What determines player salary and team valuation?: an emprical result from 2016-2017 NBA season

Renxiong Liu and Lingfei Zhao

April 16, 2018

Overview

Background

If you are the manager of a National Basketball Association (NBA) basketball team,

1. Whats the reasonable salary to sign or re-sign players?
2. What can you do if you want to increase your teams value?

Background

If you are the manager of a National Basketball Association (NBA) basketball team,

1. Whats the reasonable salary to sign or re-sign players?
2. What can you do if you want to increase your teams value?

Answer:Regression analysis!

Dataset: Source

- Dataset on player performances, social power and team valuation in 2016-2017 NBA season
  - **Player performance**: basic performance statistics(e.g. points, assists) and advanced performance statistics(e.g. player impact estimate)
  - **Player social power**: Wikipedia page views, Twitter favorite counts, Twitter retweet counts
  - **Team valuation**: Forbes NBA team value 2017
- Regional GDP data in 2016 and 2017 for cities with NBA teams.
- Average tickets price for each team in 2016-2017 NBA season

Dataset: Exploratory Data Analysis

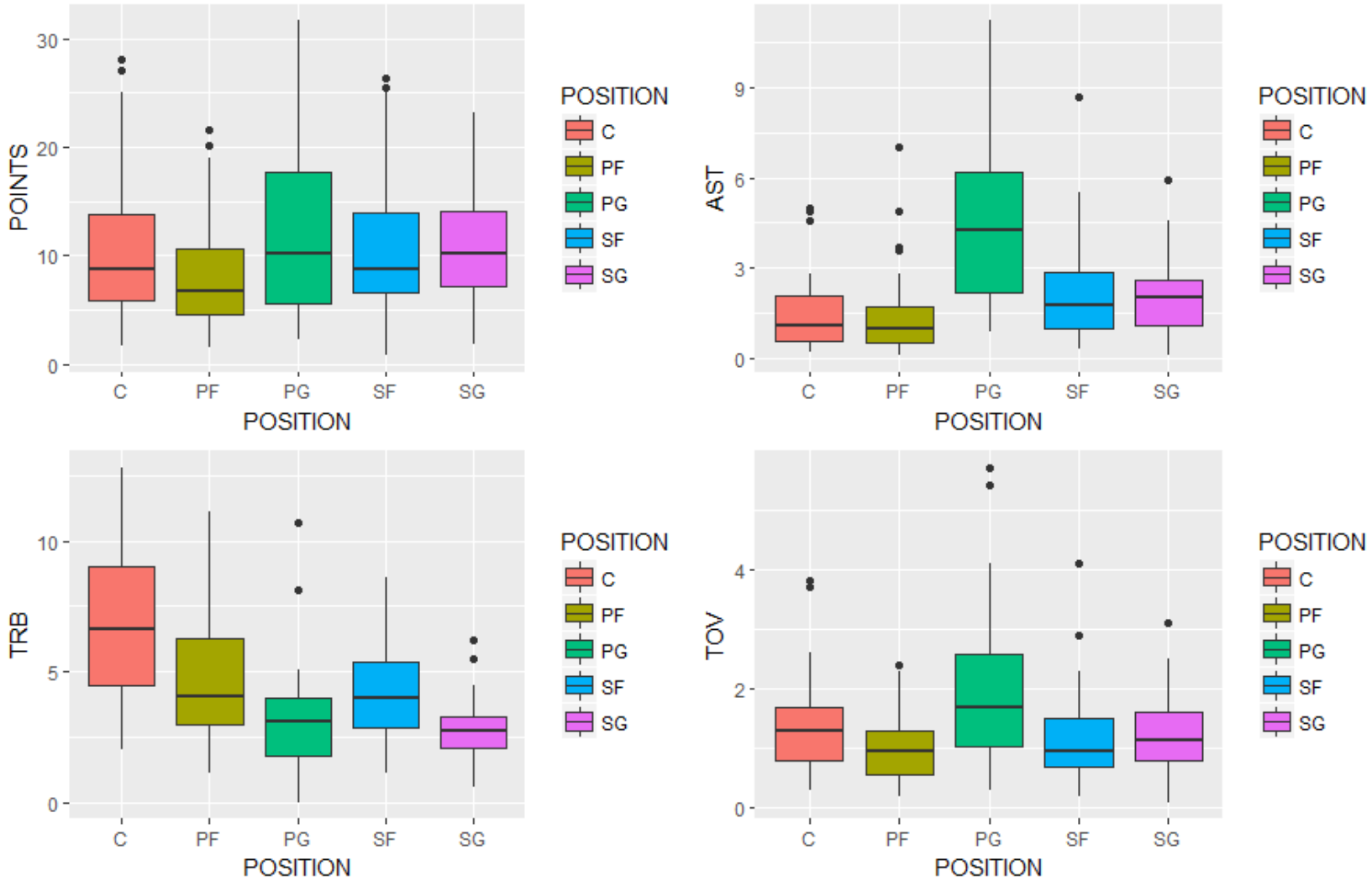


Figure: Box plot for basic performance statistics in different positions

