Influencing Factors of NBA Player's Salary Based on Mixed Linear Model

Li'ao Qian *

Department of Mathematics, University of Nottingham, Ningbo, China

* Corresponding Author Email: ssylq5@nottingham.edu.com

Abstract. With the intensification of the global commercial influence of NBA, the salary of its player has always been a hot topic. This paper aims to offer a more practical insight for stakeholders within NBA ecosystem and provide a reference for the revolution and development of China's professional basketball league by further analyzing the main factors that impact the salary level of NBA player based on previous study. Although many of the influencing factors considered in this paper have been mentioned in previous literature, they have not been systematically sorted out and analyzed based on data. A total of 8626 valid datasets of 454 NBA players in 2017-2018 season are selected as the research object for the study. By constructing a mixed linear model, which considers both fixed effects and random effects in NBA player's salary influencing factors, two conclusions are drawn. Firstly, there are multiple factors that affect the salary of NBA players, among which age, morphology, draft ranking, game experience and usage have obvious influence on the salary of players. On the other hand, nationality, education and efficiency (related to team performance) have no significant correlation with player's salary.

Keywords: NBA player; salary; mixed linear model.

1. Introduction

The National Basketball Association (NBA) is one of the most prominent and lucrative sports leagues globally with massive fan base and widespread popularity. In the field of commercial sports, professional athletes are the most important resource in most cases [1]. Therefore, the salary of NBA players has become a topic of immense interest and scrutiny among sports enthusiasts, economists, and researchers alike. Exploring the factors that influence NBA player salaries through statistical analysis can provide valuable insights into the economics of professional basketball and shed light on the dynamics governing player compensation [2]. Mixed linear model, a statistical technique used to identify underlying factors or dimensions from a set of observed variables [3, 4], offers a comprehensive approach to examining the determinants of NBA player's salary. By analyzing a multitude of variables such as player's age, morphology and draft ranking, a deeper understanding of the key factors contributing to salary disparities among NBA players could be obtained.

The significance of this research lies in its potential to uncover the complex interplay of factors affecting player's salaries, aiding team managers, agents, and players in strategic decision-making regarding contracts, negotiations, and career choices [6]. Furthermore, the findings can provide insights into the broader economic context of professional sports and contribute to the ongoing dialogue surrounding income inequality and labor market dynamics in the sports industry [7, 8].

To further strengthen the understanding of NBA player's salary, this paper builds upon existing research and incorporates data-driven analysis. In similar directions, Carvajal and Mellado conducted factor analysis on a comprehensive dataset of NBA player salaries and various performance metrics. The authors identified three primary factors influencing player salaries: on-court performance, marketability/popularity, and team market size. The findings suggest that factors beyond on-court performance, such as endorsements and market exposure, play a vital role in salary determination [5]. Lin and Simmon employed factor analysis by multiple regression model to evaluate the impact of player statistics on NBA salaries. The study found that a combination of offensive and defensive performance factors, including points per game, assists, rebounds, and blocks, strongly influenced player’s salary. Additionally, the use of advanced statistics, such as Player Efficiency Rating (PER) and Win Shares, further improved the model's accuracy in predicting salaries [6].

Li and He focused on reputation-related factors, this study applied factor analysis to investigate their influence on NBA player’s salary. The authors identified three key reputation factors: All-Star appearances, MVP awards, and media rankings. The results indicated that players with higher reputation scores in these categories tended to earn higher salaries, reflecting the value placed on public recognition in salary negotiations [7]. In summation, previous studies have delved into various aspects of athlete compensation, highlighting the influence of factors like player performance, market size, endorsements, and team success [5-7]. However, this study aims to offer a comprehensive and updated analysis, leveraging advanced statistical techniques to explore the factors behind NBA player’s salary in greater depth. By employing mixed linear model, this study will examine several potential and undeveloped factors and variables derived from publicly available data sources, including player statistics in field of morphology, draft ranking and on court performance, team performance records [8], nationality [9, 10] and education background [11, 12]. The findings will not only contribute to the existing body of research but also provide practical insights for stakeholders within the NBA ecosystem. Furthermore, China's basketball professional league is in a critical period of reform at present stage, which needs to introduce a large quantity of high-level basketball players. Studying the key factors affecting the salary of NBA players especially the player's salary management has great reference value to the development of China's professional basketball league.

