```
#####################################
## Author:   Minke Li
## Date:     2018-04-1
## Title:    ps3.R
## Purpose:  Analyze the data in file hprice1.csv, beveridge.csv, JTRAIN.csv
#####################################

rm(list=ls()) #drop all variables

library(data.table)
library(sandwich)
library(lmtest)
library(tseries)
library(plm)

##Question1

## Data import
#####################################
##import commands
context1 <- fread('hprice1.csv')

##              storage   display    value
##variable name   type    format     label      variable label
##------------------------------------------------------------------------------------------------
##price           float   %9.0g                 house price, $1000s
##assess          float   %9.0g                 assessed value, $1000s
##bdrms           byte    %9.0g                 number of bdrms
##lotsize         float   %9.0g                 size of lot in square feet
##sqrft           int     %9.0g                 size of house in square feet
##colonial        byte    %9.0g                 =1 if home is colonial style

##Run the linear model
model1 <- lm(price~bdrms+lotsize+sqrft,data = context1)

##OLS test
coeftest(model1,vcov.=vcov)
##t test of coefficients:
##               Estimate  Std. Error t value  Pr(>|t|)
##(Intercept) -2.1770e+01  2.9475e+01 -0.7386  0.462208
##bdrms        1.3853e+01  9.0101e+00  1.5374  0.127945
##lotsize      2.0677e-03  6.4213e-04  3.2201  0.001823 **
##sqrft        1.2278e-01  1.3237e-02  9.2751 1.658e-14 ***
##  ---
##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##White-corrected test
coeftest(model1,vcov.=vcovHC)
##t test of coefficients:
##  Estimate  Std. Error t value Pr(>|t|)
##(Intercept) -21.7703086  41.0326944 -0.5306 0.597124
##bdrms        13.8525219  11.5617901  1.1981 0.234236
##lotsize       0.0020677   0.0071485  0.2893 0.773101
##sqrft         0.1227782   0.0407325  3.0143 0.003406 **
##  ---
##  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

##Run the linear model
model2 <- lm(log(price)~bdrms+log(lotsize)+log(sqrft),data = context1)

##OLS test
coeftest(model2,vcov.=vcov)
##t test of coefficients:
##               Estimate Std. Error t value  Pr(>|t|)
##(Intercept)  -1.297042   0.651284 -1.9915   0.04967 *
##bdrms         0.036958   0.027531  1.3424   0.18308
##log(lotsize)  0.167967   0.038281  4.3877 3.307e-05 ***
##log(sqrft)    0.700232   0.092865  7.5403 5.006e-11 ***
##  ---
##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##White-corrected test
coeftest(model2,vcov.=vcovHC)
##t test of coefficients:
##Estimate Std. Error t value  Pr(>|t|)
##(Intercept)  -1.297042   0.850457 -1.5251  0.130988
##bdrms         0.036958   0.035576  1.0389  0.301845
##log(lotsize)  0.167967   0.053275  3.1528  0.002243 **
##log(sqrft)    0.700232   0.121392  5.7683 1.298e-07 ***
##  ---
##  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

##Interpretations
##a. lotsize and sqrft are significant
##b. sqrft is still significant
```

```
##c. ln(lotsize) and ln(sqrft) are significant
##d. both ln(lotsize) and ln(sqrft) are still significant
##e. it may reduce heteroskedasticity and reduce the over-rejection of t statistics


rm(list=ls()) #drop all variables

##Question2

## Data import
#####################################
##import commands
context2 <- fread('beveridge.csv')

##Run the linear model
model3 <- lm(urate~vrate,data = context2)

##OLS test
coeftest(model3,vcov.=vcov)
##t test of coefficients:
##  Estimate Std. Error t value  Pr(>|t|)
##(Intercept) 17.11942    0.59200  28.918 < 2.2e-16 ***
##  vrate       -3.74145    0.20681 -18.091 < 2.2e-16 ***
##  ---
##  Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##Newey-West-corrected significance test
coeftest(model3,vcov=NeweyWest(model3,lag=5))
##t test of coefficients:
##Estimate Std. Error t value  Pr(>|t|)
##(Intercept) 17.11942    1.36561  12.536 < 2.2e-16 ***
##  vrate       -3.74145    0.39575  -9.454 < 2.2e-16 ***
##  ---
##  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

##KPSS
kpss.test(context2$urate,null="Level")
##KPSS Level = 2.6835, Truncation lag parameter = 2, p-value = 0.01
kpss.test(context2$urate,null="Trend")
##KPSS Trend = 0.74085, Truncation lag parameter = 2, p-value = 0.01
kpss.test(context2$vrate,null="Level")
##KPSS Level = 1.2735, Truncation lag parameter = 2, p-value = 0.01
kpss.test(context2$vrate,null="Trend")
##KPSS Trend = 0.43255, Truncation lag parameter = 2, p-value = 0.01

kpss.test(diff(context2$urate),null="Level")
##KPSS Level = 0.25265, Truncation lag parameter = 2, p-value = 0.1
