Does Having More Players Scoring Greater than 15 Points Per Game Benefit NBA Teams' Success?

Devin Nordlund

2024-04-02

0.1 Introduction

0.1.0.1 Overview In sports, data analytics has been a growing part of the way teams operate. Teams have been leveraging a variety of analytics to help make informed decisions. The market for sports analytics has increased significantly since 2010. In the NBA, the game of basketball has evolved over time with a shift toward three point shooting and higher scoring outputs. Scoring points is one of the most important keys to winning basketball games, and understanding the relationship between higher scoring output from multiple players could be essential to making in-game decisions and team construction. Individual player performance is a key component of increasing overall team points (Choo et al. 2020), which subsequently can often lead to winning and successful outcomes. For a professional sports team, the main goal is to win a championship. This can be achieved by first having enough regular season success to make a playoff appearance. The NBA regular season format consists of 82 games. The playoffs currently include 16 of the 30 teams in the NBA, with the top eight teams from each of the Eastern and Western Conferences. The best regular season teams are not always the ultimate winners of the championship, so many teams are vying for a playoff spot for the opportunity. Team executives have used many strategies in team-building, with varying results. Some examples of this include "superteams", in which a team has at least three perennial superstars on the team. In a league that is typically offensively oriented, we would expect a positive correlation between the number of players scoring above 15 points per game and the corresponding team success metrics. A higher output from more players would appear to be signs of a more cohesive team, with multiple options to score points and ultimately win games in the face of different strategies.

0.1.0.2 Existing Works

The article “Data-Driven Approaches to NBA Team Evaluation and Building” (Choo et al. 2020) states that teams at “different stages of the success cycle and with different resources, will use these strategies differently”. The more successful teams at a point in time may allocate their resources towards maintaining current success, which may involve trading away draft picks or younger, unproven players, for proven players in the league who can benefit the team’s current performance. This article also discussed an evaluation of player performance based on a tree model, with the possible outcomes of a possession being taken into consideration. This was used to gather an estimated impact on expected team points. In this article, the cohesion of a roster is also mentioned to have a positive impact on winning percentage. A 2022 study (Cabarkapa et al. 2022) focused on finding the variables that made the largest differences in winning and losing games. There was a statistically significant difference found between scoring in wins versus losses. This was true both in the regular season and the postseason, but the difference was larger in the postseason, which could be attributed to the more competitive nature of the games. Shooting efficiently was also found to lead to more winning outcomes. In this study, the winning teams were found to have made more shots with higher shooting percentages on all field goals, three-pointers, and free throws. Shooting the ball more efficiently has been shown to lead to more shots made, points scored, and more wins. In the case of three-pointers, teams that shoot them efficiently can score more efficiently, with more points scored on less possessions and attempts. This aligns with the notion that shooting efficiency is key to determining game outcomes. A 2020 study (Terner and Franks 2021) claims that scoring ability is the most important part of player performance. This not only includes skill, but the shot selection of the player. Shot selection takes many things into account, such as the defensive pressure, situational factors, teammate ability, and the specific skillset of the player. Players will typically choose shots that align with their skillset, with an approximately linear relationship between the log relative frequency of two-point and three-point shot attempts. Coaching and team strategy also comes into play here, with good coaching being able to give players better opportunities to take good shots. A 2010 study (Lorenzo et al. 2010) corroborates the importance of scoring to winning and team success. This study was focused on the under-16 male basketball games. These players were not professionals, but similarities were found between their play and that of NBA players. A discriminant analysis was used to find the most important variables in final outcomes. For games that were decided by a margin between 10 and 29 points and those decided by more than 30 points, the main discriminants were successful two-point field goals. At a younger age, the game is played differently, but the same principles can be applied to all levels of basketball. Making more shots and subsequently scoring more points is shown to be related with winning games.

