Research on the influencing factors of the value of NBA players

Chengjie Shi

North Cross School Shanghai, Shanghai, 200000, China

michaelshicj@outlook.com

Abstract. NBA players are among the highest-paid athletes globally. Salaries can range from league minimums of around $1 million per year for new or less experienced players to over $40 million annually for superstars. In addition to salaries, players can earn significant amounts from endorsements and other business ventures. Researching NBA players' salaries is important for several reasons: Understanding how player salaries affect the overall economy of the sport; Insights into how talent is valued and compensated in a highly competitive industry; Teams must strategically manage their payroll to stay within salary cap limits while building competitive rosters. This article will use multiple linear regression model to analyze some factors such as turnover, age, points, personal fault, block, total rebound, assist, and effective Field Goal percentage that relate to NBA player salaries from 2022-23 season. Model shows the model of R value of 0.485. All factor (turnover, age, points, personal fault, block, total rebound, assist, effective Field Goal percentage) can explain 48.5% of the Salary change. In addition, Age and Points Per Game (PTS) have a significant positive impact on a player's salary. This indicates that both a player's age and their ability to score points are crucial determinants of their earning potential in the NBA.

Keywords: Multiple linear regression model, NBA players, salary.

1. Introduction

NBA players' salaries have seen significant growth over the years due to several factors, including lucrative television deals, merchandise sales, and overall popularity of the league. The collective bargaining Agreement (CBA) between the NBA and the Players' Association (NBPA) sets out the structure of salaries, including the salary cap, maximum and minimum salaries for players, and the length of contracts. Superstar players can command large contracts, often including endorsements, while rookie contracts and veteran minimums provide a salary structure for new and seasoned players.

The increasing global appeal of the NBA has also contributed to rising salaries and team revenues. According to ESPN expert John Hollinger's player efficiency formula, Shen concluded that the overall ability of NBA players has a strong positive relationship with their salary. player comprehensive ability according to the expert John Hollinger's player efficiency calculation formula are players can improve the comprehensive ability value calculation formula stated that the overall ability of NBA players has a strong positive relationship with their salary [1].

In addition, Wen calculated the amount of marginal revenue contribution to the team, with the player's salary comparative analysis it is concluded that the team wins the marginal revenue for

$1404000, came to the conclusion too low to pay most of the players in the league is Rudy Gobert. Extra pay most of the players is Chandler Parsons. From 2016-2017, all players average excess to pay 2164862, including the free agent is too low to pay $998716 on average, average excess free agent to pay 2875070 dollars [2]. Liu used random forest regression model and XGBoost regression model empirically analyze the impact of player performance on their salary. The results of this combination model show that the most important factors affecting the salary of players are the average playing time and ball age per game, whose importance ratio is more than 10%, followed by the number of starts, steals, points, rebounds and shots, whose importance ratio is more than 5%, and other performance variables have less influence [3].

Liu et al. used quantitative analysis methods (including least square method, White test, weighted least square method, stepwise regression, etc.) to analyze the data from the NBA official website and Hupu.com in the 2010-2011 season, and came to the final conclusion as follows: Although the compensation of players is linked to their on-field performance, that is, the better their on-field performance, the higher their compensation, not all the factors measuring the overall performance of players will have a significant impact on the compensation of players; In the context of league rules and limited funds, older players are often paid more than new players; In general, the compensation of players in teams with good financial management is higher than that in teams with relatively poor financial management [4].

Gou and Zhang selected the NBA league 2009/2010-2018/2019, 997 players as the research object, by constructing mixed linear model, it is concluded that the influence of NBA players pay factors are various, including nationality, educational background, age, body shape, draft pick, game experience and competitive performance (playing time, the efficiency value) It has a profound effect on the players' compensation. And NBA players' winning percentage and playoff status have no significant effect on their compensation, NBA player compensation is also affected by seasons [5].

Li and Yang, according to the classical management theory that competition theory and the theory of relative exploitation, the players wages and individual performance pay gap was significantly positively related to the hypothesis and team and individual performance are positively related hypothesis [6]. Lei and Wu used Pearson product moment correlation, Spearman 's rank order data analysis, players' income from 2002 to 2009, 8 of the season, the U.S. professional baseball major au salaries in the top 10 teams, with 60% of the team record also ranked in the top 10. 2. From 2002 to 2009, the relationship between the total salary of MLB teams and wins was significantly positive, except for 2008.3. From 2002 to 2009, the relationship between the salary ranking of MLB teams and the record ranking of the year was also significantly positive [7].