kpss.test(diff(context2$urate),null="Trend")
##KPSS Trend = 0.2695, Truncation lag parameter = 2, p-value = 0.01
kpss.test(diff(context2$vrate),null="Level")
##KPSS Level = 0.30923, Truncation lag parameter = 2, p-value = 0.1
kpss.test(diff(context2$vrate),null="Trend")
##KPSS Trend = 0.20946, Truncation lag parameter = 2, p-value =0.01245

kpss.test(diff(diff(context2$urate)),null="Level")
##KPSS Level = 0.023309, Truncation lag parameter = 2, p-value = 0.1
kpss.test(diff(diff(context2$urate)),null="Trend")
##KPSS Trend = 0.019768, Truncation lag parameter = 2, p-value = 0.1
kpss.test(diff(diff(context2$vrate)),null="Level")
##KPSS Level = 0.018448, Truncation lag parameter = 2, p-value = 0.1
kpss.test(diff(diff(context2$vrate)),null="Trend")
##KPSS Trend = 0.017123, Truncation lag parameter = 2, p-value = 0.1

##Run the linear model
model4 <- lm(diff(urate)~diff(vrate),data = context2)

##OLS test
coeftest(model4,vcov.=vcov)
##t test of coefficients:
##  Estimate Std. Error t value Pr(>|t|)
##(Intercept)  0.037046   0.017809  2.0802  0.03944 *
##  diff(vrate) -0.027599   0.107318 -0.2572  0.79745
##---
##  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

##Newey-West-corrected significance test
coeftest(model4,vcov=NeweyWest(model4,lag=5))
##             Estimate Std. Error t value Pr(>|t|)
##(Intercept)  0.037046   0.030041  1.2332   0.2197
##diff(vrate) -0.027599   0.081122 -0.3402   0.7342

##Interpretations
##f. the coefficient on the vanancy rate is significant
##g. △(△urate)is level stationary
##h. △(△vrate)is level stationary
```

```
##i. from 3 to 4, vrate become not significant
##j. model3 describes the data better

rm(list=ls()) #drop all variables

##Question3

## Data import
#####################################
##import commands
context3 <- fread('JTRAIN.csv')
context3 <- plm.data(context3,index=c("fcode","year"))

##creat new variables
context3$d88=ifelse(context3$year==1988,1,0)
context3$d89=ifelse(context3$year==1989,1,0)

for(i in 1:nrow(context3))
{if(i%%3==1){context3$grant_1[i] <- 0}else{
  context3$grant_1[i] <- context3$grant[i-1]}
}

##Run the linear model
model5 <- plm(log(scrap)~d88+d89+grant+grant_1,model="pooling",data = context3)
model6 <- plm(log(scrap)~d88+d89+grant+grant_1,model="within",data = context3)

##OLS test
coeftest(model5,vcov.=vcov)
##t test of coefficients:
##  Estimate Std. Error t value Pr(>|t|)
##(Intercept)  0.597434   0.203063  2.9421 0.003754 **
##  d88         -0.239370   0.310864 -0.7700 0.442447
##d89         -0.496524   0.337928 -1.4693 0.143748
##grant        0.200020   0.338285  0.5913 0.555186
##grant_1      0.048936   0.436066  0.1122 0.910792
##---
##  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

coeftest(model6,vcov.=vcov)
##t test of coefficients:
##Estimate Std. Error t value Pr(>|t|)
##d88     -0.080216   0.109475 -0.7327  0.46537
##d89     -0.247203   0.133218 -1.8556  0.06634 .
##grant   -0.252315   0.150629 -1.6751  0.09692 .
##grant_1 -0.421590   0.210200 -2.0057  0.04749 *
##  ---
##  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

##HAC-corrected significance test
coeftest(model6,vcov=vcovHC(model6, method = "arellano"))
##t test of coefficients:
##  Estimate Std. Error t value Pr(>|t|)
##d88     -0.080216   0.095719 -0.8380  0.40393
##d89     -0.247203   0.192514 -1.2841  0.20197
##grant   -0.252315   0.140329 -1.7980  0.07507 .
##grant_1 -0.421590   0.276335 -1.5256  0.13013
##---
##  Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

##interpretation
##k. the coefficient on grant in model5 is 0.20,
##that means if the firm have a grant this year, it will have 0.2 more of scrap rates than the year without a grant
##l. the coefficient on grant last year in model5 is 0.04
##that means if the firm had a grant last year, it will have 0.04 more of scrap rates than the last year without a
grant
##m.the signs are both positive which means if the grant is received this year or last year,the scrap rate for a firm
will increase.
##n. the coefficient on grant in model6 is -0.24
##that means if the firm have a grant this year, it will have 0.24 less of scrap rates than the year without a grant
##o. the coefficient on grant last year in model6 is -0.42
##that means if the firm had a grant last year, it will have 0.42 less of scrap rates than the last year without a
grant
##p. the signs are both negative, which means if the grant is received this year or last year,the scrap rate for a
firm will decrease.
##q. by using OLS, the variable grant got last year is significant, while it is not significant by using HAC
significance test
```