# Relationship between Three-Pointers and Wins in the NBA

Abeyan Thirunavukkarasu

20569014

Abstract: This analysis evaluates the significance of three pointers in the NBA verifying whether or not the hype of the significance of three pointers is true. The data used consists of an aggregate of individual player statistics covering a wide range of variables that may or may not be related to a team's success. A multi-linear regression model was constructed with R-squared 0.6753 leading towards concluding that the model is relatively accurate. From the estimates generated as well as the p-values which were used to determine parameter significance, it was found that three-pointers made was insignificant in explaining a team's success. Furthermore, the more three-pointers made was found to relate towards losing as a negative beta estimate was presented with high significance (low p-value).

## Introduction

The movie Moneyball presented the significance of statistical analysis for team managers in developing and organizing their teams. After the growth of Sabermetrics, team managers in sports other than baseball began to look at Statistics to make predictions as well as evaluate different team models that would provide them the highest chance of success. After seeing the success of Sabermetrics in sports gambling as well as team modelling, APBRmetrics began to grow, as the significance of data in basketball began to be realized. While there are several models that make win/loss predictions using different and varying methods, this paper aims to look specifically at the relationship between three pointer shots and a team's success.

The nature of basketball has changed over the years, and visibly, players have shifted to taking more three pointers relative to the past (Mather, 2018). With more regulations on fouling and defending, the sport is less aggressive then it was before. Thus, it makes sense that players are more reliant on taking three pointers since they are not as heavily guarded as they used to be in previous generations (McMenamin, 2018). At the same time, it is understandable that ceteris paribus, a three-pointer shot is harder to make then a field goal or free throw. By analyzing the three pointers and its relation to the team's results, coaches and team managers can use this simple analysis to restructure their team and prioritize the kind of plays its team members make.

From the data set which outlines aggregated team stats over games played since 2014 and up to 2018, after constructing a multi-linear regression model, we find that three pointers made is not a strong indicator of victory. The regression model provided an estimate of 1.351e-03 with a p-value of 0.505 for this variate, and thus leads to the conclusion that it is not a significant variate in explaining the winning or losing results of a team. On the other hand, the number of three pointers taken is estimated to be -2.862e-03 with a corresponding p-value $< 0.001$. This allows

us to reject the null hypothesis and accept three pointers taken as a significant variate. The hype surrounding players such as Stephen Curry and Klay Thompson, who are known for being very efficient with their three-point shots (43.6% and 41.9% respectively) led to a reasonable hypothesis that three-pointers are a major component of a team's success (Basketball Reference, 2019). The findings from this study contradict this hypothesis, as a negative beta estimate outputted from the regression model for three pointers taken illustrates that the more threes a team takes on average, the less likely they are to win accounting for all other variates. Based off this as well as a negligible three-points made estimate, it is concluded that three-pointers are in actuality not an essential component to winning a game and it is actually better for a team to focus on other types of scoring such as offensive rebounds which from the regression model provided a low p-value.

The data provided opponent stats as well per game, and it is observable that as expected, chances of winning is also dependent on the opponent's performance. According to the regression model made to analyze the data set, opposing team three-pointers attempted is a significant variate in estimating whether a team will win or lose. This mirrors what was found with the team being analyzed and supports the conclusion that it is not beneficial for a team to on average take more threes and should not pivot total strategy away from other areas. From the perspective of endogenous variables that a team manager can control, focus should be placed on non-scoring factors such as offensive rebounds.