0.1.0.3 Gap in Research These findings demonstrate the importance of individual and team scoring to winning and overall team success. The gap in research lies in the exact team scoring structure and how this relates to winning. Existing literature has explored the relationships between scoring efficiency, individual player performance, and team success in basketball. There remains a gap in understanding how having more players scoring greater than 15 points per game impacts team success. The distribution of scoring output across players on teams in the NBA has not been studied in depth. This is necessary to address because basketball is a team sport, and individual starpower is not enough to make a team successful. As three-point shooting has become more prevalent in the NBA, the total scoring outputs have increased, and the need for more high-scoring players has risen. The teams with more players consistently scoring 20 points per game may have an advantage over more one-dimensional teams, with only one or two players scoring most of the points. With the increasing usage of data analytics in basketball and sports as a whole, teams are looking to find an advantage in any way possible. Understanding the impacts of having multiple high-scoring players on team success can benefit teams by informing their strategies on the court and in team construction in order to maximize their chances at winning games, making playoff appearances, and winning championships.

0.1.0.4 Contributions This study aims to fill existing gaps in literature by examining the relationships between the number of players scoring greater than 15 points per game and how successful NBA teams are. The information gained from this study will have practical applications in the sport of basketball, by informing team decisions and coaching strategies, leading to a greater competitive balance in the NBA.

0.1.0.5 Summary Following this section, the details of the dataset will be explained. This will discuss the sourcing of the data, data preprocessing, sampling techniques used, and the usefulness of the data to the field of basketball. Exploratory data analysis is done using visualizations of the overall scoring trends in the NBA. The average points scored per game over time is shown, and a comparative box plot is used to show the relationship between average team points scored and making the playoffs. Then, the regression analysis techniques will be discussed in detail, using multiple regression to model the relationship between the number of high-scoring players and winning outcomes.

0.2 Data

The dataset used in this study is sourced from Kaggle, compiled by Sumitro Datta. The data contains observations spanning from 1947 to 2024 across the NBA and ABA, but for ease of use and validity of data, we will only be including observations from 1980 to 2023. The ABA was a competing league with the NBA from 1967-1976, but then merged with the NBA in 1976. In 1980, the three-point shot was introduced, and the game style corresponds much more with the game played today. This dataset contains exhaustive information about player and team statistics per game in professional basketball, as well as end of season summaries which provides insights into player performance and team success. There are 35 total variables included in the table of player statistics, however the variables of interest from this include: points per game, team, and year, as the goal is to calculate the count of players on each team scoring 15 points per game or more each year. The player statistics table includes observations from 5,153 players. The team statistics table contains 1681 team observations, which is due to the fluctuation in league format and team count over time. The variables of interest from this table are team points per game, and year. The team summaries table contains information about end of season success, and the variables of interest from this table include playoffs (boolean), wins, and year. In order to prepare this data for the study, the outlined variables of interest will be combined into one table, with the player and corresponding team statistics organized by year. Using the comprehensive statistics from a wide span of years will allow us to view the impact of having multiple players scoring 15 points across multiple eras of basketball.

