Panel Data

Jayashree Raman

9/18/2018

library(plm)

## Loading required package: Formula

library(prediction)  
library(Metrics)  
library(tseries)

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

train\_data <- na.omit(read.csv(file="usersessions-with-char-sec-train.csv", header=TRUE, row.names = NULL, sep="|"))  
  
panel.data.train <- plm.data(train\_data, index = c("session\_start","userid"))

## Warning: use of 'plm.data' is discouraged, better use 'pdata.frame' instead

mdl\_pooled <-plm(session\_length~age+session\_length\_mvavg, data = panel.data.train, model = "pooling")  
  
mdl\_random <-plm(session\_length~age+session\_length\_mvavg, data = panel.data.train, model = "random")  
  
mdl\_fe <-plm(session\_length~age+session\_length\_mvavg, data = panel.data.train, model = "within")  
  
##Summaries  
  
#summary(mdl\_fe)  
  
#summary(mdl\_random)  
  
#summary(mdl\_pooled)  
  
##Hausman test - To decide between the fixed and random effects model. Null hypothesis is that Random effects is a better fit  
  
phtest(mdl\_fe, mdl\_random)

##   
## Hausman Test  
##   
## data: session\_length ~ age + session\_length\_mvavg  
## chisq = 2.5693, df = 2, p-value = 0.2767  
## alternative hypothesis: one model is inconsistent

# Breusch-Pagan test - Testing between random effects regression and a fixed effect regression - Null is than random effects is better.   
plmtest(mdl\_fe, c("session\_start"), effect = c("twoways"), type=("bp"))

##   
## Lagrange Multiplier Test - two-ways effects (Breusch-Pagan) for  
## unbalanced panels  
##   
## data: session\_length ~ age + session\_length\_mvavg  
## chisq = 46468, df = 2, p-value < 2.2e-16  
## alternative hypothesis: significant effects

# Breusch-Pagan test - Testing between random effects regression and a simple OLS regression - Null is than random effects is better  
plmtest(mdl\_pooled, effect = c("twoways"), type=("bp"))

##   
## Lagrange Multiplier Test - two-ways effects (Breusch-Pagan) for  
## unbalanced panels  
##   
## data: session\_length ~ age + session\_length\_mvavg  
## chisq = 46468, df = 2, p-value < 2.2e-16  
## alternative hypothesis: significant effects

# Breusch--Godfrey Test For Panel Models - Test of serial correlation for (the idiosyncratic component of) the errors in panel models.  
pbgtest(mdl\_fe)

##   
## Breusch-Godfrey/Wooldridge test for serial correlation in panel  
## models  
##   
## data: session\_length ~ age + session\_length\_mvavg  
## chisq = 55542, df = 1, p-value < 2.2e-16  
## alternative hypothesis: serial correlation in idiosyncratic errors

# F Test For Individual And/Or Time Effects- Test of individual and/or time effects based on the comparison of the within and the pooling model  
  
pFtest(mdl\_fe, mdl\_pooled)

##   
## F test for individual effects  
##   
## data: session\_length ~ age + session\_length\_mvavg  
## F = 1.3334, df1 = 221240, df2 = 250, p-value = 0.00119  
## alternative hypothesis: significant effects

# The Dickey-Fuller test to check for stochastic trends. The null hypothesis is that the series has a unit root (i.e. non-stationary). If unit root is present you can take the first difference of the variable.  
  
adf.test(panel.data.train$session\_length, k=3)

## Warning in adf.test(panel.data.train$session\_length, k = 3): p-value  
## smaller than printed p-value

##   
## Augmented Dickey-Fuller Test  
##   
## data: panel.data.train$session\_length  
## Dickey-Fuller = -234.1, Lag order = 3, p-value = 0.01  
## alternative hypothesis: stationary