NBA\_Fixed\_Effects.R

utsav

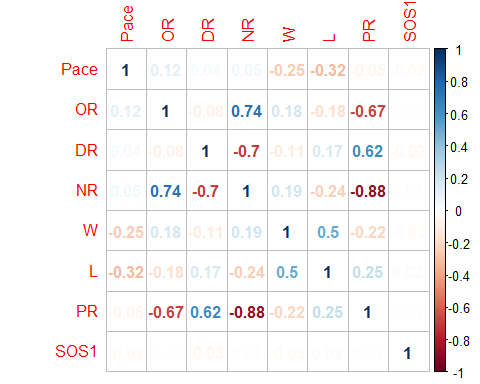
2020-05-05

setwd("/Users/Utsav/Desktop/SDM")  
  
library(readxl)  
# Reading file in a dataframe  
nba\_f <- read\_excel("Power\_Ranking\_Final\_Data.xlsx")  
attach(nba\_f)  
  
# Making week as a factor  
# nba\_f$Week <- as.factor(nba\_f$Week)  
  
# Checking correlations of variables   
m <- cor(cbind(nba\_f[, 2:9]))  
# Correlation Matrix  
library(corrplot)

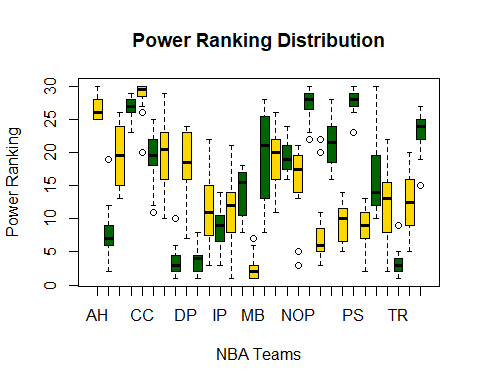
## Warning: package 'corrplot' was built under R version 3.6.3

## corrplot 0.84 loaded

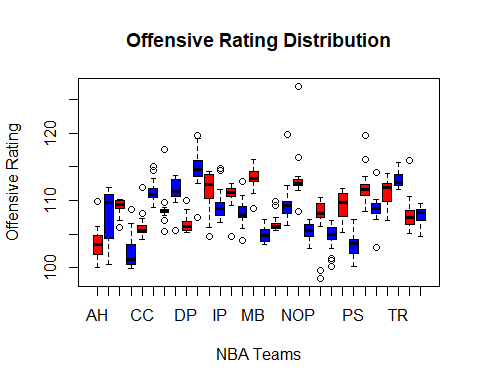
corrplot(m, method="number")



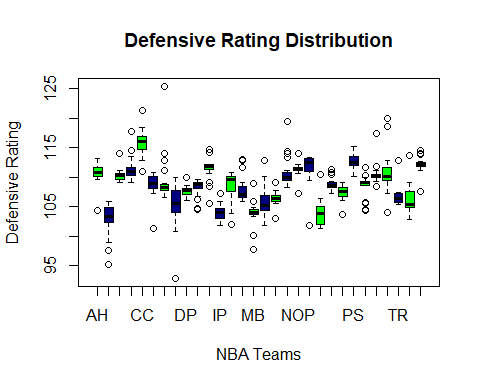
# Plotting boxplot  
  
boxplot(PR~Team,data=nba\_f, main="Power Ranking Distribution",col=(c("gold","darkgreen")),  
 xlab="NBA Teams", ylab="Power Ranking")



boxplot(OR~Team,data=nba\_f, main="Offensive Rating Distribution",col=(c("red","blue")),  
 xlab="NBA Teams", ylab="Offensive Rating")