Dataset: Exploratory Data Analysis

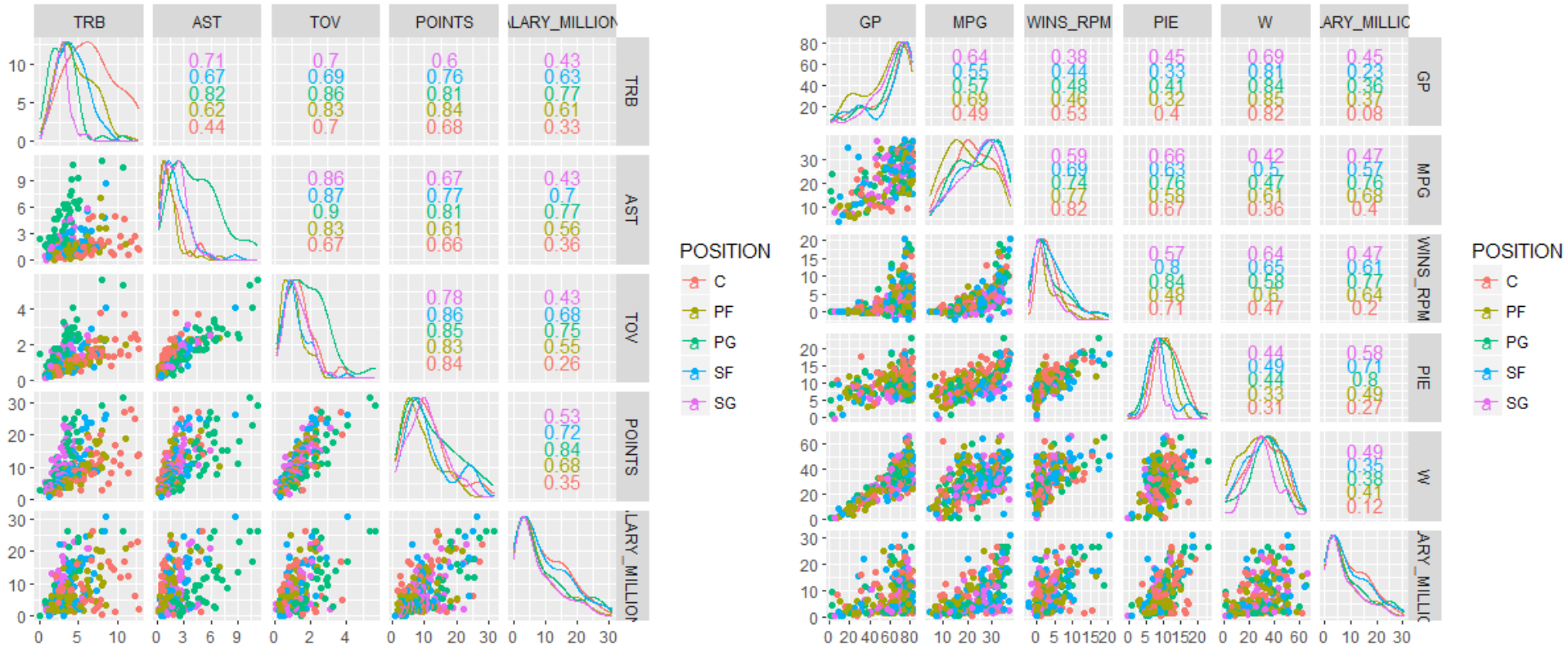


Figure: Scatterplot matrix for basic/advanced performance statistics

Dataset: Exploratory Data Analysis

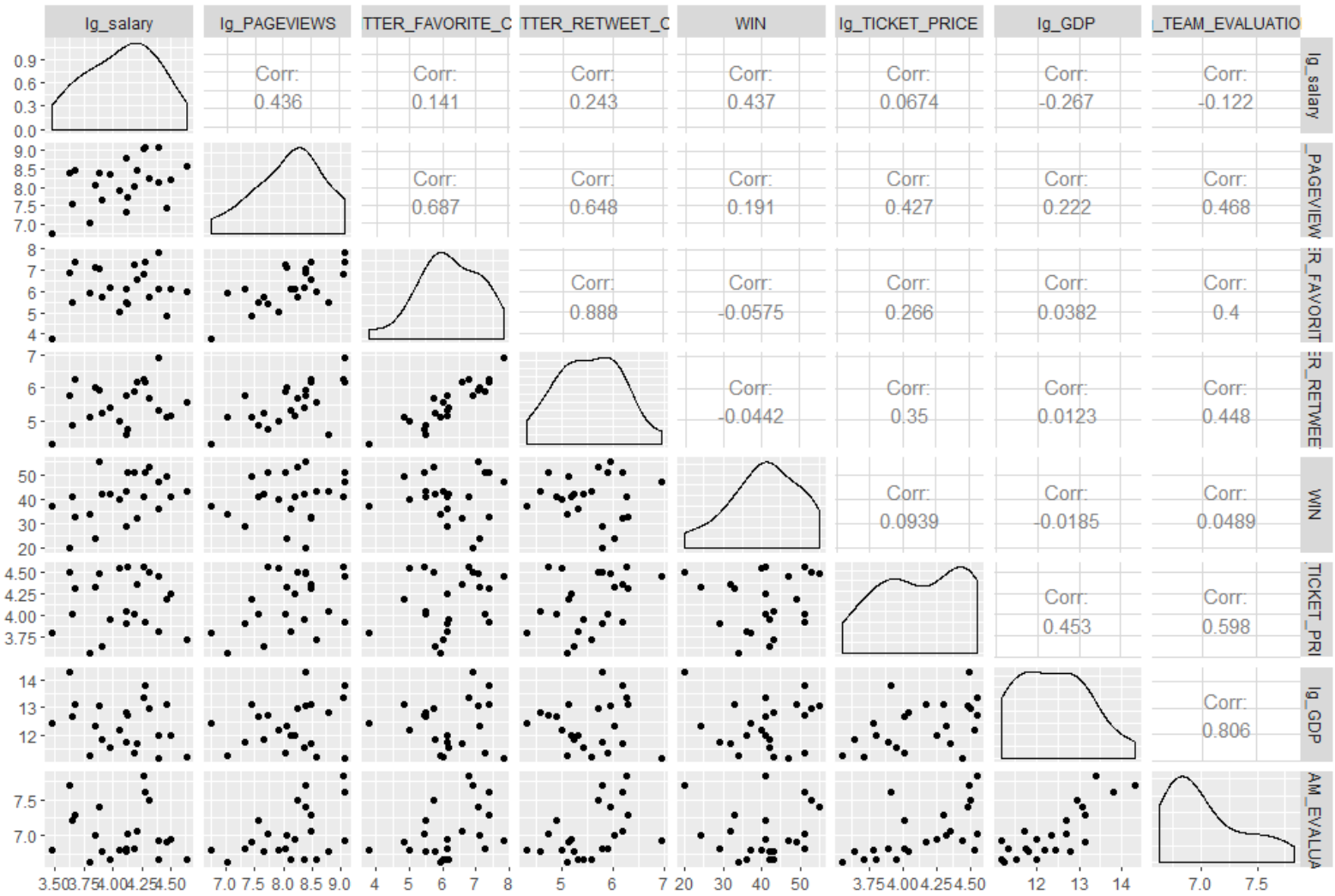


Figure: Scatterplot matrix of team variables

Method: Team Valuation

- Potential variables include basic performance statistics, team salaries, ticket price, games winning and local GDP
- A backward selection method with BIC criterion is employed to select "best" linear regression model
- Model assumptions are checked via regular tool, e.g., qq plot, (standardized/studentized) residual plot