2. Methodology

2.1. Data Source

A total of 454 players from 30 teams in the regular seasons of NBA in 2017-2018 season were selected as the research objects. The number of games played by the selected players must be greater than 5 in a single season (player data less than this number is considered invalid data. After screening and excluding missing values, 8626 valid data are finally included.

2.2. Data Processing

Data in this paper is extracted from the Kaggle website, which was compiled by Justinas Cirtautas from the NBA players’ Biometric, biographic and basic box score features from 1996 to 2021 season, which was published and updated in 2022 and NBA salary dataset (2017-2018) published by Koki Ando in 2018. All data processing is done by EXCEL 16 and SPSSAU.

2.3. Variable Description

2.3.1 Dependent variable

There is a significant distinction of salary among NBA players, which can be exemplified as the phenomenon that some superstar players can earn tens of millions of dollars a year, while some players can only earn millions of dollars or even less, making player compensation fluctuate greatly. Therefore, in order to construct the player salary model more properly, the natural logarithm is used to convert the NBA player salary.

2.3.2 Fixed effect

Categorical variables include nationality, education, morphology features and draft ranking, as follows. Nationality: Nationalities of players are divided into two categories: American players and international players. Education: Education background of players is divided into two categories: players with college degree and players without college degree (when they were drafted). K-Means clustering was used to divide height and weight of players into three categories (Table 1).

Table 1. Morphology category

Category	Means
Low height & weight	187.13cm, 87.54kg
Medium height & weight	199.63cm, 95.80kg
High height & weight	209.23cm, 112.63kg

Draft ranking: Players are divided into four groups according to the NBA draft round ranking (Table 2).

Table 2. Draft ranking category

Category	Draft pick
First-class	1st-15th
Second-class	16th-30th
Third-class	31st-45th
Fourth-class	46th-60th

2.3.3 Random effect

The random effect is the NBA season, which can be expressed as the general fluctuation of salary caused by other factors among seasons.

2.4. Method Introduction

According to the purpose of the study, through the analysis of fixed effects and random effects, the mixed linear model of NBA player, compensation is established as follows.

$$Y_{salary} = \beta_0 + \beta_{Nationality}x_1 + \beta_{Education}x_2 + \cdots + +\beta_{Experience}x_8 + T + \varepsilon \tag{1}$$

In the formula above, Y_{salary} is the dependent variable (NBA player's salary), $\beta_{Nationality}$, $\beta_{Education}$, β_{Age}, $\beta_{Morphology}$, β_{Draft}, β_{Time}, $\beta_{Efficiency}$, $\beta_{Usage}\beta_{Experience}$ indicate the fixed effect factors respectively. T is a random effect (season) and ε is a random error.

By comparing the information statistics of different random effect covariance structure models (Akaechi information criterion AlC and Schwarz Bayes Criterion BIC), it is found that the NBA player salary model with variance component covariance structure (AIC=8273.918, BIC=8292.389) is the best to fit the model).

3. Results and Discussion

3.1. Model-based underlying Data Characteristics

The descriptive statistical values of dependent variables and partial covariates in the mixed linear model of player compensation in NBA season 2017-2018 are manifested in Table 1.

Table 3. The basic data statistics of NBA players in 2017-2018 season

	n	minimum	maximum	average	S.D.	median
Age	454	20	38	26.51	4.148	25
player height	454	182.88	220.98	201.473	8.295	203.2
player weight	454	73.028	131.542	100.126	13.239	97.522
draft number	454	1	58	19.753	14.811	18
game played	454	1	82	53.63	24.253	61
Points	454	0.6	28.1	8.746	6.3	7.3
Rebound	454	0.4	16	3.825	2.907	3.25
Assist	454	0.1	8.2	1.962	1.805	1.35
Efficiency	454	-30.7	13.2	-2.701	8.327	-1.75
usage rate	454	0.079	0.318	0.183	0.051	0.179
Salary	454	119010	29512900	7637613.7	7975621.1	5086072.5
LN salary	454	11.687	17.2	15.222	1.25	15.442

3.2. Output of the Model on NBA Player's Compensation

In table 4, B refers to the regression coefficient value, and the SE, which is called the standard error represents the fluctuation of B value. In the normalized coefficient, β is the value regression coefficient when the constant is 0. In the significance coefficient, t is a value used to compute p (meaningful only if it is less than 0.5), which symbolizes the significance of the considered factors impacting the dependent variable. The VIF, which is the reciprocal of Tolerance, indicates the collinearity problem of variables in the model when the value is less than 0.2.