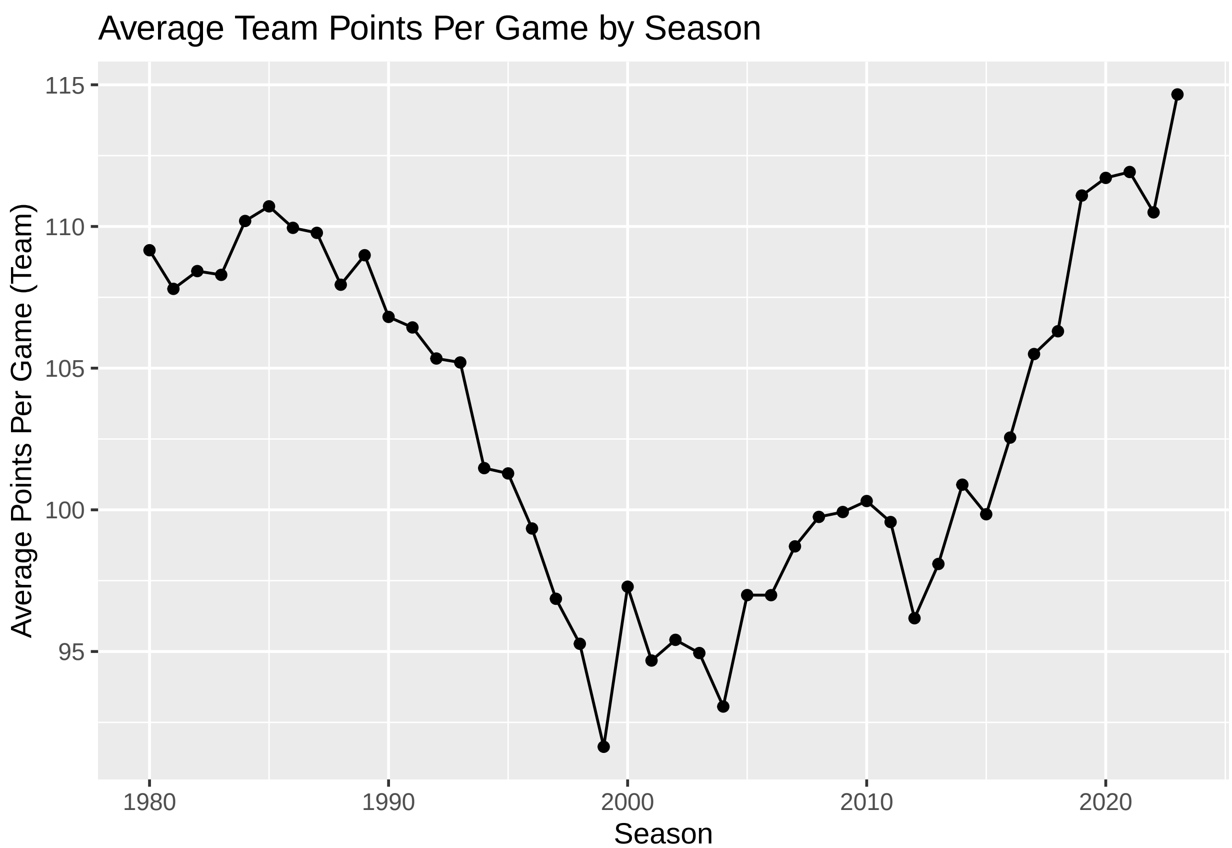
```
library(dplyr)
```

0.2.0.1 Exploratory Analysis

```
##
## Attaching package: 'dplyr'

## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```r
nba_data<- read.csv("bball_current.csv")

season_avg_ppg<- nba_data %>%

  group_by(season) %>%

 summarize(avg_pts_per_game_x = mean(pts_per_game.x, na.rm = TRUE),

            avg_pts_per_game_y = mean(pts_per_game.y, na.rm = TRUE))
```

```r
library(ggplot2)

ggplot(season_avg_ppg, aes(x=season, y = avg_pts_per_game_y)) +

  geom_line() +

  geom_point() +

  labs(x = "Season", y = "Average Points Per Game (Team)", title = "Average Team Points Per Game by Seas
```

