Model Prediction of Factors Influencing NBA Players' Salaries Based on Multiple Linear Regression

Yang Zhao

School of Mathematics and Statistics, Southwest University, Chongqing, China, 400715
**Corresponding author. Email: E-mail: yangzh2633@163.com*

ABSTRACT

With the commercialization of sports, the main concern of the various leagues is how much to pay a player, especially a star player or superstar. This paper uses the NBA league as an example to investigate the main factors affecting players' salary levels. 14 indicators such as height, weight and Win Share of players who played more than or equal to 10 times in the NBA league in the 2017-2018 season were selected. The aim is to build a regression model to select the independent variables that have a significant effect on the salary level. The result shows that NBA players' individual abilities have a strong positive relationship with their salaries and their team-related indicators do not have a significant impact on their salaries. Finally, based on the NBA players' salary model, some suggestions on salary management were provided to the Chinese Basketball Association.

Keywords: *NBA, professional player, salary management, Multiple linear regression*

1. INTRODUCTION

Since its inception in 1995, the Chinese Basketball Association has made good progress, but at the same time the league has revealed many problems in the management of players' salaries. The NBA basketball league is the most market-oriented basketball league in the world. The analysis of the current situation of NBA players' salaries may bring some inspiration to the future development of CBA.

The NBA basketball league in the USA is the most market-oriented basketball league in the world. The same problem has occurred in the NBA, even causing the 2011 NBA lockout. Firstly, the author reviews the relevant research in this research field. The literature from different research angles lays a theoretical foundation for the author's analysis. This paper selects NBA players' salary as the research object and analyzes it according to 14 indicators of players. The multiple linear regression and ANOVA methods were used to establish the linear regression relationship between players' salaries and these 14 variables, and the variables that had a more significant impact on salary were identified through the analysis.

The analyses of this paper provides a reference for learning from the NBA's advanced experience in player salary management, which will improve the healthy and orderly development of CBA league.

2. LITERATURE REVIEW

Lin H.and Y,Yang [1] examined the length of contracts and player asset specificity in MLB, dividing uncertainty into intrinsic and extrinsic, and concluded that the length of a player's contract decreases as uncertainty increases, and that longer contracts help players share risk with their teams.

Silvia Grdinaru [2] uses the example of a basketball team in Romagna to study the impact of motivation on management and to analyse the most important key factors influencing the performance of the basketball team. The results of the study concluded that goal setting is influenced by financial and human resource factors as well as control and supervision, and that motivation plays a key role in the performance of athletes. Athletes' performances play a key role.

Bernd Frick [3], in his analysis of data from 29 seasons in the Bundesliga, examines player salaries and the length of players' contracts. The study concluded that moral hazard is a widespread phenomenon, even in professional football, and that player wage differences can be explained to a large extent by differences in individual performance, with differences in player

Y. Jiang et al. (Eds.): ICEDBC 2022, AEBMR 225, pp. 1439–1445, 2022.
https://doi.org/10.2991/978-94-6463-036-7_213

performance significantly reduced in the final year of the contract.

Mark R. Frascatore [4]analyzes and proves that the strategy of offering amazing long-term contracts to star players can reduce the wage bill of existing leagues.

Scott Bukstein [5], by analyzing the current NBA salary arbitration system and the salary systems of other leagues such as the NFL, suggests that the NBA's use of salary arbitration is limiting player freedom and, therefore, a new salary system is needed.

Johnny Ducking [6], Groothuis PA, Hill J.R. argue that there is no significant statistical evidence that raising the minimum wage for veteran players negatively affects their careers in the NBA by analyzing the impact of multiple data on NBA players.

Michael [7] used least squares to compare and analyze player data and wages using race as a differentiating factor, and concluded that skin colour did not have a significant effect on player wages at higher levels of scoring and rebounding.

Berri [8] conducted an empirical analysis of player salaries in the NBA and found that a player's scoring ability plays a major role in player evaluation, which makes it possible for players to make decisions on the court that are inconsistent with the team's goals.

Kevin [9] used a salary contract-related analysis to find that players perform well in the final year of their contracts.

Shen Chen [10] collected the data of NBA 2014-2015, and established a multiple linear regression model of NBA players' salaries with players' overall ability value, age, attendance rate, team winning percentage, and development prospect as independent variables, and found that NBA players' overall ability has a strong positive relationship with their salary, and players' development potential has a negative relationship with their salary.

Chen S.T., Cheng Chen and Chai, Y.J [11] regressed player data and player salary by the OLS method and used correlation coefficients to detect the correlation between the variables and concluded that there was a significant correlation between rebounding average and scoring average and player salary when exploring player salary and player related data indicators.