Results: Team Valuation

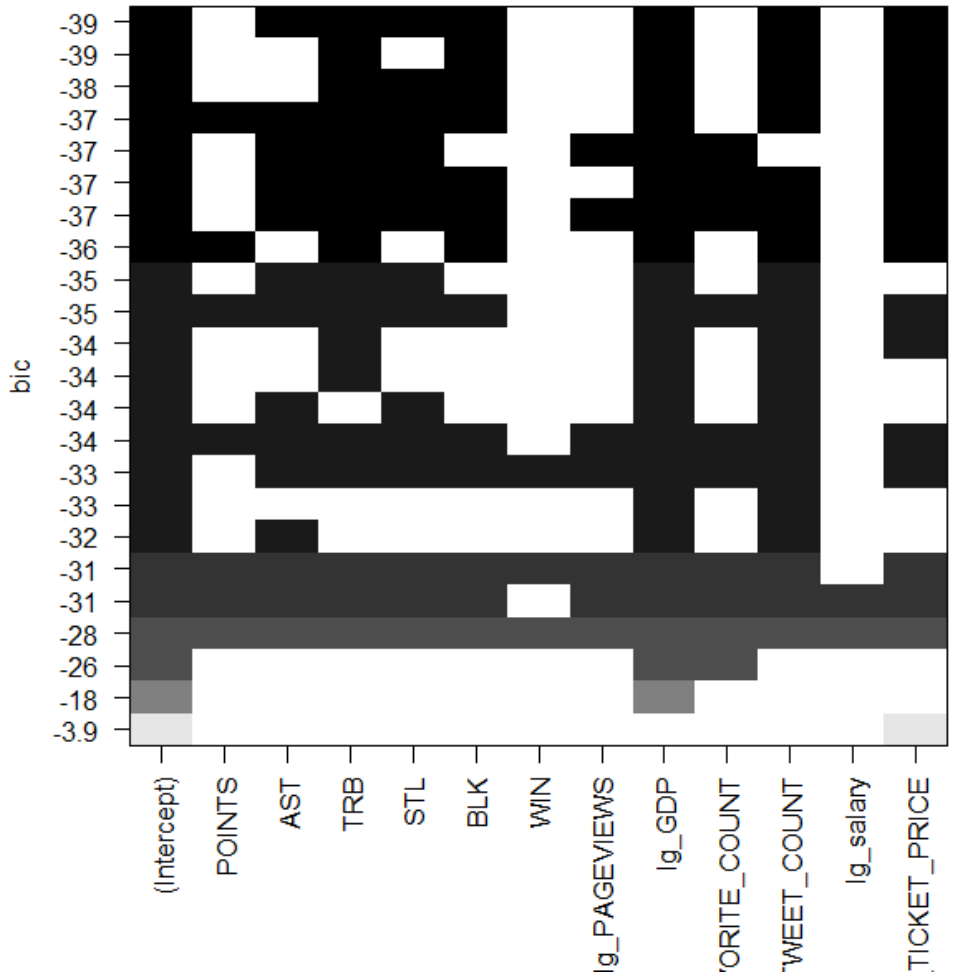
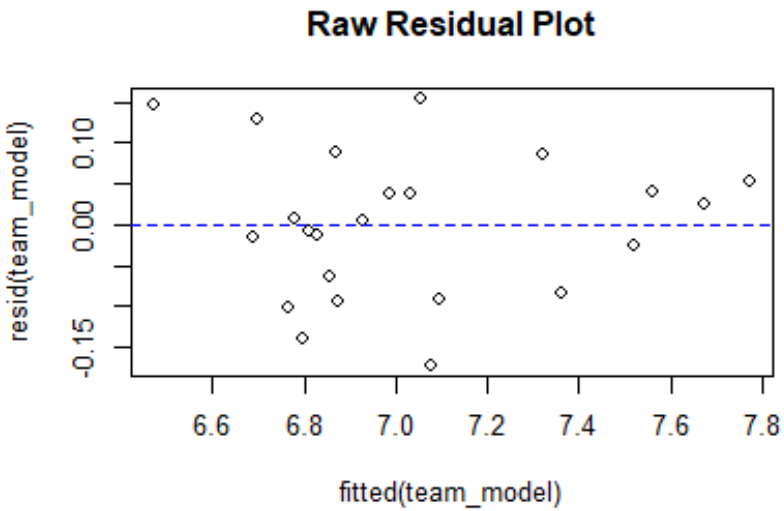