## Methodology/Data

**Figure 1: Dataset**

| | X | Team | Game | Date | Home | Opponent | WINorLOSS | TeamPoints | OpponentPoints | FieldGoals | FieldG |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | ATL | 1 | 2014-10-29 | Away | TOR | L | 102 | 109 | 40 | |
| 2 | 2 | ATL | 2 | 2014-11-01 | Home | IND | W | 102 | 92 | 35 | |
| 3 | 3 | ATL | 3 | 2014-11-05 | Away | SAS | L | 92 | 94 | 38 | |
| 4 | 4 | ATL | 4 | 2014-11-07 | Away | CHO | L | 119 | 122 | 43 | |
| 5 | 5 | ATL | 5 | 2014-11-08 | Home | NYK | W | 103 | 96 | 33 | |
| 6 | 6 | ATL | 6 | 2014-11-10 | Away | NYK | W | 91 | 85 | 27 | |

The data used was found on the website Kaggle and is an aggregation by Kelpouris of individual player statistics for every game played between 2014 and 2018 (Kelepouris, 2018). For example, the first row of the data set lists Atlanta's overall statistics against Toronto on October 29, 2014. The aggregated data includes: date the game was played, the results (win or loss), the team's total points, field goals made, field goals attempted, three-pointers made, three-pointers attempted, free throws, offensive rebounds, total rebounds, assists, steals, blocks, turnovers, and total fouls. Likewise, the same variates were provided for the opposing team which as expected will influence the results. The total number of games played over this time period was 9841 games, which is a relatively large sample size when looking for modern data. Note that for the three-pointers analysis, taking an aggregate of individual players and averaging them out as a team (which is what the data set has done), is a fair method for analyzing three-pointers since while teams are made of different role players (point guard, shooting guard, center, power-forward, small-forward), players who take threes usually are role players who are expected to make threes.

Due to the nature of data, it was simple to construct a multi-linear regression model and see the estimates produced for the parameters included in the data set. Since the response variate analyzed was the success of the team (win or loss), the categorical variates were assigned binary

values of 1 representing a win and 0 representing a loss. Through a regression model, the estimates would regress to contribute towards reaching a value of 1 permissive of winning-related events taking place. Negative estimates on the other hand contributed towards a response variate of 0 representing a loss. Logically, from the constructed model, 0.5 represents no indication of whether the model estimates a loss or win. If the response variate is greater than 0.5, a win is estimated and if the response variate is less than 0.5, a loss is estimated. The multi-linear regression model presented an R-squared value of 0.6753, meaning around 67.53% of the data can be explained by the results of the regression conducted. All of the variates from the data set except the percentage shots made results (results in singularity since other variates cover this component) and the time factor of games played were included due to observing them as being potential influencers on overall win. Time (date game played) was not included since all teams play with similar schedules not leading it to be a significant or useful indicator of estimating team victory in this case.

While the focus of this study is to evaluate three-pointers specifically, the model presents parameters that can be analyzed by team managers to figure out where attention should be placed with regards to strategy and tactics. By analyzing the significance value (P-value) calculated by the model, it is determined whether the estimate generated for each parameter was of value or not (mainly interested in whether positive or negative and relative magnitude since lots of variates that influence overall win). The model is not very flexible, however since it has a relatively large R-squared value it is effective at outlining from a general perspective, areas that team managers should take a closer look at as well as areas that may have hype surrounding it but is insignificant in leading one’s team towards victory.