Jin S.Y. [12] found that there is a large gap between players' salaries. The total amount of salaries paid by clubs is not exactly proportional to their performance. There is a relationship between players' salary levels and their playing age.

Zhang Chenying and Li Xiliang [13] argue that the salary cap system has a greater effect on promoting competition than its anti-competitive consequences, which to a certain extent provides a stronger endorsement for the legality of the salary cap system and also provides a model for professional sports leagues in other countries to follow. The salary cap system of the CBA has to a certain extent borrowed from the salary cap system of the US basketball league, but the specific rules are far more restrictive of competition than the latter, constituting a fixed price practice and violating the provisions of the Anti-Monopoly Law. In the future, the CBA league should further improve the structure of the salary cap system and devote itself to designing a new scheme that does not violate the Anti-Monopoly Law but also promotes balanced competition, so as to promote a healthy and orderly transformation of Chinese basketball towards commercialization.

3. METHODOLOGY

3.1. Source of data and description of variables

The player data used in this article was sourced from the basketball-reference [14] website and was selected from 14 data items (as shown in table1), including height and weight, for players who started more than or equal to 10 times in the 2017-2018 NBA season.

Player Efficiency Rating(PER): ESPN expert John Hollinger's data on the use of in-game stats such as shooting percentage and assists to assess a player's game performance

Box Plus/Minus(BPM): Data presented by Daniel Myers for estimating the contribution of basketball players to the team while they are on the court.

Value Over Replacement Player(VORP): the difference between a player's BPM score and the BPM score of the player's theoretical "replacement player" as suggested by Tom Tango.

Win Share(WS): A player statistic that attempts to assign credit for the team's success to individuals within the team the sum of player win shares on a given team will be roughly equal to that team's win total for the season.

Table1. Variable description

Variable	Description Of Details	Range
Salary(Y)	Unit: $million	1.07 ~ 34.68
Height (X1)	Unit: cm	175 ~ 221
Weight (X2)	Unit: kg	73 ~ 131

Age(X3)	Unit: year	19 ~ 41
Player's playing age (X4)	Unit: year	0 ~ 18
Players' playing time per game(X5)	Unit: minute	3.29 ~ 37.15
Player Efficiency Rating(X6)	Unit: percentage	4.5 ~ 30.2
Win Share(X7)	Unit: Score	-0.7 ~ 15
Box Plus/Minus(X8)	Unit: Score	-11.8 ~ 11.1
Value Over Replacement Player(X9)	Unit: Score	-1.3 ~ 8.6
Team Winning Percentage (X10)	Categorical variables	High, medium, low
Player's Continent (X11)	Categorical variables	Asia, Africa, Europe, Oceania, Latin and South America, USA&CAN
Position(X12)	Categorical variables	C,PF,PG,SF,SG
Market value of the team (X13)	Categorical variables	High, Medium, Low
Whether he have been selected for the All-Star team (X14)	Categorical variables	Yes,No

3.2. Descriptive statistical analysis

To obtain an overall impression of the data distribution, descriptive statistical analysis was performed on the data. The scatter plots of X1, X2 and player salaries are shown in figure 1 and the box plots of X1, X2 and player salaries are shown in figure 2.