Table 4. Analysis of the fixed effect impact the NBA player salary in 2017-2018

MLR results (n=454)							
	Nonnormalized coefficient		Normalized coefficient	Significance coefficient		Collinearity diagnostic	
	B	SE	Beta β	t	p	VIF	Tolerance
Constant	7.779	1.549	-	5.023	0.000**	-	-
Nationality	-0.018	0.244	-0.006	-0.074	0.941	1.622	0.617
Education	-0.002	0.247	-0.001	-0.009	0.992	1.53	0.654
Age	0.096	0.022	0.319	4.416	0.000**	1.145	0.873
Morphology	0.012	0.004	0.203	2.853	0.005**	1.113	0.898
Draft	0.021	0.005	-0.337	4.268	0.000**	1.366	0.732
Efficiency	0.015	0.012	0.101	1.263	0.21	1.394	0.717
Usage	5.831	1.889	0.237	3.087	0.003**	1.299	0.77
Experience	0.014	0.004	0.266	3.444	0.001**	1.31	0.763
R^2	0.586						
adjusted R^2	0.55						
F	F (8,91) =16.097, p=0.000						
D-W	1.981						
* p<0.05 ** p<0.01							

3.3. Explanation and Evaluation of the Model

The result of the multiple linear regression analysis with Nationality, Education, Age, Morphology, Draft, Efficiency, Usage, Experience as independent variables, with the Salary as the dependent variable, can be seen from the chart. Model R^2 value is 0.586, means that the Nationality, Education, Age, Morphology, Draft, Efficiency, the Usage, Experience can explain 58.6% of the Salary change. It was found that the model passed the F test (F=16.097, p=0.000<0.05), which shows that at least one of Nationality, Education, Age, Morphology, Draft, Efficiency, Usage and Experience would affect Salary relations. In addition, in view of the model, the multicollinearity test of the model is tested and found that all VIF values in the model are less than 5, which means that there is no collinearity problem. Moreover, the D-W value is near the number 2, which indicates that there is no autocorrelation in the model, and there is no correlation between the sample data, and the model is good. The final concrete analysis shows that:

The regression coefficient of Age was 0.096(t=4.416, p=0.000<0.01), which means that Age has a significant positive influence on Salary. Morphology regression coefficient value was 0.012(t=2.853, p=0.005<0.01), which means that Morphology has a significant positive influence on Salary. The regression coefficient of Draft was 0.021(t=4.268, p=0.000<0.01), which means that Draft has a significant positive impact on Salary. The regression coefficient of Usage was 5.831(t=3.087, p=0.003<0.01), which means that Usage has a significant positive impact on Salary. The regression coefficient of Experience was 0.014(t=3.444, p=0.001<0.01), which means that Experience has a significant positive impact on Salary.

On the other hand, the regression coefficient value of Nationality was -0.018(t=-0.074, p=0.941>0.05), which means that Nationality has no obvious correlation to Salary. The regression coefficient of Education was -0.002(t=-0.009, p=0.992>0.05), which means that Education has faint influence on Salary. The regression coefficient of Efficiency was 0.015(t=1.263, p=0.210>0.05), which means that Efficiency has a slight positive influence on Salary.

In summation, the analysis shows that Age, Morphology, Draft, Usage and Experience have significant positive influences on Salary. However, Nationality, Education and Efficiency have faint influence on Salary.

3.4. Discussion and Assessment

3.4.1 About the mixed linear model

Mixed linear model is an extension of the general linear model, which retains the assumption of normality of the dependent variable but waives the assumption of homogeneity of variance and independence, allowing the data to exhibit correlated, non-constant variability [13]. In addition, the mixed linear model can analyze both random effects and fixed effects, which contains continuous variables and categorical variables at the same time [14]. Thus, this model is relatively suitable for analyzing the complex basketball player data.

3.4.2 About the influence factors of the NBA player's compensation

Based on the result of the mixed linear model, Age, Morphology, Draft, Usage and Experience all have significant positive influences on Salary. However, Nationality, Education and Efficiency have faint influence on Salary.

According to NBA compensation rule, most young player in NBA (aged under 24) could only be offered the rookie contract, which is lower than NBA average income. With the rise of the career years, competition experience would be accumulated, and the competitive level would be improved. Thus, the elder player could get much larger contract [15].

