Importance of the Correlation between Points and Defensive Rebound*

Zheng(Zeb) Yang

April 17, 2024

In the realm of American sports, the NBA (National Basketball Association) stands out as one of the most popular leagues. This study delves into the linear relationship between the average points scored and average defensive rebounds per game among NBA teams over the past four seasons. By analyzing regular season data from the NBA's recent four seasons, this research aims to explore the correlation between team scoring and defensive rebounding. Additionally, a linear regression model will be constructed to predict this relationship. Results show that the playing defence is crucial as the average points per game increases when the defensive rebound increases. The findings will contribute to a deeper understanding of the interplay between team offense and defense in the NBA.

1 Introduction

Among the major sports leagues in the United States, the NBA has secured a prominent position, drawing substantial attention from fans worldwide. In recent years, there has been increasing interest in understanding the dynamics between team scoring and defensive performance within the league. Rebounding is one of the most important aspects of winning games. In basketball, a rebound is the ball that bounces back after an unsuccessful basket attempt. The player who subsequently catches the ball (rebounding) is awarded the rebound. If the defending team grabs it, they secured a defensive rebound (n.d.a). Offensive rebounding gives your team extra chances, and free throw opportunities, and frustrates the defense. Defensive rebounding is a key part of good defense, limiting the offense to one shot@Gels_2024. As such, this study focuses on examining the linear relationship between two crucial metrics: average points scored and average defensive rebounds per game by NBA teams over the past four seasons. By investigating these variables, we aim to uncover insights into how offensive productivity relates to defensive resilience in the NBA context. This research is pivotal to

*Code and data are available at: https://github.com/iloveyz12/NBAdefenceplay.

uncover the interactions between scoring and defensive strategies employed by NBA teams, ultimately enhancing our comprehension of the league's gameplay dynamics.

I used R(R Core Team 2023) and to further enable the analysis, I employed the use of the following packages: dataverse(Kuriwaki, Beasley, and Leeper 2023), ggplot(Wickham 2016), tidyverse(Wickham et al. 2019), arrow(Richardson et al. 2024), rstanarm(Goodrich et al. 2022), modelsummary(Arel-Bundock 2022) and here(Müller 2020).

2 Data

The data utilized in this paper focus on the defensive rebounds per game and points per game, which is accessible from the open database Basketball Reference (n.d.b). Specifically, I collected data spanning the NBA regular seasons from 2020 to 2023, capturing the average per game statistics for both defensive rebounds and points. With thirty NBA teams over the course of four years, this study encompasses a total of 120 data points.

After downloading the datasets for four seasons, I cleaned them, discarding any irrelevant data for this particular study, and combined them into one data file. The choice to acquire data from the Basketball Reference open data platform stemmed from its user-friendly interface and ease of data extraction compared to the NBA's official website. Additionally, the platform offered convenience in data storage and retrieval, facilitating a smoother research process.

3 Model

Through the data analysis conducted in this study, I uncovered a significant correlation between defensive rebounds and points scored in basketball games. This finding suggests that defensive rebounding ability may influence a team's scoring ability, thereby affecting overall game performance. To further understand this relationship and to infer the impact of defensive rebounds on scoring, I constructed a linear regression model. This model allows me to examine how changes in defensive rebounding statistics can predict variations in points scored, providing valuable insight into the dynamics of basketball games.

3.1 Model set-up

Define Y as the points scored per game by a team. Then $\text{Defensive Rebound}_i$ is the defensive rebound of each team per game. β_0 is the intercept, which indicates the points when defensive rebounds are 0, and β_1 represent the slope indicating the change in points for each additional defensive rebound. ϵ is the error term, representing the unexplained random variation in the model.

$$Y_i = \beta_0 + \beta_1 \text{Defensive Rebound}_i + \epsilon \tag{1}$$

In the model, the independent variable is the defensive rebound per game. The dependent variable is the points of each team per game.

3.1.1 Model justification

I anticipate a linear relationship between defensive rebounds and points scored in basketball games. This expectation arises from the nature of defensive rebounds, which occur when the opposing team fails to complete their offensive possession. Securing a defensive rebound signifies a defensive stop and provides the defending team with an opportunity to transition into offense, potentially resulting in higher scoring opportunities. Therefore, I hypothesize that as the number of defensive rebounds increases, points scored will also increase linearly.

