

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-genre-train.csv", header=TRUE, row.names = NULL))

panel.data.train <- plm.data(train_data, index = c("userid", "session_start"))

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

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

##Summaries

summary(mdl_random)

## Oneway (individual) effect Random Effect Model
##      (Swamy-Arora's transformation)
##
## Call:
```

```

## plm(formula = session_length ~ age + session_length_mvavg, data = panel.data.train,
##       model = "random")
##
## Unbalanced Panel: n = 174, T = 19-4312, N = 169556
##
## Effects:
##           var   std.dev share
## idiosyncratic 77321011     8793 0.981
## individual     1459817     1208 0.019
## theta:
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
## 0.1421  0.7582  0.8111  0.7932  0.8496  0.8898
##
## Residuals:
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
## -40554 -3393 -1546      -7    1458  580163
##
## Coefficients:
##                   Estimate Std. Error t-value Pr(>|t|)
## (Intercept) 1639.944911 365.035713 4.4926 7.042e-06 ***
## age          -42.181798 12.832928 -3.2870 0.001013 **
## session_length_mvavg 0.875920  0.013643 64.2042 < 2.2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares: 1.3495e+13
## Residual Sum of Squares: 1.3123e+13
## R-Squared: 0.027574
## Adj. R-Squared: 0.027563
## F-statistic: 2403.81 on 2 and 169553 DF, p-value: < 2.22e-16

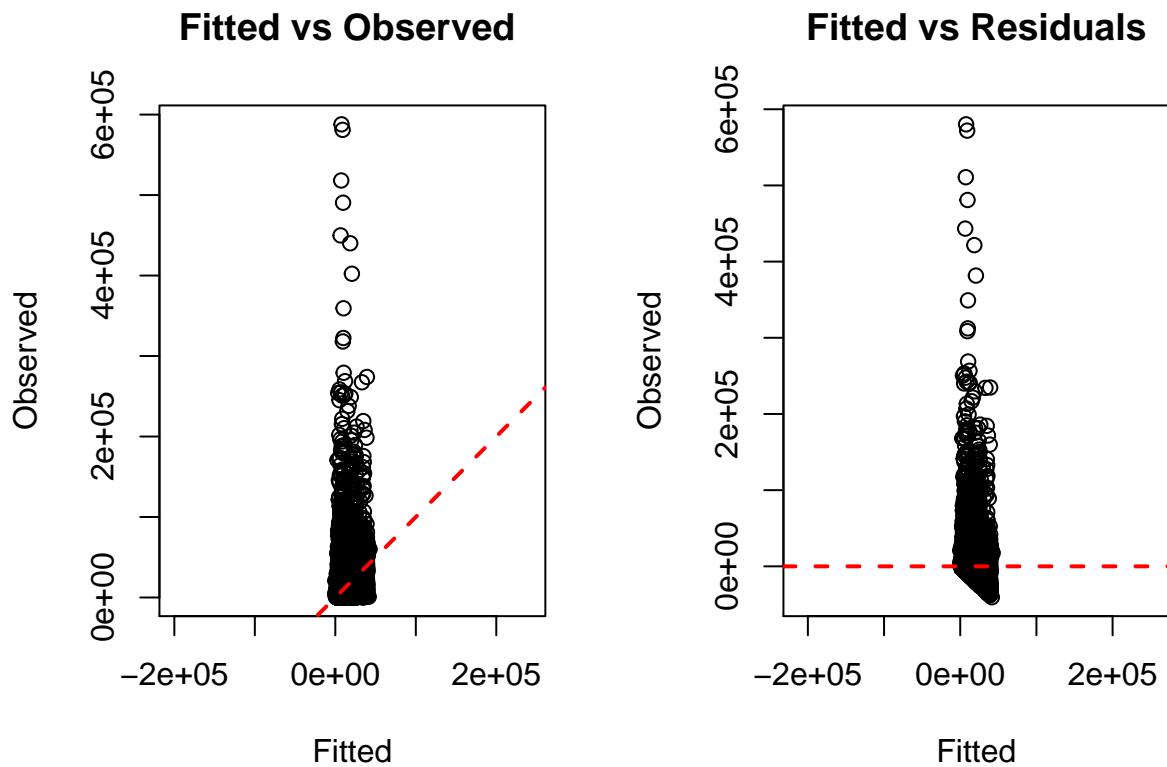
```

```

#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$session_length~mdl_random$residuals, panel.data.train$session_length, asp=1, ylab
abline(0,1, col='red', lty='dashed', lwd=2)

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

```



```
## MAE and RMSE

mae_random = mean(abs(mdl_random$residuals))
rmse_random = sqrt(mean(abs(mdl_random$residuals)^2))

cat('MAE = ', mae_random, ', RMSE = ', rmse_random)

## MAE = 4340.695 , RMSE = 8797.43
```