

Panel Data

Jayashree Raman

9/18/2018

```
library(plm)

## Warning: package 'plm' was built under R version 3.5.1

## Loading required package: Formula

library(prediction)

## Warning: package 'prediction' was built under R version 3.5.1

library(Metrics)

## Warning: package 'Metrics' was built under R version 3.5.1

library(tseries)

## Warning: package 'tseries' was built under R version 3.5.1
```

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))

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_fe <- plm(session_length~age+session_length_mvavg, data = panel.data.train, model = "within")

##Summaries

summary(mdl_fe)

## Oneway (individual) effect Within Model
##
## Call:
## plm(formula = session_length ~ age + session_length_mvavg, data = panel.data.train,
```

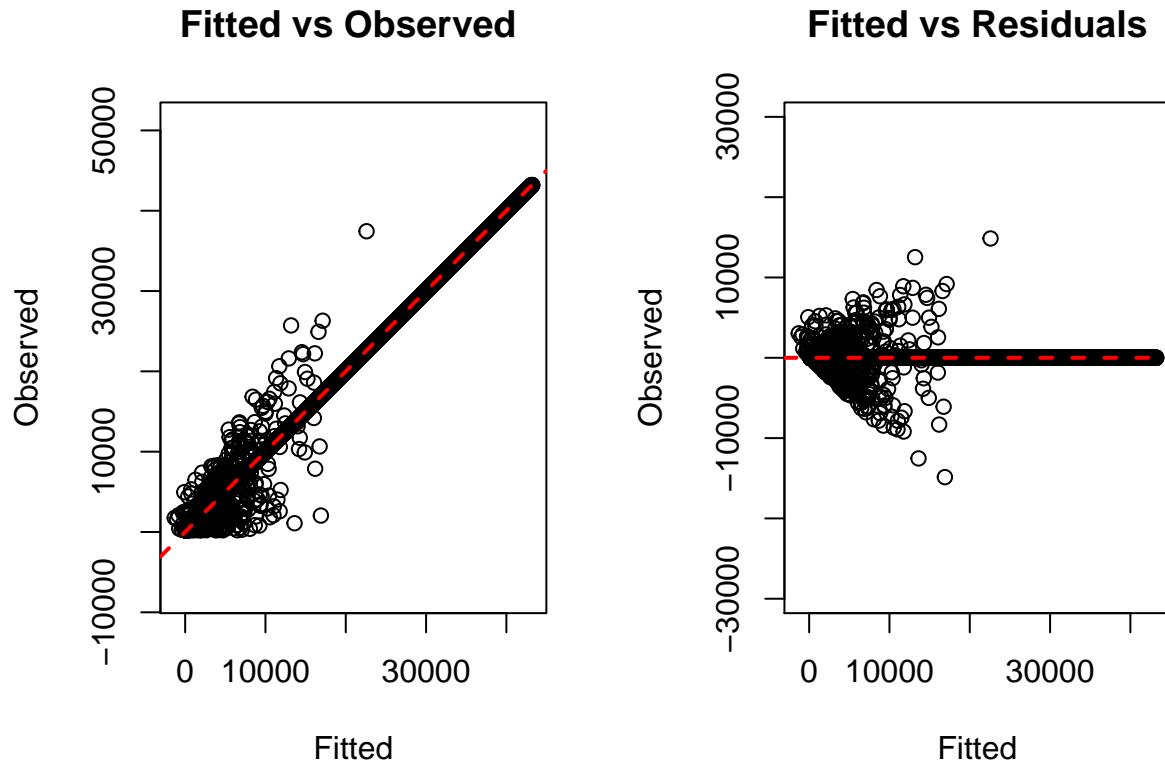
```

##      model = "within")
##
## Unbalanced Panel: n = 221244, T = 1-2, N = 221496
##
## Residuals:
##      Min. 1st Qu. Median 3rd Qu.    Max.
## -14858       0     0     0   14858
##
## Coefficients:
##                               Estimate Std. Error t-value Pr(>|t|)
## age                  -78.79559   56.92435 -1.3842   0.1675
## session_length_mvavg  0.73620   0.12892  5.7107 3.182e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares: 6461100000
## Residual Sum of Squares: 5710500000
## R-Squared: 0.11618
## Adj. R-Squared: -782.04
## F-statistic: 16.4321 on 2 and 250 DF, p-value: 1.9736e-07

#cat(length(panel.data.train$session_length), length(mdl_fd$residuals))
# Fitted vs Observed and Fitted vs Residuals plots
par(mfrow=c(1,2))
plot(panel.data.train$session_length-mdl_fe$residuals, panel.data.train$session_length, asp=1, ylab = "Observed",
abline(0,1, col='red', lty='dashed', lwd=2)

## Fitted vs Residuals plots
plot(panel.data.train$session_length-mdl_fe$residuals,mdl_fe$residuals, asp=1, ylab = "Observed", xlab =
abline(0,0, col='red', lty='dashed', lwd=2)

```



```
## MAE and RMSE

mae_fe = mean(abs(mdl_fe$residuals))
rmse_fe = sqrt(mean(abs(mdl_fe$residuals)^2))

cat('MAE = ', mae_fe, ', RMSE = ', rmse_fe)

## MAE = 5.639383 , RMSE = 160.5657
```