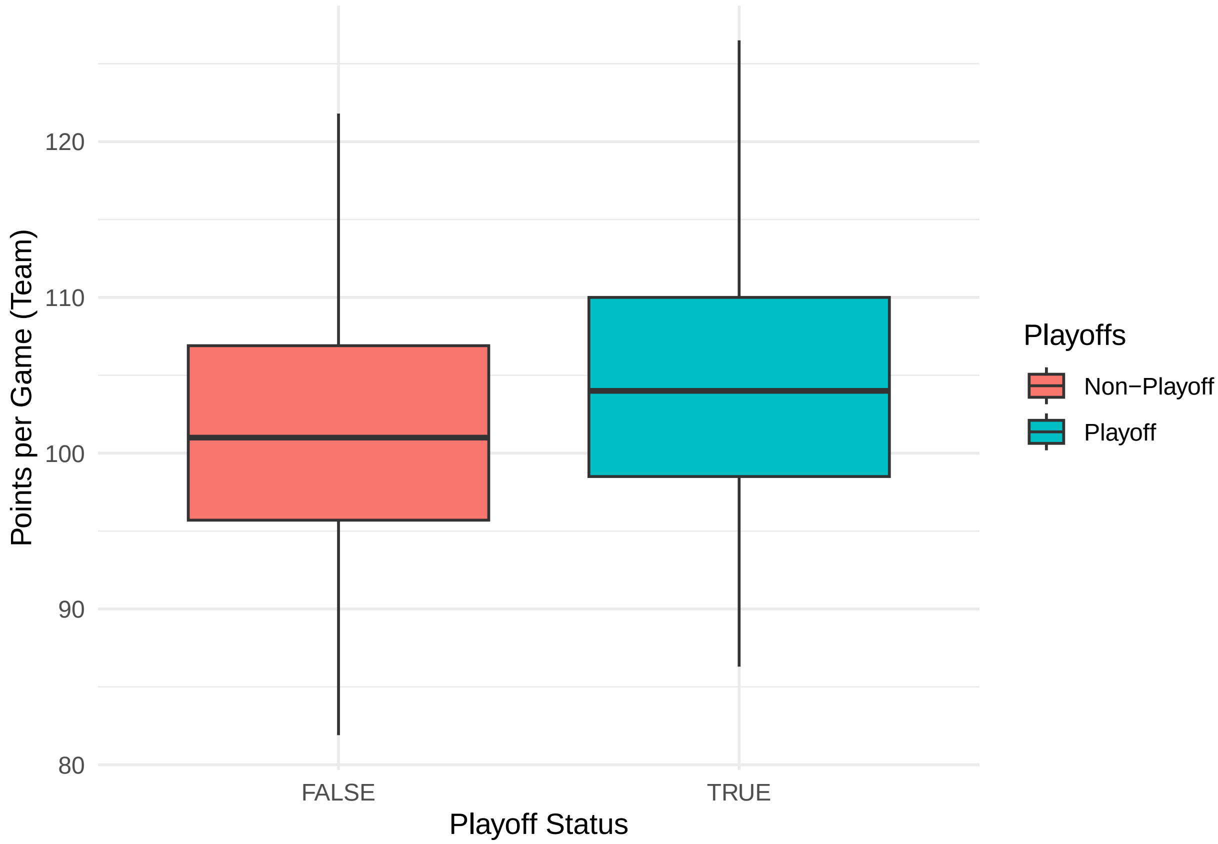



```r
library(dplyr)

team_totals <- nba_data %>%

  distinct(tm, season, playoffs) %>%

  group_by(tm) %>%
```

```r
  summarize(total_playoffs = sum(playoffs, na.rm = TRUE)) %>%
  arrange(desc(total_playoffs))
top_10_teams<-head(team_totals, 10)
print(top_10_teams)
```

```
## # A tibble: 10 x 2
##    tm    total_playoffs
##    <chr>          <int>
##  1 SAS               36
##  2 LAL               34
##  3 POR               34
##  4 BOS               33
##  5 UTA               31
##  6 HOU               30
##  7 ATL               28
##  8 OKC               28
##  9 CHI               27
## 10 IND               27
```

```r
# Create a boxplot comparing pts_per_game.y between playoff and non-playoff teams
ggplot(nba_data, aes(x = playoffs, y = pts_per_game.y, fill = playoffs)) +
  geom_boxplot() +
  labs(x = "Playoff Status", y = "Points per Game (Team)") +
  scale_fill_discrete(name = "Playoffs", labels = c("Non-Playoff", "Playoff")) +
  theme_minimal()
```