Moreover, morphology is another non-artificial factor affecting the NBA player compensation. Compared to other NBA players, tall and heavy players are generally paid a higher level of salary, while low height and small weight players are paid slightly less, but there is no significant difference. On the other hand, the salary level of medium height and weight players is significantly lower, mainly because in the NBA league, the body shape of players has an important relationship with their position and offensive and defensive ability. Players of low height and small weight have better speed and agility, faster physical recovery, better outside skills, usually play as defenders, who are mainly responsible for the tactical organization and scoring of the whole team on the field; Tall and heavy players have stronger muscle strength and generally play the role of center or power forward. They can occupy a relatively favorable offensive and defensive position (closer to the basket) in physical confrontation by virtue of their size, score goals and compete for rebounds with solid interior skills. These two types of players have a greater contribution in controlling the ball and fighting for scoring opportunities, which can create more direct value for the team, so they can get more compensation. The conclusion was also dropped by Norton et al., who found that many clubs are pursuing tall players, based on the 1993 NBA league data found that every 1.0 cm or 1.3 kg increase in players, they can earn an additional income of $43,000 during their career [16].

Draft is a significant system for every NBA team to select players in sequence since its establishment in 1947. With decades of reforming and perfecting, the system has matured to a great extent [17]. The salary has been depended on the draft ranking ever since 1995-1996 season. Consequently, the player with higher draft ranking would obviously get higher salary in their rookie contract. Moreover, the draft ranking can also epitomize the player's talent and potential significantly. Under most circumstances, players being selected in the first round would have higher talent, which will push them to play at a more competitive level and earn more in the future.

The NBA player's game experience and usage can also significantly affect their salary. From the perspective of sports economics, as a kind of labor product, professional sports competition is mainly produced by athletes, and the quality and attractiveness of sports competition products are affected by the level of athletes' competitive performance. The number of games played, and the usage rate represent the working experience and importance in work accordingly, which can contribute to the level of compensation. However, it is amazingly seen that the efficiency of player only have faint effect on the salary level. Based on the algorithm of efficiency adopted in this paper, the player's efficiency is highly correlated to the performance of the entire team. Thus, according to the compensation rule of NBA, with the same salary cap of every team, outstanding players in an overall poor team would be well paid. Vice versa, players with worse performance would be poorly paid in an excellent team.

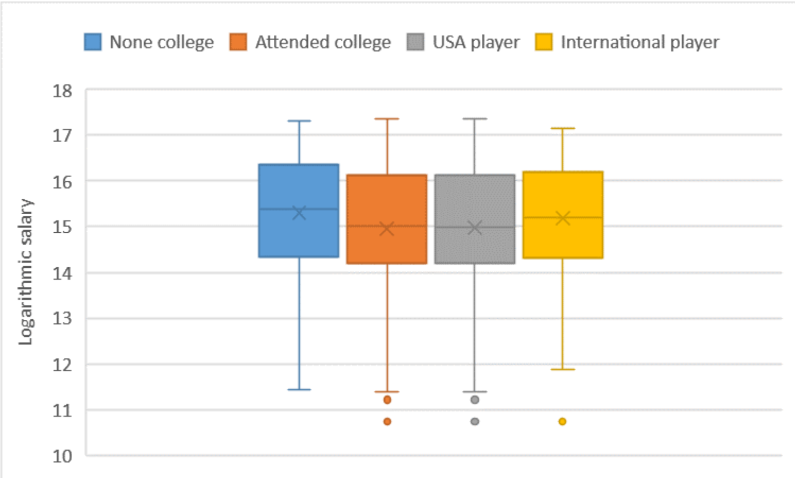

Fig. 1 The salary distribution of NBA players from different background

The box plot in Figure 1 manifests that two of the categorical variables: nationality and education does not affect the salary distribution of NBA player distinctly. However, the previous study indicates the existence of racial discrimination in NBA player compensation [9]. It's mainly because NBA was enhancing its global influence in recent years by recruiting increasing number of international players. Otherwise, with the perfection of the college sports training system, more talented players in high school choose to go to university instead of entering the NBA directly, which narrow the salary gap between the players attended the university and the players without university experience [17].

3.5. Suggestions

Player's data in field of morphology and individual skills should be well collected by teams to construct more precise mathematical model of salary. Teams should consider their on-field performance and importance more than their team performance when signing free agents. With the prosperity and development of world basketball, NBA teams should recruit high-level international players at a higher salary level, so as to improve the competitiveness of the team and global influence.