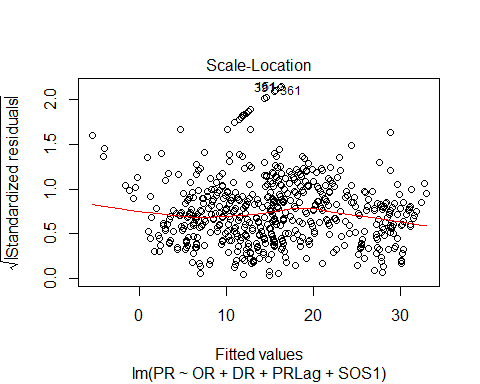
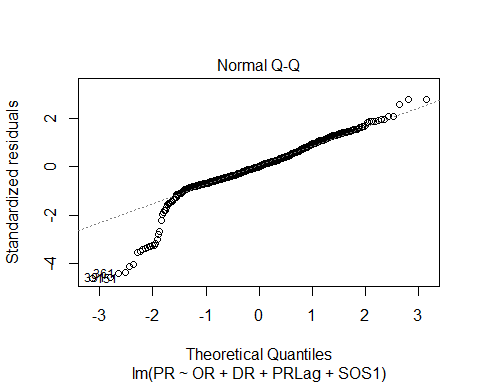
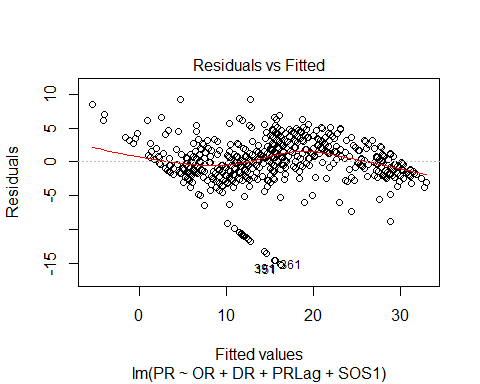
boxplot(DR~Team,data=nba\_f, main="Defensive Rating Distribution",col=(c("green","navyblue")),  
 xlab="NBA Teams", ylab="Defensive Rating")



# Creating a win% column  
nba\_f$winp= round((nba\_f$W/(nba\_f$W + nba\_f$L))\*100,1)  
  
# Creating sequence to make lags  
n <- nrow(nba\_f)  
  
nba\_f$PRLag <- c(NA, nba\_f$PR[1:n-1])  
nba\_f$SOS1Lag <- c(NA,nba\_f$SOS1[1:n-1])  
   
# Linear model  
ols\_base <- lm(PR ~ OR+DR+PRLag+SOS1, data = nba\_f)  
summary(ols\_base)

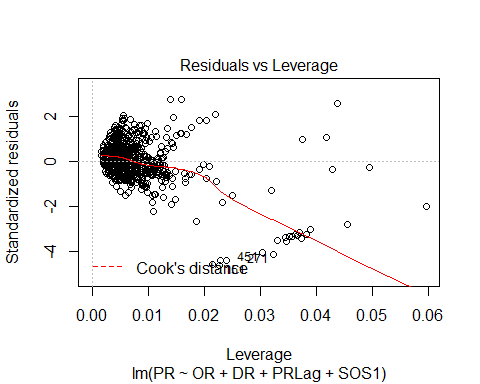
##   
## Call:  
## lm(formula = PR ~ OR + DR + PRLag + SOS1, data = nba\_f)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -15.3084 -1.5956 0.0076 1.9515 9.2238   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 18.65225 5.88004 3.172 0.00159 \*\*   
## OR -1.01052 0.04386 -23.037 < 2e-16 \*\*\*  
## DR 0.92614 0.04560 20.310 < 2e-16 \*\*\*  
## PRLag 0.37594 0.02264 16.605 < 2e-16 \*\*\*  
## SOS1 0.01310 0.01605 0.816 0.41456   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 3.352 on 594 degrees of freedom  
## (1 observation deleted due to missingness)  
## Multiple R-squared: 0.8508, Adjusted R-squared: 0.8498   
## F-statistic: 846.6 on 4 and 594 DF, p-value: < 2.2e-16

plot(ols\_base)



# The residual plot shows that the model is violating linearity and   
# independence assumptions. The residual plot is also heteroskedastic.  
  
  
library(plm)

## Warning: package 'plm' was built under R version 3.6.3



# Panel Data Analysis  
# Fixed Effect Model With Lag  
fixed\_1 <- plm(PR ~ OR + DR + PRLag + SOS1, data=nba\_f,   
 index=c("Team","Week"), model="within")  
summary(fixed\_1)

## Oneway (individual) effect Within Model  
##   
## Call:  
## plm(formula = PR ~ OR + DR + PRLag + SOS1, data = nba\_f, model = "within",   
## index = c("Team", "Week"))  
##   
## Unbalanced Panel: n = 30, T = 19-20, N = 599  
##   
## Residuals:  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -12.952113 -1.434572 0.095455 1.628025 7.556587   
##   
## Coefficients:  
## Estimate Std. Error t-value Pr(>|t|)   
## OR -0.7645397 0.0560805 -13.6329 <2e-16 \*\*\*  
## DR 0.6986344 0.0559342 12.4903 <2e-16 \*\*\*  
## PRLag 0.2658874 0.0210328 12.6416 <2e-16 \*\*\*  
## SOS1 0.0067195 0.0128702 0.5221 0.6018   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 8361.8  
## Residual Sum of Squares: 3915  
## R-Squared: 0.53181  
## Adj. R-Squared: 0.50446  
## F-statistic: 160.442 on 4 and 565 DF, p-value: < 2.22e-16

df\_nba<- fixef(fixed\_1)  
  