4 Results

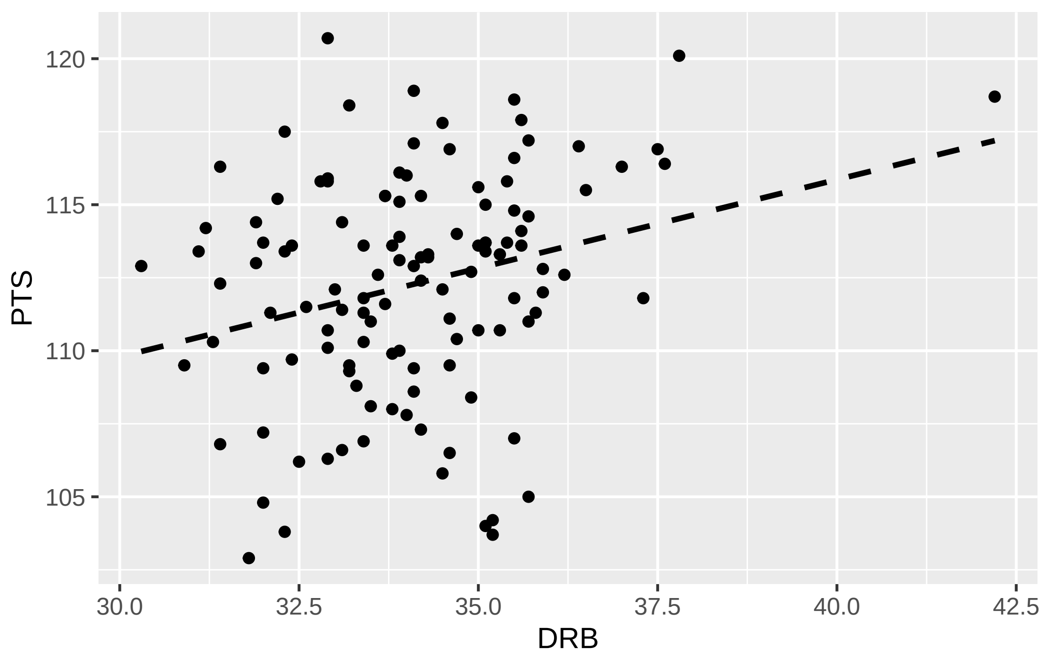

Figure 1: The distribution of points per game, by defensive rebounds per game

My results are summarized in Figure 1. Figure 1 illustrates the correlation between defensive rebounds per game and points scored per game in NBA basketball games. The scatter plot

Table 1: Model summary of the points per game, by defensive rebounds per game by each team

	(1)
(Intercept)	91.591
	(6.822)
DRB	0.607
	(0.200)
Num.Obs.	120
R2	0.072
R2 Adj.	0.065
AIC	660.2
BIC	668.6
Log.Lik.	−327.119
RMSE	3.70

portrays the general trend of defensive rebounding and scoring as the game progresses, further accentuated by the regression line. The model suggests that as the number of defensive rebounds of each team securing per game increases, there is a linear growth in points scored.

Table 1 shows the coefficients of the predictor variables of my model. The coefficient for DRB (defensive rebounds) suggests that for each increase of defensive rebounds, the predicted value of the dependent variable, points scored per game, increases by approximately 0.607, with a standard error of 0.200, reflecting the uncertainty in the estimated coefficient.

5 Discussion

References

n.d.b. *Basketball*. https://www.basketball-reference.com/.

———. n.d.a. *Owayo Custom Sports*. https://www.owayo.com/magazine/rebounds-in-basketball-us.htm.

Arel-Bundock, Vincent. 2022. "modelsummary: Data and Model Summaries in R." *Journal of Statistical Software* 103 (1): 1–23. https://doi.org/10.18637/jss.v103.i01.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "Rstanarm: Bayesian Applied Regression Modeling via Stan." https://mc-stan.org/rstanarm/.

Kuriwaki, Shiro, Will Beasley, and Thomas J. Leeper. 2023. *Dataverse: R Client for Dataverse 4+ Repositories*.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Richardson, Neal, Ian Cook, Nic Crane, Dewey Dunnington, Romain François, Jonathan Keane, Dragoș Moldovan-Grünfeld, Jeroen Ooms, Jacob Wujciak-Jens, and Apache Arrow. 2024. *Arrow: Integration to 'Apache' 'Arrow'*. https://CRAN.R-project.org/package=arrow.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.