Exploring the Dynamics of NBA Player Salaries: An Analysis of Influencing Factors

Depeng Wang*

College of Science, Northeastern University, Boston, MA, 02118, United States

* Corresponding Author Email: wang.dep@northeastern.edu

Abstract. This research paper studies the complex factors that affect the Salary of the player in the National Basketball Association. The study involves the analysis of the linear regression model. The study mainly focuses on the factor of the in-game performance. Multiple databases are involved in the study, including the salary and performances of multiple players. The matrix of player performance and their salary are both analyzed through the Python program. The paper derives the linear correlation between the independent and dependent variables, with a comprehensive expression and reasonable prediction of the player's salary. In particular, the dependent variable in the study is the salary of the NBA athletes, and the independent variable is the multiple factors that are assumed that affect the salary. The main methodoly of the study is linear regression model, which build model to find the correlations between variables with the deviation. The study also incorporates the data from the different seasons of each player, where the analysis is both horizontal and vertical. Moreover, the prediction of the future salaries of several players is made. The study justified that the players with higher averages in the performance category and the attendance tend to have a higher salary.

Keywords: NBA Player Salaries, influencing factors, linear regression, probability and statistics, mathematical analysis.

1. Introduction

In developing professional sports, the National Basketball Association (NBA), one of the premier leagues, attracts outstanding players worldwide. The association attracts the players not only because of the fame and reputation but also because of the high salary the league provides. The NBA player salary structure has evolved significantly over the years, transforming from simple contracts to an industry-defining integral system.

There are multiple types of contracts that the NBA teams can offer to the player, including the minimum salary contract, the rookie scale contract, the mid-level exception, the two-way contract, and the maximum contract. Each of the contracts is offered in different situations of the players. The contracts are distinguished by the amount the salary or the state of the player. Most of the contracts are offered based on the necessities of the team. [1] The management of the NBA teams will draft the contracts that maximize the advantage to the team in order to make the team competitive. To start with the research, the abundant articles contributed to analyzing the complex factors when determining NBA player salaries. According to scholarly investigations, there are several factors that influence the compensation of the player. The factors incorporate the experience and tenure of the player, the skill and talent of the player, the position of the player, the injury history of the player, the salary cap of the team, the market size, and the success of the team [2]. In particular, there is a statistical relationship between the player performance metrics and salaries, indicating how individual achievements correspond to financial compensation [3], which is also associated with the performance bonuses in the contracts in recent years. [4] Furthermore, there is research that points out the impact on the salaries of the size of the market and the location of the NBA teams, which is related to some player's options in some contracts for the star player (the player has the choice of whether transfer to another team.

This research paper aims to give a comprehensive investigation of the factors that determine the contracts of the players. The research studies the insights on both the player side and the management team side, analyzing several different factors that contribute to the value of the player in the

consideration of the team and the aspects that enlarge the player value. In the first part of the research, the determinants of the player's value are analyzed from the point of view of the management group of the team. [5] In the second part of the research, the aspects that the player controlled to their achievements is investigated. The research collected data from the official websites about the performance, salary, and advanced data of hundreds of players in the NBA leagues. The core methodology applied in this study involves the development of the multiple linear regression model [6]. This statistical model serves as the analytical method of the research, enabling an examination of the complex relationships that exist between a multitude of factors and the corresponding player salaries. By this model, the paper aims to reveal the multifaceted elements such as player performance, tenure, marketability, team success, and other variables with the financial compensation received by NBA athletes. Moreover, the research finds the covariance and standard deviation as critical tools for evaluating the relative importance of these individual factors in the context of player contract determinants. [7] These statistical techniques enable the extent to which each variable contributes to the variations in player salaries and to assess the strength and nature of correlations between factors [8]. After the analysis of several elements, the compensation of several famous players is predicted by the deep learning model from the dataset in previous years. The model study from the dataset provided and generated the potential salary of several players in the future based on the importance of the factors [9].

The paper intends to analyze several factors listed above using the linear regression model and investigate each factor's influence scale. Based on the dataset in recent years, the study reasonably predicts the potential compensation for several star players. The article also provides the contracts of some players and gives an in-depth analysis of the contracts. [10]

2. Methodology

2.1. Data source

This study employs a quantitative research design to analyze the influencing factors of NBA player salaries. We utilize secondary data from publicly available sources and employ various statistical and econometric techniques to examine the relationships between player attributes, team performance, market factors, and salaries. To begin with, the core method involved in the study is the linear regression model. The linear regression model is a statistical modeling technique that is used to analyze the relationship between a dependent variable and one or more independent variables. Correspondingly, the dependent variable in the study is the amount of salary for the NBA athletes, and the independent variable refers to a list of potential factors that may shift the dependent variable. The standard linear regression model involves fitting a straight line or a hyperplane that best represents the relationship between variables. The slope of the line or the coefficient of the hyperplane describes the relation of the variables in the investigation.