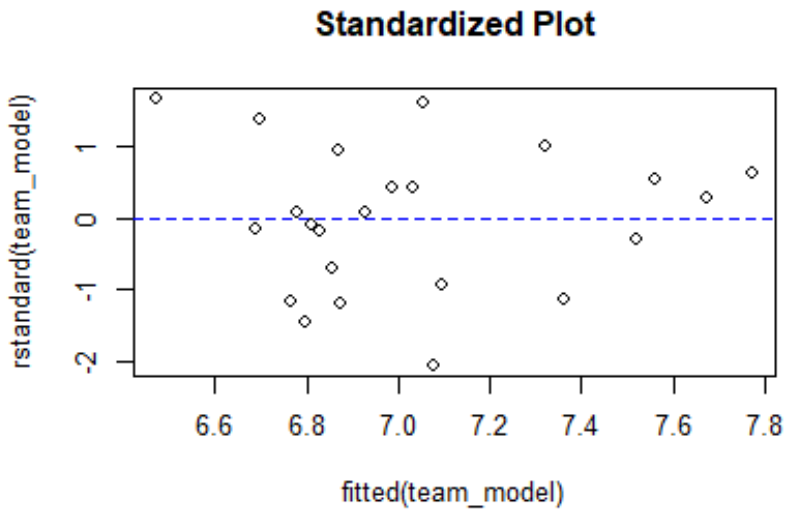
Figure: Variable selection plot with BIC criterion