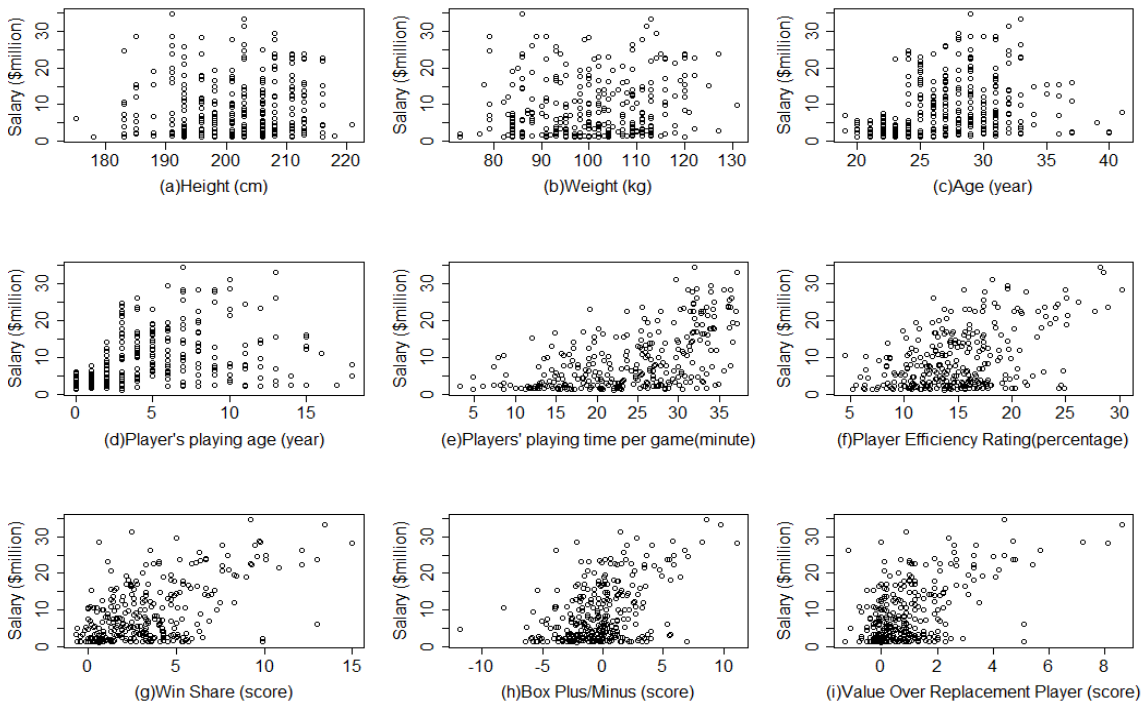

Figure 1 Scatterplot of categorical variables and salaries

The following obvious conclusions can be seen in figure 1.

(a) The linear correlation between player salary and player height is not significant.

(b) The linear correlation between player salary and player weight is insignificant.

(c) The linear correlation between player salary and player age is not significant, but the higher paid players are in the 25-35 age range.

(d) The linear correlation between player salary and player age was not significant, with the highest paid players being between 3-13 years old, which is consistent with the findings on player age.

(e) There is a positive correlation between player salary and playing time, with the highest paid players playing more than 30 minutes.

(f) There is a positive correlation between player salary and player PER.

(g) There is a positive correlation between player salary and player WS, with the majority of players having a WS score of 5 or less, including some of the highest paid players.

(h) The linear correlation between player salary and player BPM is not significant.

(i) There is a positive correlation between player salary and player VORP. Although the vast majority of players have a VORP score of below 2, those with a score of above 2 have a significantly higher salary than other players.