## Results

**Figure 2: Linear Regression Output**

```
Call:
lm(formula = win ~ data$TeamPoints + data$OpponentPoints + data$FieldGoals +
    data$FieldGoalsAttempted + data$X3PointShots + data$X3PointShotsAttempted
+
    data$FreeThrowsAttempted + data$OffRebounds + data$TotalRebounds +
    data$Assists + data$Steals + data$Blocks + data$Turnovers +
    data$TotalFouls + data$Opp.FieldGoals + data$Opp.FieldGoalsAttempted +
    data$Opp.3PointShots + data$Opp.3PointShotsAttempted + data$Opp.FreeThrow
sAttempted +
    data$Opp.OffRebounds + data$Opp.TotalRebounds + data$Opp.Assists +
    data$Opp.Steals + data$Opp.Blocks + data$Opp.Turnovers +
    data$Opp.TotalFouls)

Residuals:
    Min      1Q  Median      3Q     Max
-1.2391 -0.2377  0.0000  0.2377  1.2391

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                   5.000e-01  5.252e-02   9.520  < 2e-16 ***
data$TeamPoints               2.838e-02  1.401e-03  20.256  < 2e-16 ***
data$OpponentPoints          -2.838e-02  1.401e-03 -20.256  < 2e-16 ***
data$FieldGoals              -1.472e-03  2.672e-03  -0.551  0.58169
data$FieldGoalsAttempted     -9.515e-03  1.420e-03  -6.703 2.16e-11 ***
data$X3PointShots             1.351e-03  2.024e-03   0.667  0.50468
data$X3PointShotsAttempted   -2.862e-03  7.215e-04  -3.967 7.34e-05 ***
data$FreeThrowsAttempted     -4.155e-03  1.348e-03  -3.082  0.00207 **
data$OffRebounds              5.691e-03  1.891e-03   3.010  0.00262 **
data$TotalRebounds            2.078e-03  1.298e-03   1.601  0.10939
data$Assists                 -7.307e-05  7.850e-04  -0.093  0.92584
data$Steals                  -4.197e-04  1.554e-03  -0.270  0.78711
data$Blocks                   1.869e-03  1.284e-03   1.455  0.14558
data$Turnovers               -1.258e-02  1.699e-03  -7.402 1.45e-13 ***
data$TotalFouls              -9.220e-03  1.194e-03  -7.725 1.23e-14 ***
data$Opp.FieldGoals           1.472e-03  2.672e-03   0.551  0.58169
data$Opp.FieldGoalsAttempted  9.515e-03  1.420e-03   6.703 2.16e-11 ***
data$Opp.3PointShots         -1.351e-03  2.024e-03  -0.667  0.50468
data$Opp.3PointShotsAttempted 2.862e-03  7.215e-04   3.967 7.34e-05 ***
data$Opp.FreeThrowsAttempted  4.155e-03  1.348e-03   3.082  0.00207 **
data$Opp.OffRebounds         -5.691e-03  1.891e-03  -3.010  0.00262 **
data$Opp.TotalRebounds       -2.078e-03  1.298e-03  -1.601  0.10939
data$Opp.Assists              7.307e-05  7.850e-04   0.093  0.92584
data$Opp.Steals               4.197e-04  1.554e-03   0.270  0.78711
data$Opp.Blocks              -1.869e-03  1.284e-03  -1.455  0.14558
data$Opp.Turnovers            1.258e-02  1.699e-03   7.402 1.45e-13 ***
data$Opp.TotalFouls           9.220e-03  1.194e-03   7.725 1.23e-14 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.2853 on 9813 degrees of freedom
Multiple R-squared:  0.6753,   Adjusted R-squared:  0.6745
F-statistic: 785.1 on 26 and 9813 DF,  p-value: < 2.2e-16
```

As observed from the above linear regression, the model captures a significant portion of variates relating to the response variate. With a relatively high R-squared value and the respective p-values presented for individual variates, the model is quite significant.

What is observed first and foremost is that three-pointers made is insignificant. A p-value of 0.505 leading to a fair assumption that we can accept the null hypothesis that three-pointers made is insignificant and has a beta estimate of 0. This debunks the hype presented in the articles referenced in the introduction. Note however, this model includes both open and guarded three-pointers and does not reflect specific cases of three-pointers.

If you look at three points attempted, a negative beta estimate is presented with relatively high significance. This indicates that taking more threes is related to winning and when opposing team takes more threes, it is related to your team winning. This is another interesting observation indicating more caution when taking threes. Perhaps only highly efficient three-point shooters should be taking these shots.

With regards to the other variates, the variates with stars represent being significant in estimating whether or not a team will win, and the ones without stars can be regarded as being insignificant towards estimating success. Overall, the model indicates that coaches should focus on non-scoring factors such as reducing number of fouls, reducing turnovers, and improving offensive rebounds as these were all significant (small p-values).