4. Conclusion

By constructing the mixed linear model of NBA player's salary. This paper has indicated two conclusions. Firstly, there are many factors that affect the salary of NBA players, among which age, morphology, draft ranking, game experience and usage have obvious influence on the salary of players. On the other hand, nationality, education and efficiency (related to team performance) have no significant correlation with player's salary.

The mathematical model of salary analysis proposed in this paper quantizes and adds several new indicators such as nationality, education background, morphology and draft ranking based on previous scholars' models, enriching the research perspective of players' salary. In the future, the above indicators and model can be combined to conduct more detailed research and analyze the data of different leagues. Furthermore, some higher-order data that can more comprehensively and objectively reflect players' on-court performance can be added to obtain a more accurate player salary model. However, due to the collinearity problem of some higher-order data, principal component analysis and ridge regression analysis are required, so they are not included in this paper. Otherwise, although the random effect "season" has been proposed in this paper, the impact of random effect on NBA player's salary is inconclusive because of the lack of supporting data. In future studies, if the player's data in different seasons are sufficiently collected, the influence of random effects on player's salary can be further explained.

References

[1] Stewart B, Smith A. The special features of sport. Annals of Leisure Research, 1999, 2(1): 87-99.

[2] Hinton, Alexander, Sun Yiguo. The sunk-cost fallacy in the National Basketball Association: evidence using player salary and playing game. Empirical economics, 2020, 59 (2): 1019-1036.

[3] Casals Martí, Girabent-Farrés Montserrat, Carrasco Josep L. Methodological Quality and Reporting of Generalized Linear Mixed Models in Clinical Medicine (2000–2012): A Systematic Review. United States: Public Library of Science PloS one, 2014, 9 (11): 112653-112653.

[4] Wang Xu, et al. Statistical primer: an introduction to the application of linear mixed-effects models in cardiothoracic surgery outcomes research-a case study using homograft pulmonary valve replacement data. Oxford University Press, European journal of cardio-thoracic surgery, 2022, 62(4).

[5] Lyons Jr R, Jackson Jr EN, Livingston A. Determinants of NBA player salaries. The Sport Journal. 2015, 18.

[6] Lin Y C, Simmons R. Exploring the factors influencing NBA player salaries: A principal component analysis. Journal of Applied Sport Management, 2019, 11(3): 27-43.

[7] Papadaki I, Tsagris M. Are NBA Players' Salaries in Accordance with Their Performance on Court?. In Advances in Econometrics, Operational Research, Data Science and Actuarial Studies: Techniques and Theories, 2022, 405-428.

[8] Kralj B, Zakyne A. Multivariate analysis of determinants of NBA player salaries. Journal of Quantitative Analysis in Sports, 2017, 13(1): 1-18.

[9] Kahn L M, Sherer P D. Racial differences in professional basketball players' compensation. Journal of Labor Economics, 1988, 6(1): 40-61.

[10] Yang C H, Lin H Y. Is there salary discrimination by nationality in the NBA? Foreign talent or foreign market. Journal of Sports Economics, 2012, 13(1): 53-75.

[11] Koba T H, et al. An exploration of professional U.S. -based basketball players competing in Turkey. Journal of Global Sport Management, 2020, 1-22.

[12] Groothuis P A, Hill J R, Perri T J. Early entry in the NBA draft: The influence of unraveling, human capital, and option value. Journal of Sports Economics, 2004, 8(3): 223-243.

[13] Piepho, Hans-Peter. A coefficient of determination (R2) for generalized linear mixed models. Biometrical journal, 2019, 61 (4): 860-872.

[14] Moore R, et al. A linear mixed-model approach to study multivariate gene-environment interactions. Nature genetics, 2019, 51 (1): 180-186.

[15] Hill James R, Jolly Nicholas A. Salary Distribution and Collective Bargaining Agreements: A Case Study of the NBA. Industrial relations (Berkeley), 2012, 51 (2): 342-363.

[16] Norton K, Olds T. Morphological evolution of athletes over the 20th century. Sports Medicine, 2001, 31(11): 763-78.

[17] Berger Tobias, Daumann, Frank. Jumping to conclusions-an analysis of the NBA Draft Combine athleticism data and its influence on managerial decision-making. Sport, business and management, 2021, 11 (5): 515-534.