

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

##Summaries

summary(mdl_pooled)

## Pooling Model
##
## Call:
## plm(formula = session_length ~ age + session_length_mvavg, data = panel.data.train,
```

```

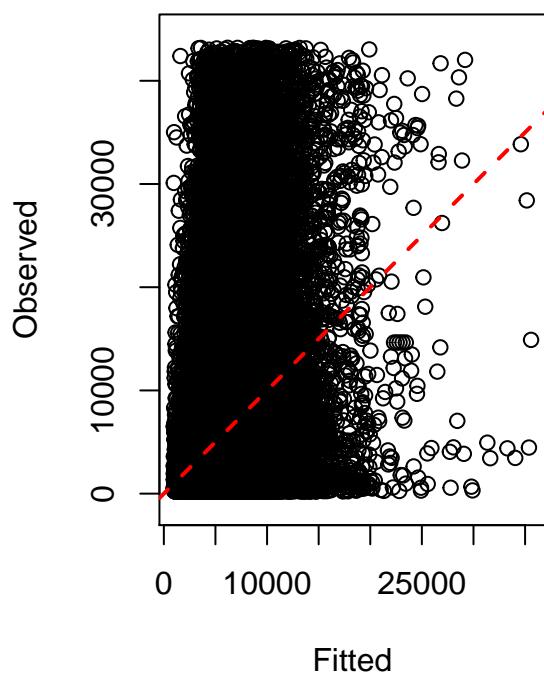
##      model = "pooling")
##
## Unbalanced Panel: n = 221244, T = 1-2, N = 221496
##
## Residuals:
##      Min.    1st Qu.     Median    3rd Qu.     Max.
## -30909.1   -3092.5   -1345.2    1535.7   40794.8
##
## Coefficients:
##                               Estimate Std. Error t-value Pr(>|t|)
## (Intercept)            7.7994e+02 5.4125e+01 14.4100 <2e-16 ***
## age                    1.1362e+00 1.9318e+00  0.5881  0.5564
## session_length_mvavg  8.0628e-01 4.4583e-03 180.8501 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Total Sum of Squares:  7.7579e+12
## Residual Sum of Squares: 6.7444e+12
## R-Squared: 0.13065
## Adj. R-Squared: 0.13064
## F-statistic: 16643.2 on 2 and 221493 DF, p-value: < 2.22e-16

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

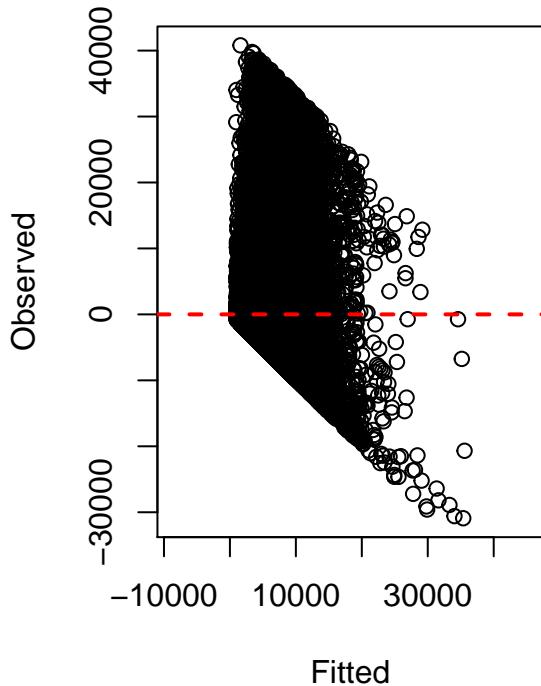
## Fitted vs Residuals plots
plot(panel.data.train$session_length-mdl_pooled$residuals,mdl_pooled$residuals, asp=1, ylab = "Residuals")
abline(0,0, col='red', lty='dashed', lwd=2)

```

Fitted vs Observed



Fitted vs Residuals



```
## MAE and RMSE

mae_pooled = mean(abs(mdl_pooled$residuals))
rmse_pooled = sqrt(mean(abs(mdl_pooled$residuals)^2))

cat('MAE = ', mae_pooled, ', RMSE = ', rmse_pooled)

## MAE = 3710.765 , RMSE = 5518.078
```