# Checking assumptions for auto correlation and heteroskedasticity  
  
library(lmtest)

## Loading required package: zoo

##   
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':  
##   
## as.Date, as.Date.numeric

pdwtest(fixed\_1)

##   
## Durbin-Watson test for serial correlation in panel models  
##   
## data: PR ~ OR + DR + PRLag + SOS1  
## DW = 1.1665, p-value < 2.2e-16  
## alternative hypothesis: serial correlation in idiosyncratic errors

# Checking for   
pbgtest(fixed\_1)

##   
## Breusch-Godfrey/Wooldridge test for serial correlation in panel  
## models  
##   
## data: PR ~ OR + DR + PRLag + SOS1  
## chisq = 136.7, df = 19, p-value < 2.2e-16  
## alternative hypothesis: serial correlation in idiosyncratic errors

# The test shows that our model has auto correlation and heteroskedasticity.  
# We will have to correct for these in order to do a better prediction.  
  
  
# Random Effect Model with Lag and SOS  
  
random\_1 <- plm(PR ~ OR + DR + PRLag + SOS1, data=nba\_f,   
 index=c("Team"), model="random")  
summary(random\_1)

## Oneway (individual) effect Random Effect Model   
## (Swamy-Arora's transformation)  
##   
## Call:  
## plm(formula = PR ~ OR + DR + PRLag + SOS1, data = nba\_f, model = "random",   
## index = c("Team"))  
##   
## Unbalanced Panel: n = 30, T = 19-20, N = 599  
##   
## Effects:  
## var std.dev share  
## idiosyncratic 6.929 2.632 0.688  
## individual 3.138 1.772 0.312  
## theta:  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 0.6773 0.6847 0.6847 0.6845 0.6847 0.6847   
##   
## Residuals:  
## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## -13.6070 -1.3189 0.2220 0.0008 1.6264 6.9068   
##   
## Coefficients:  
## Estimate Std. Error z-value Pr(>|z|)   
## (Intercept) 19.0954255 6.7160935 2.8432 0.004466 \*\*   
## OR -0.8734441 0.0520911 -16.7676 < 2.2e-16 \*\*\*  
## DR 0.7982889 0.0525704 15.1851 < 2.2e-16 \*\*\*  
## PRLag 0.2891779 0.0211449 13.6760 < 2.2e-16 \*\*\*  
## SOS1 0.0062245 0.0132150 0.4710 0.637628   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Total Sum of Squares: 11974  
## Residual Sum of Squares: 4360.4  
## R-Squared: 0.63584  
## Adj. R-Squared: 0.63339  
## Chisq: 1037.16 on 4 DF, p-value: < 2.22e-16

df\_nba\_rn<- ranef(random\_1)  
  
View(df\_nba\_rn)  
library(stargazer)

##   
## Please cite as:

## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.

## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

stargazer(ols\_base,fixed\_1,random\_1,type = "text")

##   
## ==================================================================================  
## Dependent variable:   
## --------------------------------------------------------------  
## PR   
## OLS panel   
## linear   
## (1) (2) (3)   
## ----------------------------------------------------------------------------------  
## OR -1.011\*\*\* -0.765\*\*\* -0.873\*\*\*   
## (0.044) (0.056) (0.052)   
##   
## DR 0.926\*\*\* 0.699\*\*\* 0.798\*\*\*   
## (0.046) (0.056) (0.053)   
##   
## PRLag 0.376\*\*\* 0.266\*\*\* 0.289\*\*\*   
## (0.023) (0.021) (0.021)   
##   
## SOS1 0.013 0.007 0.006   
## (0.016) (0.013) (0.013)   
##   
## Constant 18.652\*\*\* 19.095\*\*\*   
## (5.880) (6.716)   
##   
## ----------------------------------------------------------------------------------  
## Observations 599 599 599   
## R2 0.851 0.532 0.636   
## Adjusted R2 0.850 0.504 0.633   
## Residual Std. Error 3.352 (df = 594)   
## F Statistic 846.621\*\*\* (df = 4; 594) 160.442\*\*\* (df = 4; 565) 1,037.159\*\*\*  
## ==================================================================================  
## Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

# Hausman Test  
phtest(fixed\_1, random\_1) # Null hypothesis is random effects model is consistent

##   
## Hausman Test  
##   
## data: PR ~ OR + DR + PRLag + SOS1  
## chisq = 28.689, df = 4, p-value = 9.04e-06  
## alternative hypothesis: one model is inconsistent

# but since p-value is less than 0.05, we can reject the null hypothesis  
# and say that fixed effects model is better.