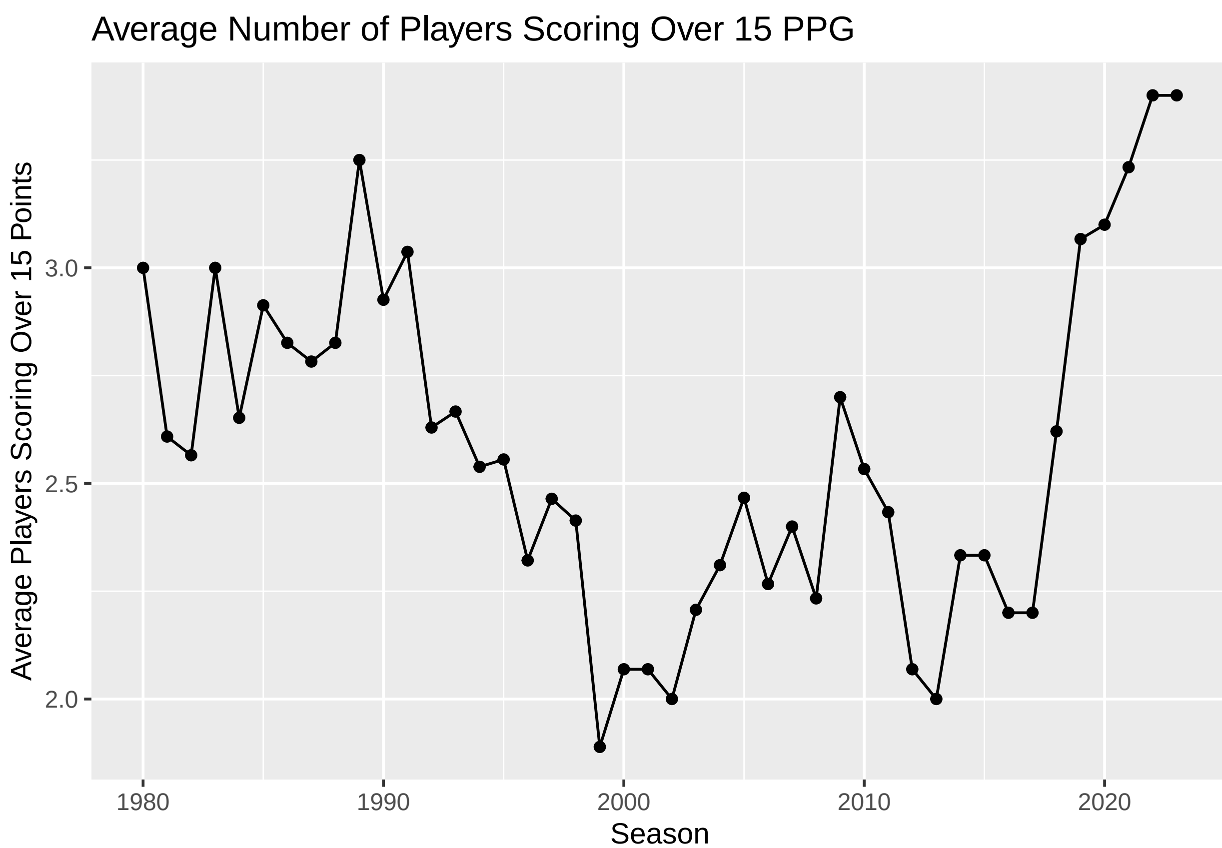



```
over_15_scorers<-nba_data %>%
  filter(pts_per_game.x >= 15) %>%
  select(season,tm,player) %>%
  distinct()
over_15_counts<-over_15_scorers %>%
  group_by(season, tm) %>%
  summarize(num_over_15 = n())
```

```
## `summarise()` has grouped output by 'season'. You can override using the
## `.groups` argument.
```

```
average_over_15<-over_15_counts %>%
  group_by(season) %>%
  summarize(average_players = mean(num_over_15))
ggplot(average_over_15, aes(x=season, y=average_players)) +
  geom_line() +
```

```
geom_point() +
labs(x = "Season", y = "Average Players Scoring Over 15 Points") +
ggtitle("Average Number of Players Scoring Over 15 PPG")
```




```
nba_updated_data<-left_join(nba_data, over_15_counts, by = c("season", "tm"))
```