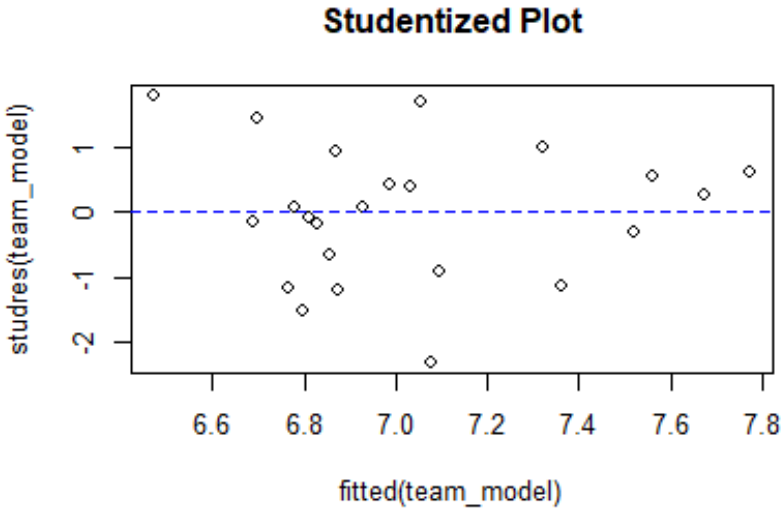
Results: Team Valuation

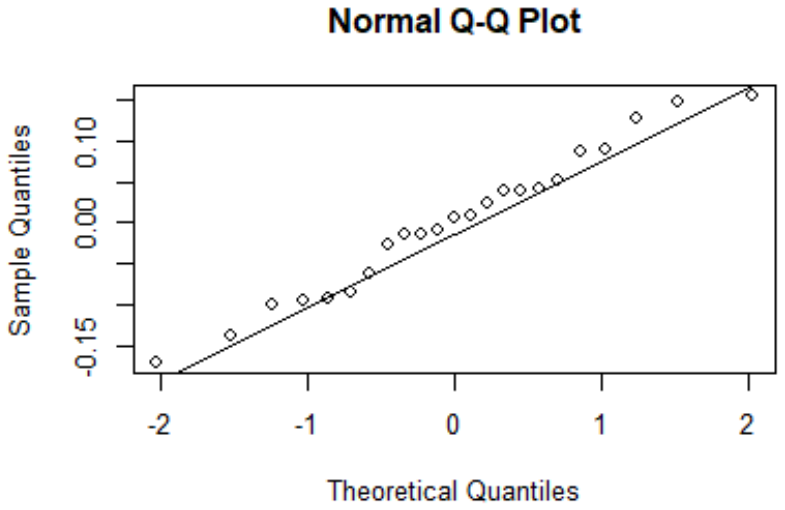
```
##
## Call:
## lm(formula = lg_TEAM_EVALUATION ~ AST + TRB + STL + BLK + lg_TWITTER_RETWEET_COUNT +
##     lg_GDP + lg_TICKET_PRICE, data = teammat)
##
## Residuals:
##       Min        1Q    Median        3Q       Max
## -0.171450 -0.074029  0.005667  0.046353  0.154039
##
## Coefficients:
##                            Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 1.18765    0.49924   2.379  0.03108 *
## AST                        -0.02764    0.01622  -1.704  0.10898
## TRB                        -0.04159    0.01315  -3.163  0.00643 **
## STL                         0.22708    0.10530   2.157  0.04767 *
## BLK                         0.12958    0.06060   2.138  0.04934 *
## lg_TWITTER_RETWEET_COUNT    0.23663    0.04443   5.326 8.47e-05 ***
## lg_GDP                      0.29129    0.03562   8.178 6.56e-07 ***
## lg_TICKET_PRICE             0.24447    0.09782   2.499  0.02454 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1076 on 15 degrees of freedom
## Multiple R-squared:  0.9373, Adjusted R-squared:  0.9081
## F-statistic: 32.06 on 7 and 15 DF,  p-value: 6.405e-08
```

Results: Team Valuation









Method: Player Salary

In our second model, we are exploring the relationship between players salary and their statistics, we will use players salary as response variables, and their statistics as explantory variables in our multivariate linear regression model:

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \cdots + \beta_{p-1} X_{p-1,i} + \varepsilon_i, \quad i = 1, 2, \cdots, n. \tag{1}$$

Method: Player Salary

I divide my analysis into 6 parts:

- Potential Predictors
- Variable Selection
- Diagnostics
- Cross Validation
- Conclusion

Method: Player Salary

Potential Predictors

- **The intercept**
- **Continuous measurements and age** All continuous measurement + **AGE**.
- **Dummy variables and factors** The **POSITION** with 4 levels are considered as *factor*.
- **Interactions** The interactions between **POSITION** and other predictors.

To simplify our model and make it interpretable, we do not consider the transformations or the polynomials of our predictors

Method: Player Salary - Variable Selection

AIC

1. **Backward Elimination** It chooses almost all potential predictors, and it gets AIC=641.89.
2. **Forward Selection** It chooses 'POINTS + AGE + DRB + PF + AST + eFG. + W + FG. + TOV + ORPM + FGA + PAGEVIEWS + TWITTER FAVORITE COUNT + TWITTER RETWEET COUNT + X2P' as predictor, and it gets AIC=671.14.
3. **Stepwise Regression** It chooses 'POINTS + AGE + DRB + PF + AST + eFG. + W + TOV + ORPM + FGA + PAGEVIEWS + TWITTER FAVORITE COUNT + TWITTER RETWEET COUNT' as predictor and it gets AIC=670.52

Method: Player Salary - Variable Selection

BIC

1. **Backward Elimination** It chooses "'AGE + MP + FT + MPG + PAGEVIEWS + TWITTER FAVORITE COUNT'" as predictors, and it gets BIC=708.97.
2. **Forward Selection** It chooses "'POINTS + AGE + DRB + PF'" as predictor, and it gets BIC=711.48.
3. **Stepwise Regression** It gets the same result as Forward Selection.

We adopt the predictors we get from the **Stepwise Regression** with BIC.

Method: Player Salary - Diagnostics

We can see that $\text{Var}(\mathbf{Y}|\mathbf{X}) \propto \text{E}(\mathbf{Y}|\mathbf{X})$, so we do the square root transformation to stablize the variance.