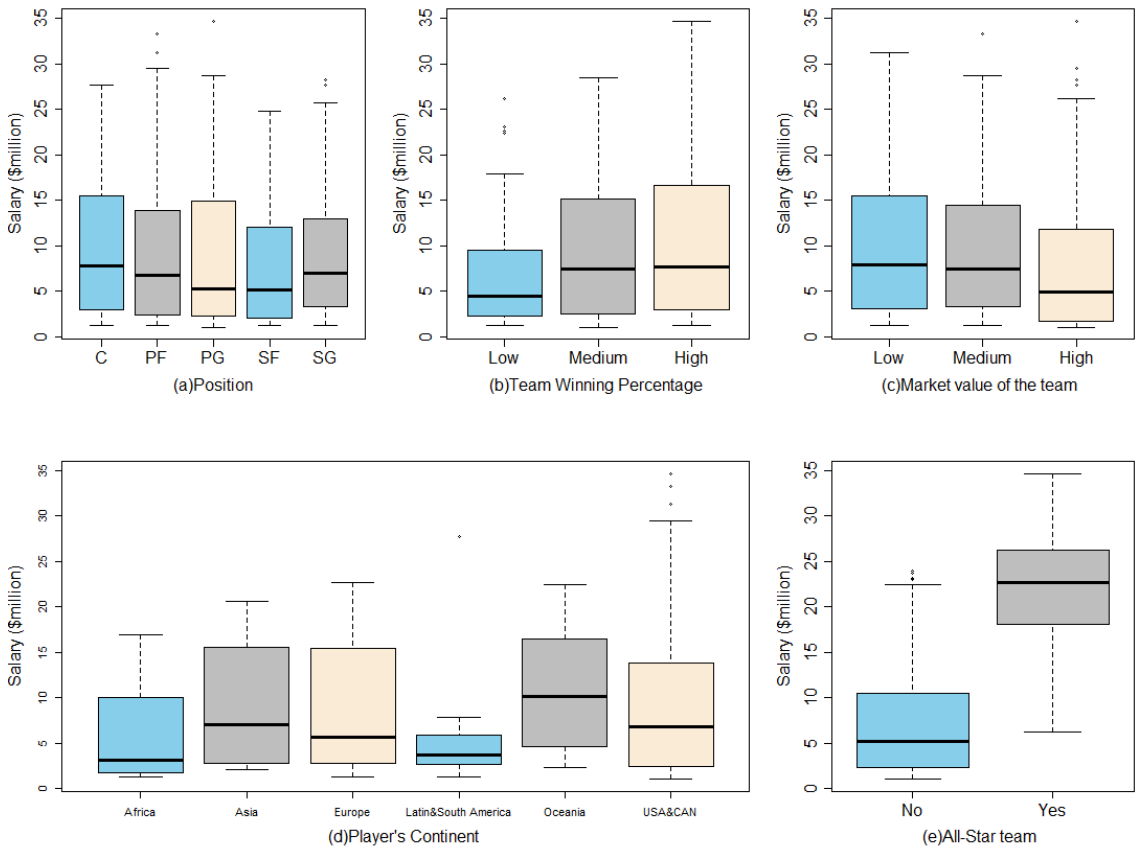

Figure 2 Box-plot of categorical variables and salaries

The following obvious conclusions can be seen in figure 2.

(a) The mean (in terms of median) salary of players at different positions varies, with C(Center) > PF(Power Forward) > SG(Shooting Guard) > PG(Point Guard) > SF(Small Forward)

(b) Players with a high and medium Team Winning Percentage have significantly higher salary averages (in terms of median) than players with a low Team Winning Percentage.

(c) Players in the team with a low or medium Market value have higher salary averages (in terms of median) than players in the team with a high Market value, which is contrary to people's common knowledge of the high salaries of the rich teams.

(d) The average(in terms of median)salary of players in developed regions such as North America, Oceania, etc. is significantly higher than that of players in developing regions such as Latin America, South America, Africa, etc.

(e) The average(in terms of median) salary of players selected for the All-Star team is significantly higher than that of players who have not been selected.

3.3. Model building and interpretation of results

The least squares method are used to develop a multiple linear regression model of NBA player salary Y for the 2017-2018 season with Xi for each influencing factor, and used stepwise regression to filter the variables.

The results of the linear regression are shown in table 2 below:

Table2. Regression model results

Variable	Coefficient	t value	Pr(>|t|)	VIF
(Intercept)	-10.079	-3.298	0.00109	
Weight(X1)	0.076	2.885	0.00421	1.07
Players' playing age (X4)	0.414	5.351	1.76e-07	1.14
Players' playing time per game (X5)	0.323	6.628	1.61e-10	1.66
Value Over Replacement Player (X9)	0.707	2.517	0.01237	1.84
Whether he have been selected for the All-Star team (X14)=Yes	8.850	8.277	4.43e-15	1.61
R-squared: 0.5908	Adjusted R-squared: 0.5839			
F-statistic: 85.18 on 5 and 295 DF	F-test p-value: < 2.2e-16			

From table 2, F = 62.16 and P = 2.2e-16 indicate that this regression equation is meaningful and the variables X1, X4, X5, X9, X14 pass the significance test at the 0.05 level of significance and the VIF coefficients are all less than 5, therefore there is no multicollinearity between the independent variables. The value of the Adjusted R-squared was 0.5839, indicating that 58.39% of the variation in pay could be explained by the model.

As can be seen from figure 1, all sample points are basically distributed around the asymptote, the scatter is basically linear, and the data and the model basically match, indicating that the residuals basically obey a normal distribution. The standardized residual distribution is scattered, and most of them lie between -2 and 2, indicating that the residuals are random and there is no heteroscedasticity, and the multiple linear regression model fits well.

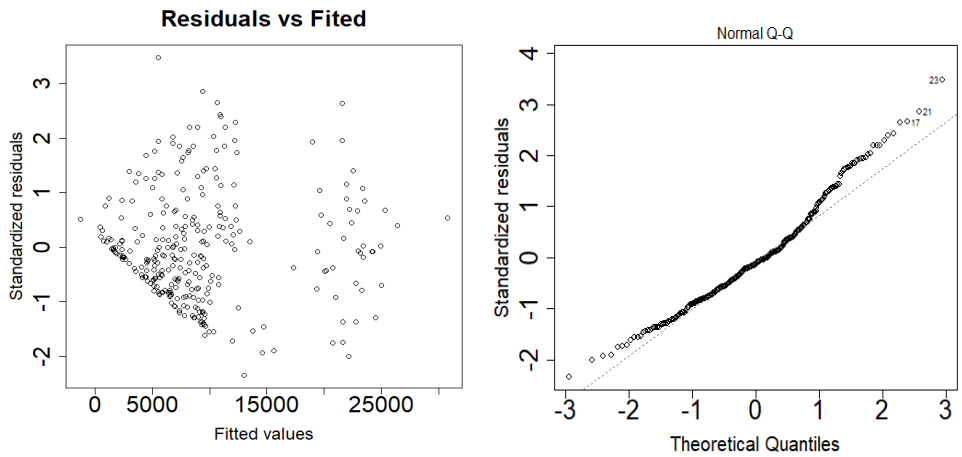

Figure 3 Plot of standardised residual and Q-QPlot

The regression equation at this point is: Y = -10.079+ 0.076X1+ 0.414 X4 +0.323 X5 + 0.707 X9 + 8.850 X14+ε