0.3 Methods

For the purposes of this study, we will be using linear regression models. In the first model, let

$$Y$$

denote the number of players scoring greater than 15 points per game on an NBA team for a given season.

$$X$$

will denote the chosen team success metric, the number of wins. The linear regression model is shown below, where

$$\epsilon$$

is the error term.

$$Y = \beta_0 + \beta_1 X + \epsilon$$

We will be estimating the parameters

$$\beta_0$$

and

$$\beta_1$$

to evaluate and understand the relationship present between the number of players scoring 15 points per game and wins. For the second linear regression model, let

$$Y$$

denote the number of players scoring greater than 15 points per game on an NBA team, and let

$$X$$

denote the corresponding playoff status for that year. The Ordinary Least Squares Method will be used to estimate

$$\beta_0$$

and

$$\beta_1$$

based on the observed data. For each of these coefficients, the standard errors of the point estimators will be calculated as a means to measure the level of uncertainty. The OLS method will estimate the variances of the point estimators, which are important for hypothesis testing. The hypothesis testing will be used to establish the null distribution of the testing statistics. We assume that there is a linear relationship between the number of players scoring greater than 15 points per game and team success metrics. We claim that the number of players scoring greater than 15 points per game significantly impacts the success of their team.

0.4 Simulation

The aim of this analysis is to investigate the relationship between the number of players on a team scoring greater than 15 points per game and team success in the NBA. Specifically, the goal is to determine if having more players scoring greater than 15 points per game is associated with more wins and playoff appearances. The data to be used in this study is real, observational data, therefore resampling and simulation are not necessary. The main factor of interest is the number of players scoring greater than 15 points per game. Other factors considered are the team success metrics, which include wins and playoff appearances. Using regression analysis, we will analyze the data to learn about the relationship between the number of players scoring greater than 15 points per game and team success. This will be done using linear regression to create a model for the relationship. Additionally, diagnostic tests will be used to check for key assumptions and the validity of the model. Evaluating the performance of this model will be done using measures such as the R-squared value, indicating how much of the variance in team success is explained by the number of players scoring more than 15 points per game.

0.5 Code

```
library(car)
```

```
## Loading required package: carData

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode
```

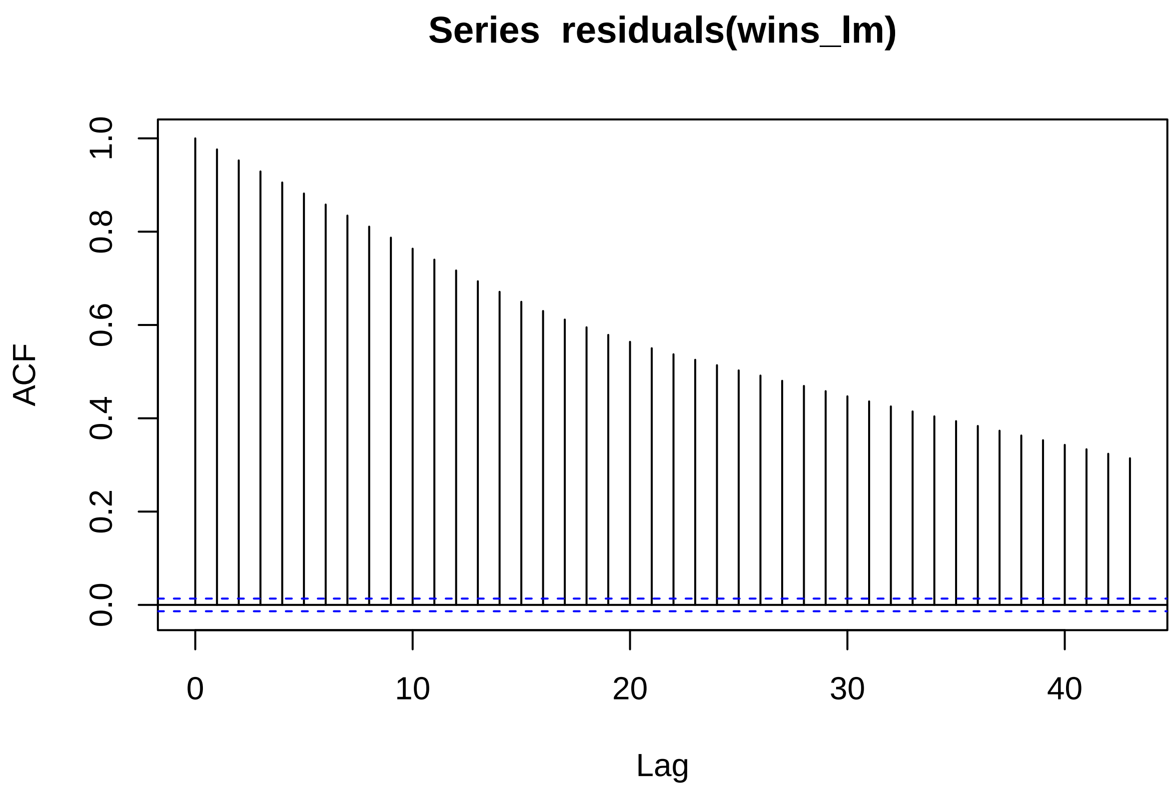
```
library(forecast)
```

```
## Registered S3 method overwritten by 'quantmod':
##   method            from
##   as.zoo.data.frame zoo
```