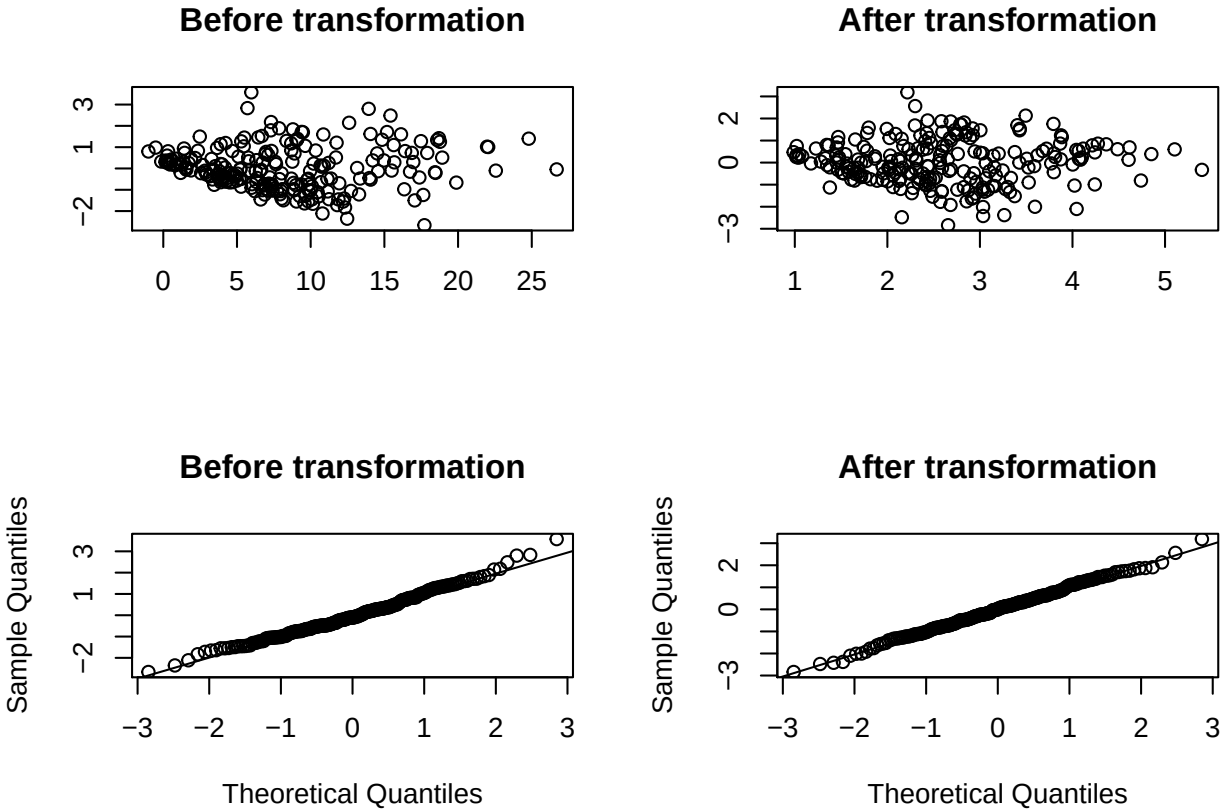


Figure: (1)Standardized residual has right-opening megaphone shape and QQ-plot does not fit very well before transformation (2) Standardized residual has no pattern and QQ-plot looks perfect after the sqrt-transformation

Method: Player Salary - K-fold Cross Validation

5-fold CV estimator is $0.6 \approx$ MSE.