Holding other influences constant,this equation indicates that:

(1) For every 1KG increase in the weight of an NBA player (X1), the player's salary will increase by $0.076 million;

(2)For every 1 year increase in the NBA player's age (X4), the player's salary will increase by $0.414 million;

(3)For each additional minute of playing time (X5), the player's salary will increase by $0.323 million;

(4)For every 1 unit increase in Value Over Replacement Player (X9), the player's salary will increase by $0.707 million;

(5)NBA players who have been selected as All-Stars (X14=Yes) will earn $8.850 million more than players who have not been selected (X14=No).

4. CONCLUSION

This paper quantifies the factors influencing the salaries of NBA players based on the methods of multiple linear regression and ANOVA. From the conclusion of the model, it can be seen that NBA players' Value Over Replacement Player and playing age have a strong

positive relationship with their salaries, while team-related indicators such as Win Share do not have a significant impact on their salaries. Therefore, a player's value should be measured first and foremost on the basis of his ability and experience on the court. However, for the CBA league, which is still in its development stage, the player salary model has not yet been solidified and there is still room for improvement. Teams should not only use these individual ability indicators as a reference basis for player salary, but also introduce more team indicators such as Win Share as a reference basis, so that the individual interests of players can be in line with the overall interests of the team, thus achieving a win-win situation for both sides.

Since the performance and salary of players will also be affected by factors such as contract term, the model can be improved by collecting long-term data and selecting independent variables such as "contract year" in the future. So as to create a more reasonable salary model.

ACKNOWLEDGMENTS

Firstly, I would like to express my gratitude to my teachers and professors at university who have provided me with invaluable guidance at every stage of the writing of this thesis. In addition, I would like to thank all my friends and parents, especially Han, without whose encouragement and support I would not have been able to complete my dissertation.

REFERENCES

[1] Lin, H. Y. , & Yang, C. H. Uncertainty, specific investment, and contract duration: evidence from the MLB player market. Empirical Economics, 50(3) (2016) 1009-1028.

[2] Grdinarus. The most important key performance indicators at BC Timisoara management. Timisoara Physical Education and Rehabilitation Journal, 27(3) (2005)13-17.

[3] Frick B. Performance, salaries, and contract length: Empirical evidence from German soccer. International Journal of Sport Finance, 6(2) (2011) 87.

[4] Frascatore MR. Staggered Contracts and Profitable Entry Deterrence: an Application to Professional Sports. Journal of Economics, 66(1) (1997) 71-86.

[5] Bukstein S. A New Solution for Salary Disputes: Implementing Salary Arbitration in the National Basketball Association. Marq. Sports L. Rev. 22, 2011, 25.

[6] Ducking J, Groothuis PA, Hill J.R. Minimum Pay Scale and Career Length in the NBA. Working Papers, 53(4) (2012) 617-635.

[7] Wallace Michael. Wage Discrimination in the National Basketball Association: Is ere Discrimination Based on Race. Work and Occupations 15(3)(1988) 294-312.

[8] Berri D J. Does One Simply Need to Score to Score? International Journal of Sport Finance, 2(4) (2007)190-205

[9] Stiroh Kevin. Playing for keeps Pay and performance in NBA. Economic Inquiry, 45(1)(2007)145-161.

[10] Shen Chen. A Study of the Factors Influencing NBA Player Salaries. Business, (26) (2016) 1.DOI: CNKI:SUN:SHNG.0.2016-26-043

[11] Chen Sitong, Cheng Chen &Chai, Yu jie. A Study of the Correlation between Salary and Team Performance and Player Performance. Sport, (19) (2015) 33-35.

[12] Jin Saiying. A Study of NBA Players and Club Wages. China Sport Science, 24(2) (2004) 9-11

[13] Zhang Chenying, Li Xiliang. An Analysis of Antitrust Law on the NBA Salary Cap Regulations and on the Legality of the CBA Salary Cap Regulations. Competition Policy Research, (06) (2020) 47-62.

[14] Basketball-reference, Data Source, https://www.basketball-reference.com/