

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_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 = 221244, T = 1-2, N = 221496
##
## Effects:
##           var  std.dev share
## idiosyncratic 22841851      4779  0.75
## individual     7607686      2758  0.25
## theta:
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
## 0.1339 0.1339 0.1339 0.1341 0.1339 0.2253
##
## Residuals:
##   Min. 1st Qu. Median   Mean 3rd Qu.   Max.
## -26772 -2679 -1165      0    1330  35333
##
## Coefficients:
##                               Estimate Std. Error t-value Pr(>|t|)
## (Intercept)            7.8026e+02 5.4127e+01 14.4153 <2e-16 ***
## age                  1.1214e+00 1.9318e+00 0.5805  0.5616
## session_length_mvavg 8.0630e-01 4.4583e-03 180.8542 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:  5.8205e+12
## Residual Sum of Squares: 5.0599e+12
## R-Squared:          0.13068
## Adj. R-Squared:        0.13067
## F-statistic: 16647.8 on 2 and 221493 DF, p-value: < 2.22e-16

```

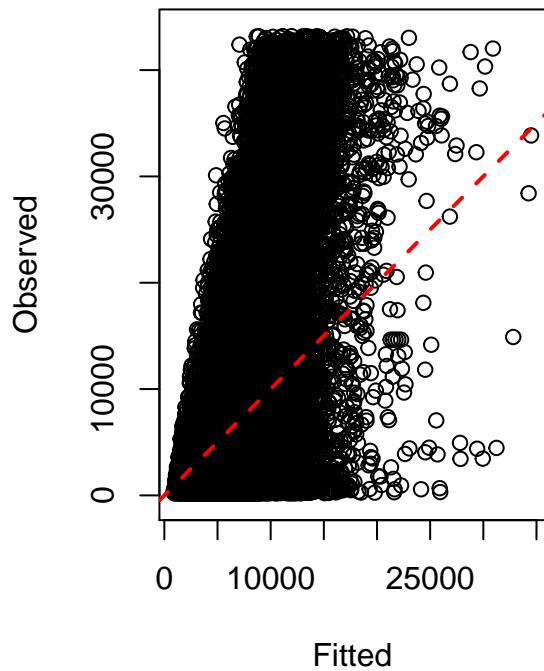
```

#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)

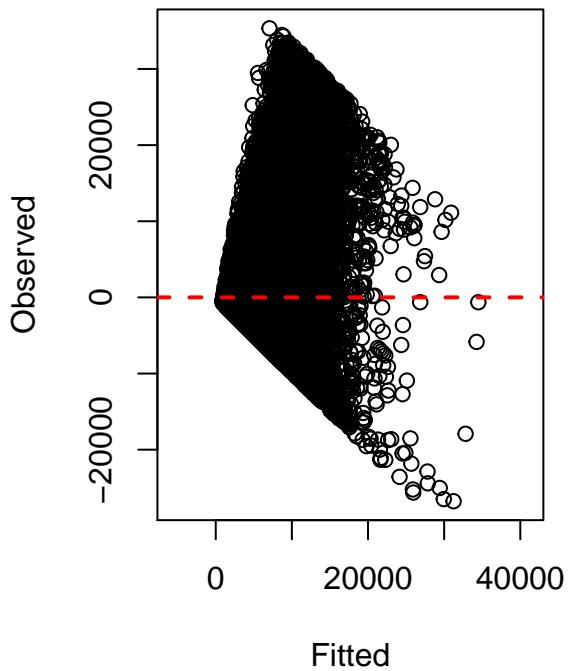
## Fitted vs Residuals plots
plot(panel.data.train$session_length~mdl_random$residuals,mdl_random$residuals, asp=1, ylab = "Observed"
abline(0,0, col='red', lty='dashed', lwd=2)

```

Fitted vs Observed



Fitted vs Residuals



```
## 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 = 3214.21 , RMSE = 4779.546
```