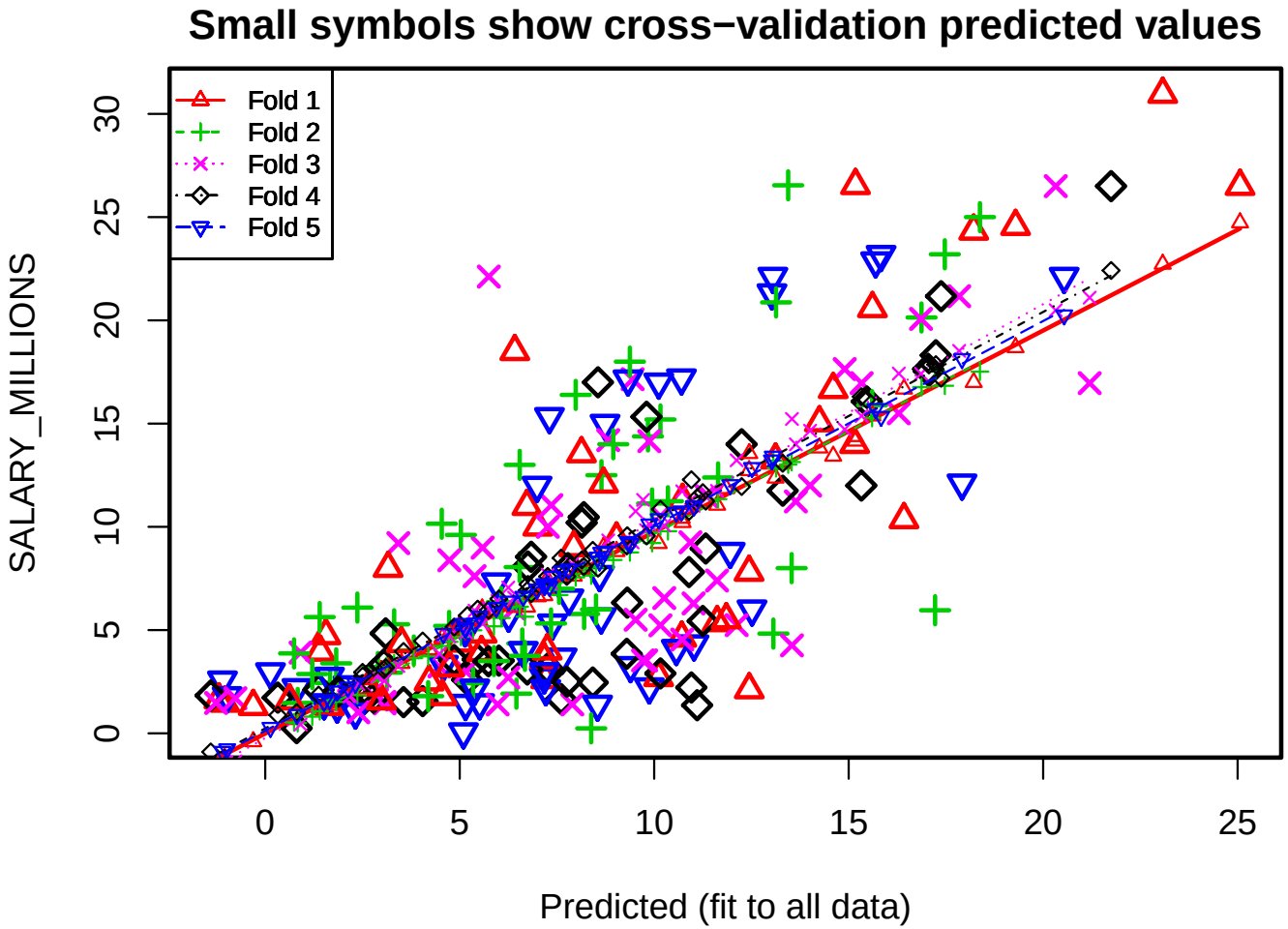


Figure: K-fold Cross Validation

Results: Player Salary

In our final model, the MLR model only contains three predictors, **POINTS + AGE + DRB** and the intercept, which is easy to interprete and understand. $R^2 \approx 0.57$

$$\sqrt{\text{Salary}} = -1.3582 + 0.0986 POINTS + 0.0975 AGE + 0.1012 DRB \quad (2)$$

The interpretation for the coefficient of **DRB** is, when Defensive Rebounds increases one unit and other predictors are fixed, the salary will approximately increase 0.1012 units. From this model, we can see that the coefficients of these predictors are all positive, which means they have positive contribution to the salary. We believe these three predictors are sufficient enough for us to do the prediction and provide a useful suggestion to the team manager.

Discussion

Limitations:

1. The transformations or the polynomials of our potential predictors.
2. How to choose appropriate variable selection methods and criteria?

Solution:

1. Considering more complex model which includes the transformation of predictors as potential predictors.
2. Asking for some expertise or getting some prior knowledge.

Future Work

We also notice that 100 players are considered as elite players based on their performance in the 2016-2017 season. We would like to predict the whether a player would be a elite player in the following season. We choose the same predictors as in our multivariate regression model.

```
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -9.38966    1.71713  -5.468 4.55e-08 ***
## POINTS       0.23350    0.04254   5.488 4.05e-08 ***
## AGE          0.12336    0.05024   2.455   0.0141 *
## DRB          0.64795    0.15119   4.286 1.82e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 272.30  on 224  degrees of freedom
## Residual deviance: 148.69  on 221  degrees of freedom
## AIC: 156.69
##
```