Predicting NBA Regular Season Results Based on some popular statistics

Chenjie Li,Qiao Qiao

November 16, 2018

1.Problem Description

1.1 Background

NBA(National Basketball Association) is one of the most successful basketball league around the world.And since it is popular, there exists tons of websites recording the statistics related to the game. They not only provide simple metrics like Field Goal Percentage, Average Points, Average Assists, but also some advanced metrics like PER,True shooting percentage and much more.

Based on those statistics, people are exploring what kinds of metrics can result in a team's success or failure as a season.

In this report, we investigated the relationship between the regular season results of the team and some popular metrics in two different perspectives:

- **Players' Effects**: **team PER**,

- **Teams' Overall Effects**: **Team's Offensive Efficiency**(TOE)

) for the past 5 seasons $(2014-2018)$.Based on the results we get, we made some predictions for the results of this uocoming season.

2.Data Source

There are many "data hubs" available, our data source is mainly from 1insider.espn.com/nba/hollinger/statistics and https://www.basketball-reference.com/.

To be able to work with data, we deveoped a "web crawler" using Python Scrapy library. By using the program we got the formatted CSV file so that we can easily input them in R.

Furthermore, since we are familiar with Database and SQL language, we also put the data in a database so we can manipulate the data easily by Queries.

3.Win Ration Analysis

3.1Team Offensive Efficieny Analysis

3.1.1 New TOE introduction and Experiments

Offensive Efficiency, as the name suggests, is a parameter measuring the efficiency of the offense.Here we explain this intuition by giving an simplified example:

Player A took 5 shots and made 4 of them; Player B took 5 five shots, and made 2 of them. Therefore, we can conlcude that player A is more efficient than Player B.

Of course, this is a oversimplified example.

There are a lot different kinds of definitions for Player and Team's Efficiency: In Shea,Stephen M's book[1], They define the Offensive Efficiency for Team as:

$$TOE = \frac{FG}{FGA - ORB + TO}$$

Where FG is the field goals made, FGA is the field goals attempts, ORB is the number of offensive rebounds, TO represents the number of turnovers.

It is of course reasonable to guess that the more efficient a team's offensive is, the more number of games a team is going to win.This idea was proved in their book: "The top 5 teams in OE in the 2012-2013 season all won at least 56 games."

This book was written and published five years ago. As time goes by, NBA statistics has been getting more and more comprehensive. Now we want to propose a new version of Team's offensive efficiency.

The formula we proposed is:

$$New\ TOE = \frac{2PTM + 1.5 * 3PTM}{FGA - ORB * ORBS\% + TO * OP\_POTS\%}$$

Where *2PTM* means 2 points field goals made;*3PTM* means 3 points field goals made; *ORBS%* means the offensie rebound scoring rate; *OP_POTS%* means opponent scoring rate off the team's turnovers.

Now let's motivate our formula. An individual offensive possesion could lead up to 4 possible results: taking a 2 point shot, taking a 3 point shot, missing a shot but getting an offensive rebound, and turning the ball over. We believe a 3 point shot is worth 1.5 2 point shot. Offensive rebound could compensate the results of missing the shot.Since most(not all of them, of course) second chance points are 2 points, we ignored the effect of the 3 point second chance points.but here we times Offensive rebound by "offensive rebound scoring rate", because different teams might have different abilities to convert a offensive rebound to box score.Here we also ignored the effect of 3 points made by opponent off turnovers.And finally, we add a factor "$TO * OP\_POTS\%$" to take the effects of turnovers into consideration.

In this section all our data is from *https://stats.nba.com/*[2].

We calculated the results from Season 2014 - Season 2017, here we only show two groups of results. For comprehensive results, please check out the Appendix part.