By using descriptive statistics, simple correlation analysis, linear regression analysis and other mathematical statistical methods, Chen Sitong, Cheng and Chai found that There was no relationship between a team's overall salary and its performance. Five indicators, including average points per game, average rebounds per game, average assists per game, average steals per game and the number of appearances, are important indicators to evaluate a player's salary [8]. Wang took the player's salary as the response variable, and the independent variable includes 21 variables such as the average score, the number of rebounds and the number of assists in a season. The purpose is to establish a regression model and select the independent variables that have a decisive effect on the wage level. This paper will compare the methods of least squares, ridge regression, Lasso, adaptive Lasso, smoothly truncated absolute deviation (SCAD), and Elastic Net. Through data analysis, it is found that Elastic Net has the smallest prediction error [9]. Wang used the method of literature and mathematical statistics to analyze the salary and competitive contribution of NBA players in 2016-2017 season as the research object, and concluded that as the gap between player salary and competitive contribution increased. There was a strong positive relationship between salary and competitive contribution, along with a notable negative relationship between salary and cost performance. Therefore, there was a dissociation between player price and competitive value [10].

In summary, the research on the important data of players' salary has attracted numerous scholar. This article will mainly use multiple linear regression model to analyze the salary of NBA players.

2. Methodology

2.1. Data source

The dataset titled "National Basketball Association Player Salaries (2022-23 Season)" was sourced from Kaggle. This dataset combines per-game and advanced statistics of NBA players from the 2022-23 season with their salary data, providing a thorough resource for analyzing both performance and financial aspects of professional basketball players. The dataset was created by web scraping player salary information from Hoopshype and downloading traditional per-game statistics.

The variables are described in Table 1. Table 1 shows that the maximum value for one AST item is three standard deviations above the mean, indicating significant data variability. This suggests that using the median is more appropriate for describing the overall level than the mean. The conclusion is that the maximum value (minimum/maximum value) of one AST data item exceeds the mean by three standard deviations, reinforcing the idea that the median is a better measure for capturing the central tendency given the high data fluctuation.

Table 1. List of variables.

name	Sample	min	maxi	mean	SD	median
Salary	100	12939848	48070014	26047328	10148315	23090000
Age	100	22	38	28.81	3.631	28
PF	100	0.4	3.8	2.2	0.647	2.15
PTS	100	5	33.1	17.356	7.298	16.8
TOV	100	0.4	4.1	1.944	0.881	1.9
AST	100	0.5	10.7	3.969	2.199	3.9
TRB	100	1.5	12.5	5.376	2.642	4.55
eFG%	100	0.443	0.671	0.552	0.048	0.555
BLK	100	0	2.5	0.556	0.461	0.4

2.2. Method introduction

Multiple Linear Regression (MLR) is a statistical approach used to analyze the relationship between an outcome factor and multiple predictor variables. The formula for MLR is:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_n X_n + \epsilon \tag{1}$$

Where Y refer to the outcome factor, β_0 refer to the y-intercept, β_1 refer to the coefficients for each independent variable $X_1, X_2, \cdots, X_n$, ϵ refer to the error tern, accounting for the variance in Y not explained by the independent variables.

3. Results and analysis

3.1. Model results

Table 2 shows that the Age, PF, PTS, TOV, AST, TRB, eFG %, using the series as the independent variables and Salary as the outcome factor, the linear regression analysis, as depicted in the chart, results in the following model formula:

$$Salary = -15177957.277 + 982084.535 * Age + \cdots + 1309858.527 \, x \, series \tag{2}$$

The R-squared value of the model is 0.485, mean Age, PF, PTS, TOV, AST, TRB, eFG %, series can explain 48.5% of the Salary diversification. An F-test was performed on the model, which determined that the model, according to the F-test,(F = 10.714, p = 0.000 < 0.05), and explains namely Age, PF, PTS, TOV, AST, TRB, eFG %, a series of at least one affect Salary relations, in addition, in view of the model, found that the multicollinearity test The VIF value in the model is greater than 5, but less than 10, which means that there may be some collinearity problem, which can be solved by ridge regression or stepwise regression. The final concrete analysis shows that:

The regression coefficient for Age is 982,084.535 (t=4.207, p=0.000<0.01), indicating a significant positive effect of Age on Salary. In contrast, the coefficient for PF is 66,916.795 (t=0.044, p=0.965>0.05), suggesting that PF does not impact Salary. The regression coefficient for PTS is 518,120.206 (t=2.935, p=0.004<0.01), demonstrating a significant positive influence on Salary. However, the coefficient for TOV is 2,143,419.190 (t=1.046, p=0.298>0.05), showing no effect on Salary. Similarly, AST has a coefficient of 507,980.458 (t=0.847, p=0.399>0.05), indicating no impact on Salary. Lastly, the coefficient for TRB is 325,469.340 (t=0.739, p=0.462>0.05), which also suggests no influence on Salary.