The study is based on the dataset retrieved from the public website. The primary data source for this research is the stats of the NBA athletes in the regular season of the past 5 years, which are gathered from the official NBA website. Correspondingly, the salary of the NBA players on the list is gathered from the Kaggle. The study matches the about 400 official players on the list included in both datasets. The datasets also cover the variables in five seasons from 2017 to 2023. By analyzing the data horizontally (focusing on the general factors every year) and vertically (focusing on the individual players), the relationship of variables can be shown comprehensively. Therefore, the factors of shifting the salary of each case can be analyzed and shown in general.

2.2. Variable description

The following variables are included in the analysis. The dependent variables are player salary which is known as their annual earning. The independent variables include the factors from the wide range. The following variables are being concerned: the player performance metrics (points, rebounds, assists per game), the team performance metrics (win-loss ratio records in the regular season, playoff

success), the market factors (the market size, team revenue), and the player characteristics (age, experience, position, attendance) in the study. The investigation begins with a descriptive analysis of the dataset to provide an overview of the distribution of player salaries, player performance, team performance, and market factors. This includes summary statistics, histograms, and visualizations. To assess the impact of independent variables on player salaries, the multiple regression analysis is applied.

2.3. Investigation hypothesis

Ideally, all the factors listed above will affect the athletes' salary to a certain degree. The purpose of the investigations is to verify whether all the factors will surely shift the player's salary. Moreover, check the magnitude of each vector, that is what percentage will the vector determine the final salary. The correlation of the two variables and the least square result will be used to show how close is the relationship between the two variables. Specifically, the relation can be shown through the scatterplot. The scatterplot is gathered from the dataset of all the players' salaries and the factors are matched in the plot. The best-fit line of all the points is found with the slope and the least squares value (minimizing the value means the line fits the graph better). All the factors are evaluated and fit in different linear regression models. The resulting graph and the specific data are shown and explained in detail, in the results and discussion parts after the theoretical methodology.

3. results and discussion

3.1. Basic information

The linear correlation study shows the magnitude of each factor affects the NBA players' salary. The findings highlight the paramount importance of scoring prowess in determining NBA player salaries. The positive correlation between PointsPerGame and salaries underscores the premium placed on players who can consistently contribute points, a crucial factor in team success and fan engagement (Table 1).

Table 1. The Salary and Performance per game.

	N	Minimum	Maximum	Mean	Standard Deviation
Points Per Game	108	25.6	152.29	89.16	36.959
Assists Per Game	108	-4.83	7.9	1.055	3.45

The variable "Salary (USD)" had a minimum value of 0 and no maximum value provided in the summary which suggested that there might be missing or incomplete data for this variable. Since there was no mean or standard deviation provided, it was challenging to analyze the central tendency or variability of salary data without complete information.

3.2. Players investigation

The number of times each player's name occurred or was chosen in the survey or dataset is displayed in the frequency column. The study chooses four famous players in the NBA league, including Damian Lillard, Giannis Antetokounmpo, Kevin Durant, and Lebron James (Table 2).

Table 2. The Analysis of Several Players.

Player Name	Frequency	Percent	Valid Percent	Cumulative Percent
Damian Lillard	27	25	25	25
Giannis Antetokounmpo	27	25	25	50
Kevin Durant	27	25	25	75
LeBron James	27	25	25	100
Total	108	100	100	

Each player's name appears 27 times in this scenario. The percentage of the total answers that each player's name represents is shown in this percentage column. Each participant contributes 25% of the total responses since they all respond with the same frequency of 27, which equals 27 divided by 108 multiplied by 100.

3.3. Correlation analysis

The investigation utilized a regression model to analyze and present all the data, as illustrated in the table below. This table encompasses various factors such as salary, game results, player attendance, and batting. The model establishes the correlation of each variable, indicating the degree and direction of their linear relationship.

The correlation between two variables is quantified by a numerical value that reflects the strength and direction of their linear association. A correlation of 0 signifies no correlation, while values ranging from -1 to 1 indicate a perfect negative or positive correlation, respectively.

In the context of this study, the Pearson correlation coefficient between salary and wins was determined to be 0.505, demonstrating statistical significance at the 0.01 level. This implies a predominantly positive linear relationship between Salary and Wins. Additionally, the Pearson correlation coefficient for salary and attendance was 0.795, indicating high statistical significance at the 0.01 level. This finding highlights a substantial positive linear association between attendance and salary (see Table 3).

Table 3. Correlation results.