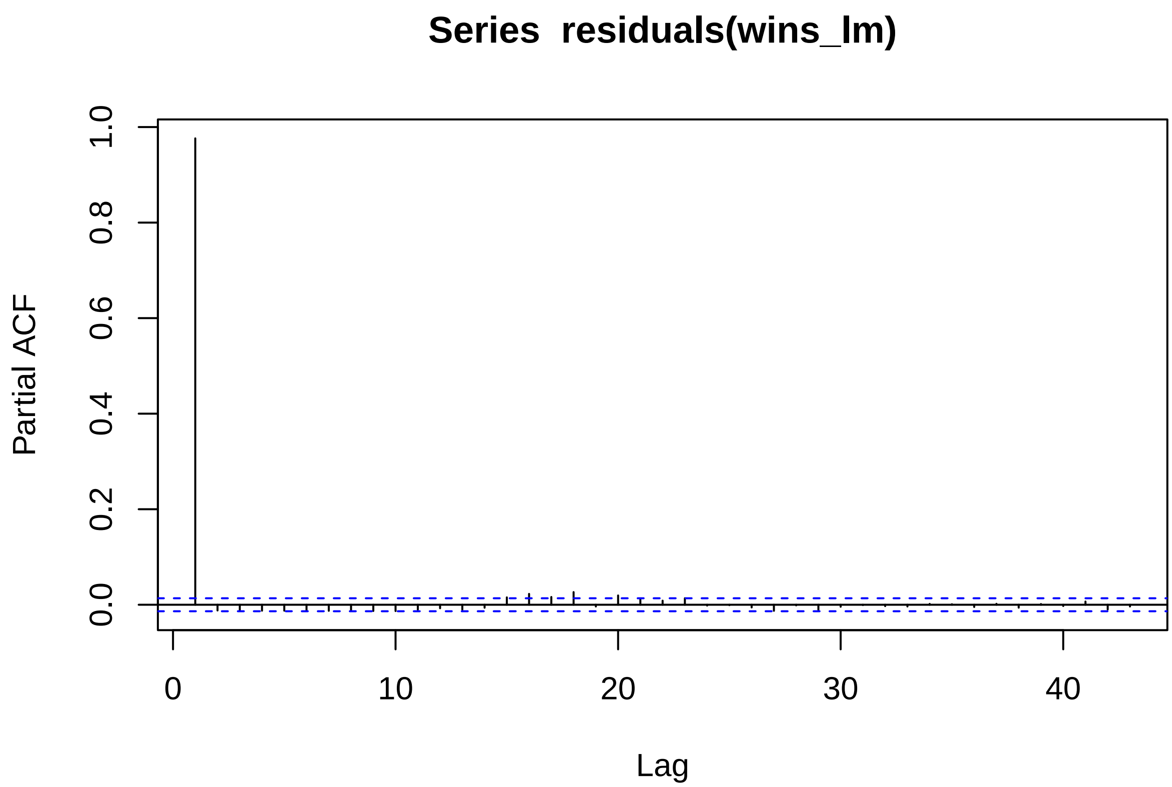
```r
wins_lm<-lm(w~num_over_15, data = nba_updated_data)
durbinWatsonTest(wins_lm)
```

```
##  lag Autocorrelation D-W Statistic p-value
##    1       0.9764156    0.04713365       0
##  Alternative hypothesis: rho != 0
```

```r
## modeling this has been very difficult due to issues with autocorrelation.
## linear regression does not seem to be a good fit, and adding lags and differencing has not seemed to
nba_updated_data$lag1 <- lag(nba_updated_data$num_over_15, 1)
nba_updated_data$lag2 <- lag(nba_updated_data$num_over_15, 2)
nba_updated_data$lag3 <- lag(nba_updated_data$num_over_15, 3)




wins_lm <- lm(w ~ num_over_15 + lag1 +lag2+lag3, data = nba_updated_data)
acf(residuals(wins_lm))
```




```
pacf(residuals(wins_lm))
```




```r
summary(wins_lm)
```

```
##
## Call:
## lm(formula = w ~ num_over_15 + lag1 + lag2 + lag3, data = nba_updated_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -32.279  -9.608   1.049   9.721  33.721
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.794e+01  2.456e-01 154.473   <2e-16 ***
## num_over_15  6.044e-01  3.205e-01   1.886   0.0593 .
## lag1        -7.885e-13  4.489e-01   0.000   1.0000
## lag2         4.518e-13  4.489e-01   0.000   1.0000
```

```
## lag3         6.703e-02  3.205e-01   0.209   0.8343
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 12.65 on 20842 degrees of freedom
##   (211 observations deleted due to missingness)
## Multiple R-squared:  0.002882,   Adjusted R-squared:  0.002691
## F-statistic: 15.06 on 4 and 20842 DF,  p-value: 2.679e-12
```