The three graphs below show that some variables behave in a non-linear method. For instance, the residuals vs. fitted plot ideally should present constant variance in errors. The Q-Q model should be closer to a straight line which is a condition for the normal distribution of errors. The residuals vs. leverage graph shows that the cook’s distance is horizontal which would normally

not be the case for a linear model. These graphs present that this model is unique and does not follow conventional model conditions (constant variance of errors, normally distributed errors, and functional form) because of the categorial nature of the response variate (fitted relative to observed value will always be difference between value and either 0 or 1). The model is still effective in answering the research question of whether or not three-pointers are significant when looking at the R-squared presented from the model.

**Figure 3: Residuals vs Fitted**

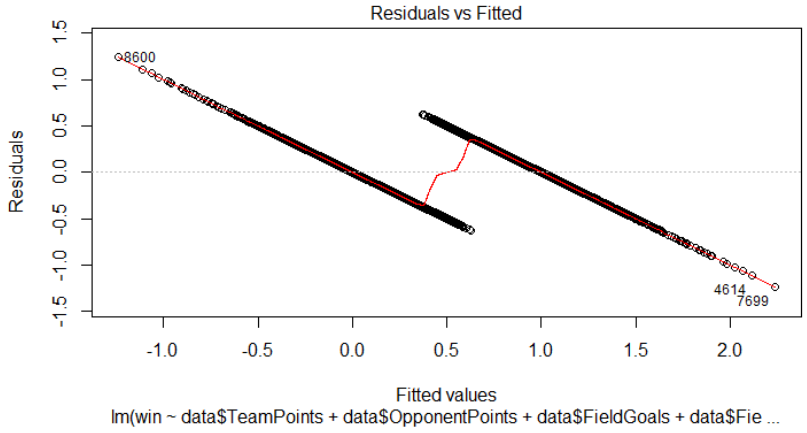



**Figure 4: Normal Q-Q**

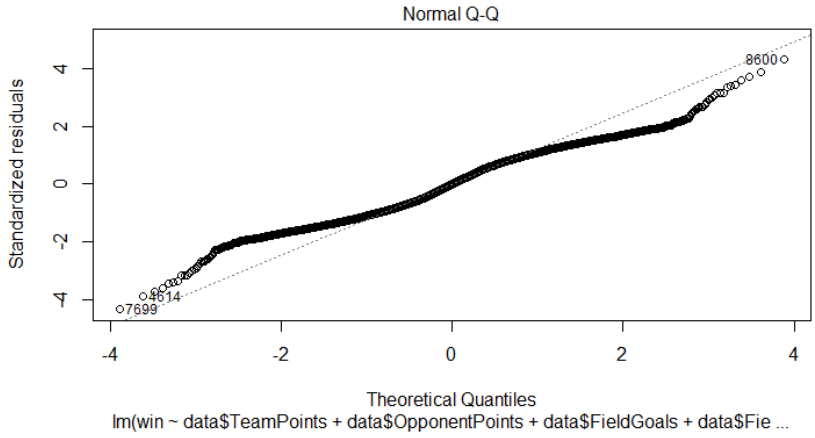
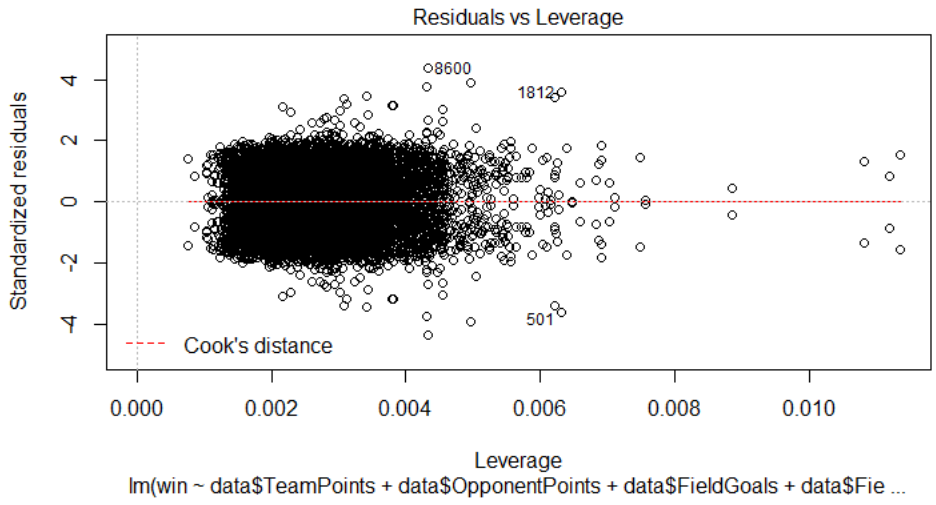