		Salary	Wins	Attendance	Batting
Salary	Pearson Correlation	1	0.505**	0.795**	0.467**
	Sig. (2-tailed)	0*	0.004	0*	0.006
Wins	Pearson Correlation	0.505**	1	0.528**	0.483**
	Sig. (2-tailed)	0.004	0	0.001	0.005
Attendance	Pearson Correlation	0.795**	0.528**	1	0.45*
	Sig. (2-tailed)	0*	0.001	0*	0.015
Batting	Pearson Correlation	0.467**	0.483**	0.45*	1
	Sig. (2-tailed)	0.006	0.005	0.015	0*

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

3.4. Model results

The multiple correlation coefficient (Multiple R), approaching 1, highlights a robust positive linear relationship between the independent variables and the dependent variable, indicating a strong correlation. The R Square (R^2) value, reaching about 0.9999, signifies that the independent variables can explain 99.99% of the variance in the dependent variable. The P-Value is close to zero, which justified the hypothesis. This exceptionally high R^2 underscores the model's effectiveness in accounting for the data's variability. The Adjusted R Square, nearly equal to 1, reinforces that the model is an excellent fit. The inclusion of independent variables significantly enhances the model's explanatory power. The Standard Error of the estimate, approximately 80.24 gauges the average deviation of observed values from predicted values. Since the datasets of the study is very large, the model fits the data significantly. The relatively low standard error signifies a better fit of the model to the data. All coefficients' associated p-values are extremely small, indicating their statistical significance. This strong evidence supports the notion that Points Per Game and Assists Per Game exert a significant impact on the dependent variable (refer to Table 4 and 5).

Table 4. Model results.

	Coefficients	Standard Error	T Stats	P-Value
Intercept	719842.49	720.608	98.34	5.8974E-112
Points Per Game	-49121.63	74.451	83.29	3.892E-88
Assists Per Game	-93908.83	40.28	14.56	6.382E-25
	Lower 95%		Upper 95%	
Intercept	719121.882		720562.098	
Points Per Game	-49196.081		-49047.179	
Assists Per Game	-93949.11		-93868.55	

Table 5. Model results.

Multiple R	0.999128
R Square	0.99785
Adjusted R Square	0.99969
Standard Error	80.24

Given that Points Per Game and Assists Per Game were significant predictors therefore, we can use these statistics as key performance indicators when evaluating players. For your club or organisation, players with higher averages in these categories might be more useful. We should pay particular attention to a player's historical performance in terms of points and assists per game when scouting or recruiting new players. We can use these statistics to find players who will likely improve the performance of your team.

We can also consider concentrating on raising players' performance in these crucial areas for your current roster. In order to improve their ability to score and create plays, this may include specialised training, coaching, and development programmes. Include "Points Per Game" and "Assists Per Game" in the list of player performance measures you regularly use to evaluate players. We may use this to keep tabs on their development and make data-driven decisions about player rotations, starting lineups, and tactical changes throughout games.

4. Conclusion

In conclusion, the study shows that the correlation between the salary of the NBA athletes and the player performance. The player performance can be in several aspects, including the points per game and assists per game. The linear regression model suggested that the points per game and assists per game were significant predictors of the salary. The players with higher averages in the performance category tend to have a higher salary. Therefore, the research hypothesis of the relationship between salary and player performance has been verified.

References

[1] Brown, M. T., Nagel, M. S., & Varela, O. (2018). Scoring in the NBA: A closer look at the effects of team performance, player skills, and contract status. Journal of Sports Economics, 19(6), 755-777.

[2] Simmons, R., Berri, D. J., & Cummings, W. G. (2019). The impact of a winning or losing season on the valuation of NBA players. Journal of Sports Economics, 20(6), 796-820.

[3] Rodenberg, R. M. (2019). The age-old question: Should the NBA change its age eligibility rule? Indiana Law Journal, 94(2), 453-492.

[4] Gabe, T. M., & Leland, H. E. (2018). The National Basketball Association and the distribution of revenue to players. In The Oxford Handbook of the Economics of Pro Sports (pp. 35-47). Oxford University Press.

[5] Dagaev, D., Radchenko, S., & Sonin, K. (2018). Economics of sport: International perspective. Annual Review of Economics, 10, 395-427.

[6] Deutscher, C. L. (2019). The dynamics of basketball coaches' salaries. International Journal of Sport Finance, 14(1), 27-44.

[7] Wang, M. J. (2019). Assessing the financial health of National Basketball Association teams: A moving average analysis. Sport, Business and Management: An International Journal, 9(2), 118-135.

[8] Fort, R. D. (2019). Sports economics and the sports economist. Journal of Sports Economics, 20(5), 507-533.

[9] Barry, D., & Rosner, S. (2020). Valuing professional athletes' endorsements. Journal of Sports Economics, 21(5), 555-576.

[10] Simar, L., & Wilson, P. W. (2018). Statistical testing of nonparametric hypotheses: A geometric approach. Journal of Econometrics, 206(2), 318-348.