrest ~ tenure + total_crimes + median_income + share_of_black + share_of_hisp + share_o
    effect = "time",
    data = panel_df,
    model = "within"
)
# estimators
# beta
fe1$coefficients[1]
```

```
##        tenure
## -3.809782e-06
```

```r
# gamma
fe1$coefficients[2:6]
```

```
##   total_crimes  median_income share_of_black  share_of_hisp share_of_white
##  -6.320360e-06  -4.910055e-07  -9.200883e-02  -1.398480e-01  -1.012976e-01
```

```r
# psi
fe1$coefficients[7:30]
```

```
##  factor(unit)2  factor(unit)3  factor(unit)4  factor(unit)5  factor(unit)6
##  -2.437791e-02  -2.001211e-02  -3.632805e-05  -1.580178e-02  -1.385754e-02
##  factor(unit)7  factor(unit)8  factor(unit)9 factor(unit)10 factor(unit)11
##  -2.108532e-02   2.342640e-02   3.109578e-03   7.681503e-03  -1.133599e-02
## factor(unit)12 factor(unit)13 factor(unit)14 factor(unit)15 factor(unit)16
##   4.412439e-04  -3.063905e-03   2.440896e-02  -2.027399e-02   7.288093e-03
## factor(unit)17 factor(unit)18 factor(unit)19 factor(unit)20 factor(unit)21
##   1.080354e-03   4.016511e-03   8.935392e-03  -1.544846e-02  -3.671653e-02
## factor(unit)22 factor(unit)23 factor(unit)24 factor(unit)25
##  -6.345306e-04  -1.097174e-02  -1.539638e-02   2.202629e-02
```

```r
# kappa
fixef(fe1)
```

```
## 2007-03-01 2007-04-01 2007-05-01 2007-06-01 2007-07-01 2007-08-01 2007-09-01
##    0.64160    0.63174    0.64581    0.62099    0.63370    0.62689    0.63900
## 2007-10-01 2007-11-01 2007-12-01 2008-01-01 2008-02-01 2008-03-01 2008-04-01
##    0.64170    0.63944    0.62765    0.62037    0.62824    0.62913    0.64405
## 2008-05-01 2008-06-01 2008-07-01 2008-08-01 2008-09-01 2008-10-01 2008-11-01
##    0.65507    0.63722    0.63615    0.63721    0.64554    0.63282    0.63463
## 2008-12-01 2009-01-01 2009-02-01 2009-03-01 2009-04-01 2009-05-01 2009-06-01
##    0.64711    0.63017    0.63039    0.62728    0.63126    0.63463    0.64038
## 2009-07-01 2009-08-01 2009-09-01 2009-10-01 2009-11-01 2009-12-01 2010-01-01
##    0.64164    0.63151    0.63138    0.63459    0.63096    0.62629    0.63799
## 2010-02-01 2010-03-01 2010-04-01 2010-05-01 2010-06-01 2010-07-01 2010-08-01
##    0.63979    0.62729    0.64325    0.63276    0.63030    0.62264    0.64605
```

```
## 2010-09-01 2010-10-01 2010-11-01 2010-12-01 2011-01-01 2011-02-01 2011-03-01
##    0.63839    0.63362    0.62700    0.63141    0.62797    0.63078    0.63162
## 2011-04-01 2011-05-01 2011-06-01 2011-07-01 2011-08-01 2011-09-01 2011-10-01
##    0.63498    0.63564    0.63629    0.64018    0.63607    0.62483    0.64667
## 2011-11-01 2011-12-01 2012-01-01 2012-02-01 2012-03-01 2012-04-01 2012-05-01
##    0.62920    0.63646    0.62572    0.63915    0.63799    0.64075    0.64053
## 2012-06-01 2012-07-01 2012-08-01 2012-09-01 2012-10-01 2012-11-01 2012-12-01
##    0.63411    0.63685    0.63105    0.63849    0.62160    0.63964    0.61930
## 2013-01-01 2013-02-01 2013-03-01 2013-04-01 2013-05-01 2013-06-01 2013-07-01
##    0.63997    0.62809    0.62032    0.63968    0.65028    0.64323    0.65370
## 2013-08-01 2013-09-01 2013-10-01 2013-11-01 2013-12-01 2014-01-01 2014-02-01
##    0.63344    0.63921    0.62747    0.63036    0.63949    0.62121    0.63867
## 2014-03-01 2014-04-01 2014-05-01 2014-06-01 2014-07-01 2014-08-01 2014-09-01
##    0.63173    0.61755    0.63027    0.63441    0.64175    0.63506    0.63723
## 2014-10-01 2014-11-01 2014-12-01 2015-01-01 2015-02-01 2015-03-01 2015-04-01
##    0.63444    0.62763    0.62915    0.63125    0.62686    0.64033    0.63265
## 2015-05-01 2015-06-01 2015-07-01 2015-08-01 2015-09-01 2015-10-01 2015-11-01
##    0.62945    0.62906    0.62754    0.63083    0.63516    0.62784    0.63849
## 2015-12-01 2016-01-01 2016-02-01 2016-03-01 2016-04-01 2016-05-01 2016-06-01
##    0.62880    0.63083    0.62907    0.62164    0.64599    0.64603    0.63103
## 2016-07-01 2016-08-01 2016-09-01 2016-10-01 2016-11-01 2016-12-01 2017-01-01
##    0.62535    0.61696    0.63463    0.63929    0.62219    0.62327    0.63626
## 2017-02-01 2017-03-01 2017-04-01 2017-05-01 2017-06-01 2017-07-01 2017-08-01
##    0.63033    0.64223    0.63043    0.64458    0.62398    0.63005    0.63351
## 2017-09-01 2017-10-01 2017-11-01 2017-12-01 2007-01-01 2007-02-01
##    0.62768    0.62649    0.61933    0.62725    0.63617    0.63892
```

Exercise 5 Panel Data: Individual fixed effects

Implement a within, between, and first difference estimator for the parameter beta. Then, compare the estimated values.

```
# within
fe2 <- plm(
    formula = arrest ~ tenure + total_crimes + median_income + share_of_black + share_of_hisp + share_o
    effect = "twoway",
    data = panel_df,
    model = "within"
)
```

```
# between
fe3 <- plm(
    formula = arrest ~ tenure + total_crimes + median_income + share_of_black + share_of_hisp + share_o
    effect = "individual",
    data = panel_df,
    model = "between"
)
```

```
# fd
fe4 <- plm(
    formula = arrest ~ tenure + total_crimes + median_income + share_of_black + share_of_hisp + share_o
    effect = "individual",
    data = panel_df,
    model = "fd"
)
```

```r
# compare beta
est_betas <- c(fe2$coefficients[1], fe3$coefficients[2], fe4$coefficients[2])
names(est_betas) <- c("within", "between", "fd")
est_betas
```

```
##        within       between            fd
## -2.767569e-04 -1.754975e-05  5.430658e-03
```

within and between estimated beta estimator are both negative, first difference estimated beta is positive.

Use a GMM approach to estimate all parameters (including fixed effects) in one step.

one-step GMM:

$$\hat{\beta}_{2SLS} = [X'Z(Z'Z)^{-1}Z'X]^{-1}X'Z(Z'Z)^{-1}Z'y$$

if X = Z, it is just estimator of OLS.