As Figure 1 and Figure 2 show, our new TOE is much better(significantly higher R^2 value) than evaluating the relationship between Win Ratio and Team Offensive Efficiency. We noticed the coefficient β_1 values keep increasing in 3 consectutive years, whereas in 2017 the β_1 drops again.
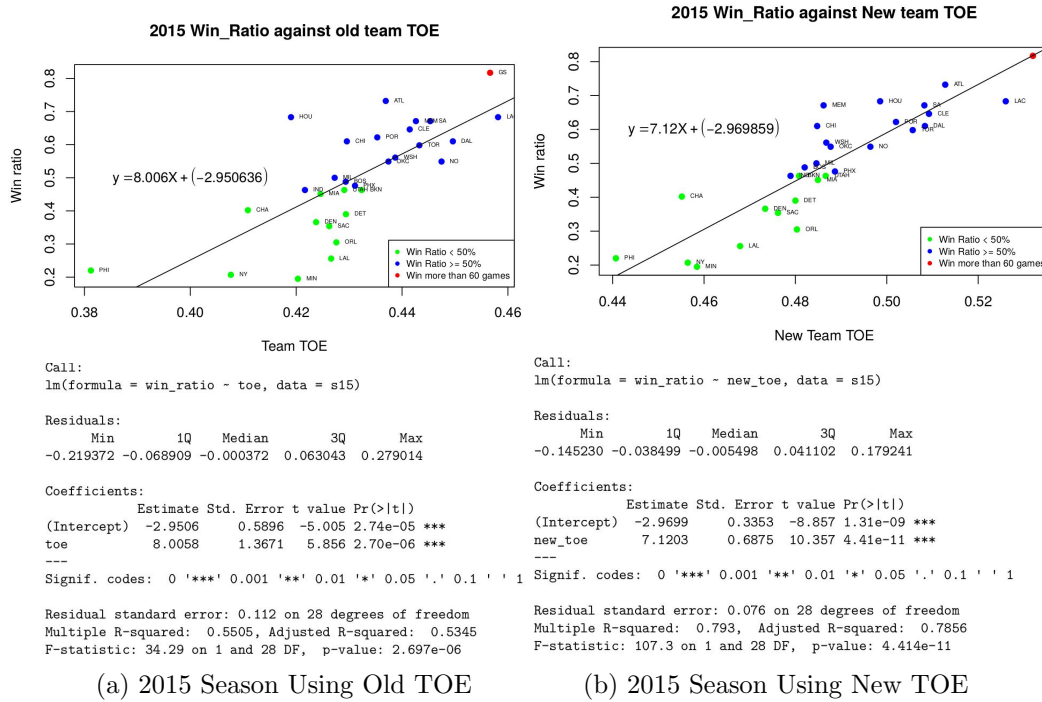
2015 Win_Ratio against old team TOE

$y = 8.006X + (-2.950636)$

Win Ratio < 50%
Win Ratio >= 50%
Win more than 60 games

```
Call:
lm(formula = win_ratio ~ toe, data = s15)

Residuals:
      Min        1Q    Median        3Q       Max
-0.219372 -0.068909 -0.000372  0.063043  0.279014

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.9506     0.5896  -5.005 2.74e-05 ***
toe           8.0058     1.3671   5.856 2.70e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.112 on 28 degrees of freedom
Multiple R-squared:  0.5505, Adjusted R-squared:  0.5345
F-statistic: 34.29 on 1 and 28 DF,  p-value: 2.697e-06
```

(a) 2015 Season Using Old TOE

2015 Win_Ratio against New team TOE

$y = 7.12X + (-2.969859)$

Win Ratio < 50%
Win Ratio >= 50%
Win more than 60 games

```
Call:
lm(formula = win_ratio ~ new_toe, data = s15)

Residuals:
      Min        1Q    Median        3Q       Max
-0.145230 -0.038499 -0.005498  0.041102  0.179241

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -2.9699     0.3353  -8.857 1.31e-09 ***
new_toe       7.1203     0.6875  10.357 4.41e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.076 on 28 degrees of freedom
Multiple R-squared:  0.793,  Adjusted R-squared:  0.7856
F-statistic: 107.3 on 1 and 28 DF,  p-value: 4.414e-11
```

(b) 2015 Season Using New TOE

Figure 1: Comparision Between TOE and New TOE in 2015 Season

Season	Old TOE	New TOE	Season	Old TOE	New TOE
2014	0.438	0.602	2014	7.558	6.623
2015	0.551	0.793	2015	8	7.12
2016	0.608	0.776	2016	8.874	7.811
2017	0.45	0.667	2017	6.645	6.399

Figure 2: Comparison of R^2(left) and β_1(right)

3.1.2 Prediction

Now we want to use first four seasons (2014-2017) data combined to make a prediction for season 2018, which has already ended.Our goals is to test how accurate our predictions are compared with the facts.

4. References

1. Shea, S. M., Baker, C. E. (2013). Basketball analytics: Objective and efficient strategies for understanding how teams win. CreateSpace Independent Pub. Platform.

2. Teams General Misc Statistics. Retrieved from *stats.nba.com*