Sources

Cabarkapa, Dimitrije, Michael A. Deane, Andrew C. Fry, Grant T. Jones, Damjana V. Cabarkapa, Nicolas M. Philipp, and Daniel Yu. 2022. "Game Statistics That Discriminate Winning and Losing at the NBA Level of Basketball Competition." *PLOS ONE*. Public Library of Science. https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0273427.

Choo, Hyunsoo, David Creegan, Olivia Majedi, Daniel Smolyak, and Brian Valcarcel. 2020. "DATA-DRIVEN APPROACHES TO NBA TEAM EVALUATION AND BUILDING." *University of Maryland DRUM*. https://api.drum.lib.umd.edu/server/api/core/bitstreams/bed7e07f-2360-4e0d-a30e-4f0b32950a7d/content.

Lorenzo, Alberto, Miguel Angel Gomez, Enrique Ortega, Sergio José Ibáñez, and Jaime Sampaio. 2010. "Game Related Statistics Which Discriminate Between Winning and Losing Under-16 Male Basketball Games." *Journal of Sports Science & Medicine*. U.S. National Library of Medicine. https://pubmed.ncbi.nlm.nih.gov/24149794/.

Terner, Zachary, and Alexander Franks. 2021. "Modeling Player and Team Performance in Basketball." *Annual Review of Statistics and Its Application*. Annual Reviews. https://www.annualreviews.org/content/journals/10.1146/annurev-statistics-040720-015536#right-ref-B38.