**Figure : Residuals vs Leverage**




## Conclusion

Overall, it can be concluded that three-pointers are not as effective as media hype indicates. With low significance values being presented for threes made by your team and the opposing team, and high significance values indicating that more threes is related towards losing, team coaches might switch their strategy from taking more threes to solely limiting high efficiency players to taking threes.

The limitations of this study include aggregating all players together. For example, a team has players who specialize in different roles. A center is not as likely to make a three as a point guard, however the data aggregates them together. The suggested strategy from the model cannot be used to develop individual strategy but is still effective in acting as a starting point in determining overall strategy. To improve on this model, it would be ideal to divide and analyze the groups by role since the regression will produce greatly differing results for each subsection.

For instance, point guards would be expected to be reliant on threes relatively more than a center. Another observation is that statisticians now use a metric called with or without you (Keating, 2012). This metric can be used to confirm outliers such as Stephen Curry and Klay Thompson and validate the three-pointer analysis for point guards and shooting guards, since three pointers only make up a part of the role these players have. Ideally, this important metric has been constructed off individual statistics, but would be valuable if paired to the team study analysis for general strategy.

A unique way that the model's effectiveness was verified was by comparing the estimates made for your team and the opposing team. In all variates, they mirrored each other. In other words, if a significant variate indicated a positive attribution towards team success, the opposing team's variate would be a negative value indicating that them achieving more of that variate relates to reduced success. An effective model that can verify these mirroring results would be to pair up individuals in games by role, however this would be an extensive study resulting in a large data set. Nonetheless, if data can be constructed in this manner, similar modelling techniques can be applied to verify the significance of the variates seen in the used data set but for individual players which is more useful to coaches and managers from a strategic perspective.

## Works Cited

McMenamin, D. (2018, December 13). Three-point revolution? LeBron's noticed. Retrieved from http://www.espn.com/nba/story/_/id/25513076/three-point-revolution-lebron-noticed

Mather, Victor. "How the N.B.A. 3-Point Shot Went From Gimmick to Game Changer." The New York Times. January 19, 2018. https://www.nytimes.com/2016/01/21/sports/basketball/how-the-nba-3-point-shot-went-from-gimmick-to-game-changer.html.

"Klay Thompson Stats." Basketball. https://www.basketball-reference.com/players/t/thompkl01.html.

Kelepouris, Ionas. "NBA Team Game Stats from 2014 to 2018." Kaggle. April 12, 2018. https://www.kaggle.com/ionaskel/nba-games-stats-from-2014-to-2018?fbclid=IwAR10xygb3ibLHHy2zN9Q037nKCW_pP05JjsM5TVEnY1L0D-yIY7KW0mF1Ss.

Keating, Peter. "On the plus Side." ESPN. August 13, 2012. http://www.espn.com/nba/story/_/id/7551067/nba-utility-plus-minus-statistic-espn-magazine.