Table 2. Results of the linear regression analysis

	Non-standardized coefficients		Standardization coefficient	t	p	collinearity diagnostics	
	B	standard error	Beta			VIF	tolerance
constant	-15177957.277	13432127.489	-	-1.130	0.261	-	-
Age	982084.535	233458.143	0.351	4.207	0.000**	1.233	0.811
PF	66916.795	1521833.244	0.004	0.044	0.965	1.663	0.601
PTS	518120.206	176507.560	0.373	2.935	0.004**	2.847	0.351
TOV	2143419.190	2048259.805	0.186	1.046	0.298	5.582	0.179
AST	507980.458	599835.669	0.110	0.847	0.399	2.984	0.335
TRB	325469.340	440128.049	0.085	0.739	0.462	2.320	0.431
eFG%	-8828493.461	19803210.321	-0.042	-0.446	0.657	1.532	0.653
BLK	1309858.527	2357475.637	0.059	0.556	0.580	2.025	0.494
R^2	0.485						
Adj R^2	0.440						
F	F (8,91)=10.714,p=0.000						
D-W value	0.970						

Dependent variable:Salary

* p<0.05 ** p<0.01

3.2. Model test

Table 3 shows that the Age, PF, PTS, TOV, AST, TRB, eFG %, using the series as the independent variables and Salary as the dependent variable, the linear regression analysis shows that the model has an R-squared value of 0.485, as depicted in the chart, Mean Age, PF, PTS, TOV, AST, TRB, eFG %, series can explain 48.5% of the Salary change.

Table 3. Model test

R	R^2	Adj R^2	RMSE	DW value	AIC value	BIC value
0.696	0.485	0.440	7245957.394	0.970	3460.979	3484.425

4. Conclusion

Using both the multiple linear regression approach and factor analysis technique, the author conducted a quantitative analysis of the factors influencing NBA salaries. According to the conclusions drawn from the model, Age and Points Per Game (PTS) have a significant positive impact on a player's salary. This indicates that both a player's age and their ability to score points are crucial determinants of their earning potential in the NBA.

Conversely, other factors such as Personal Fouls (PF), Turnovers (TOV), Assists (AST), Total Rebounds (TRB), and Effective Field Goal Percentage (eFG%) do not have a significant impact on the salary. This finding suggests that while these metrics are important for evaluating a player's overall

performance and contribution to the game, they are not major determinants of salary in comparison to scoring ability and age.

Therefore, when measuring the value of an NBA player, it is essential to prioritize the player's scoring ability. Players with higher scoring averages tend to command higher salaries. This emphasis on scoring ability aligns with the entertainment value that high-scoring players bring to the game, which in turn drives revenue and fan engagement for teams.

For players looking to maximize their salary potential, it is advisable to focus on improving their scoring skills. By enhancing their ability to score, players can increase their market value and secure more lucrative contracts. Teams, on the other hand, should consider these findings when negotiating salaries and structuring their rosters, ensuring they allocate appropriate financial resources to players who can consistently deliver high scoring performances.

In summary, the key takeaway from the analysis is that scoring ability is the most significant factor in determining NBA salaries. Both players and teams should consider this when making decisions related to contracts and player development. By focusing on enhancing scoring abilities, players can maximize their earnings, and teams can ensure they are investing wisely in talent that will deliver the greatest impact on the court.

References

[1] Shen C 2016 Research in the impacting factors of NBA player salaries. The quotient, 26, 46.

[2] Wen H 2018 An Analysis of NBA player salary and performance. Jiaotong university sports journal, 14, 45-56.

[3] Liu Y 2019 The Impact of basketball player performance on player salary: an empirical analysis based on NBA player data. Economic and Social Development Research, 18, 253-254.

[4] Liu F F, Liu Y and Wei S K 2024 Nanjing University of Finance and Economics. An empirical analysis of the influencing factors of NBA players' compensation: A case study of 2010-2011 NBA season. Working paper.

[5] Gou H C and Zhang H 2022 The NBA player salaries impact factor analysis. Sports scientific research, 43(2), 62-69.

[6] Ching L and Yang Z H 2009 NBA players of the relationship between pay and performance. Journal of Shanghai Institute of Physical Education, 33(1), 6.

[7] Lei W G and Jing Y W 2010 The major au team wages and record the related research. Journal of Sports, Health and Leisure, 9(2), 14-25.

[8] Chen S T, Cheng C and Chai Y J 2015 Research on the correlation between salary and team achievement and player performance. Sports, 19, 3.

[9] Wang H H 2018 Statistical analysis of NBA players' salary. Research the World, 3, 6.

[10] Wang Y 2021 Discussion on the relationship between NBA players' salary and competitive contribution. Fujian Sports Science and Technology, 40(2), 20-25.