Vol. 8 (2025): Proceedings of the 2025 International Conference on Social Sciences and Cultural Studies (ICSSCS 2025)
E-ISSN: 3105-0522 P-ISSN：3105-0514
DOI: https://doi.org/10.70267/icsscs.202507

Exploring the Impact of Various Factors on the Salaries of Players in the National Basketball Association

Yunlong Guo*

College of Computer Science, Beijing University of Technology, Beijing, 100124, China

Corresponding author: Yunlong Guo

Abstract

This study focuses on the factors affecting NBA player salaries. On the basis of multidimensional data from 288 players in the 2022--2023 season, a random forest model and multiple linear regression model are used to explore the impacts of individual competitive level, implicit traits, commercial endorsement value, and institutional restrictions on salaries. Research has shown that a player’s performance in the field, especially their offensive contribution, is the core determinant of salary. Multiple linear regression validation reveals that scores and team win rates have a significant positive effect on salary, whereas the number of games has a negative effect. Defensive indicators and injury absenteeism do not significantly affect player salaries. In addition, the team’s luxury tax and off-field commercial value have weak explanatory power for salaries, revealing that the current salary system is still centered around competitive performance. This research provides a quantitative basis for player salary negotiations, team salary structure optimization, and alliance policy adjustments. In the future, real-time game data and social media indicators can be further integrated to improve the prediction accuracy.

Keywords

NBA player salary, random forest model, multiple linear regression model, performance on the field, team winning percentage

1. Introduction

The National Basketball Association (NBA) is the most influential professional basketball league in the world and is considered the most competitive male professional basketball league in the world. In the NBA, significant differences in player salaries are common, which can be attributed to many factors. The salary of NBA players not only reflects the value of their individual competitive level but is also influenced by multiple factors, such as team operation strategies and league rules. In recent years, with the development of digital technology and the expansion of the sports market, the factors affecting NBA athlete salaries have changed.

The on-court performance of NBA players, including basic data such as scoring, assists, rebounds, and higher-order indicators such as player shooting percentage, is the main factor affecting player salaries. One study uses a two-stage NDEA model to evaluate whether a player’s performance in the field can be fully converted into their salary level and to assess their salary creation efficiency (Chen, 2025). There are studies that use descriptive statistics and simple correlation analysis to investigate the correlation between salary, team performance, and player performance. Some intangible traits that cannot be quantified, such as injury risk, can also affect the salaries of NBA players. The NBA has gone through three major labor wars, suspensions, and shutdowns, and its salary system has evolved from a fixed salary to a dynamic regulation model of “salary cap + luxury tax”. Some studies have used methods such as literature reviews and comparative analyses to explore the impact of the NBA salary cap and luxury tax system on player salaries(Zhao and Li, 2018, Zhang, 2015). The implementation of the NBA salary cap has limited the overall wages that teams can pay to their players, and through economic penalties such as luxury taxes, it has restricted some teams from using their abundant financial advantages to win over star players. To some extent, this has prevented players from receiving excessive salaries and made salary distribution more complex and diverse.

Most studies focus on the impact of player performance and athletic level on player salaries while neglecting the relationship between player commercial endorsements and salaries, as well as the impact mechanism of alliance systems such as labor agreement rules on player salaries. This study aims to break through the traditional linear framework of player salary evaluation and uses a random forest model and multiple linear regression model to explore the impact of four factors, namely, individual competitive level, implicit traits, commercial endorsement value, and institutional limitations, on player salary (Xu and Xu, 2026, Lyons et al., 2015). Provide data support for player salary negotiations, player career planning, team salary structure optimization, and team building strategies.

2. Data and methods

2.1 Data sources and explanations

NBA players with appearance records in the 2022--2023 season, salary data in the 2023--2024 season, and NBA players who have played multiple teams in the 2022--2023 season should be considered. The data include four dimensions: court performance, implicit traits, commercial endorsement value, and institutional limitations. All the data are numerical, and the final sample size is 288 players. The data were collected from the official Kaggle, basketball reference, Sportrac, and NBA websites.

2.2 Indicator selection and explanation

Analyze and research the on-court performance of NBA players during the 2022--2023 season, including basic metrics for each game: number of games played, minutes on the court, points scored, assists, steals, rebounds, blocks, and turnovers. Additionally, advanced metrics for each game, shooting percentage, as well as latent indicators representing injury risk, such as the duration of absences due to injury, were assessed. Analyze metrics related to NBA teams for the 2022--2023 season: team win percentages and luxury tax implications for the teams. The indicators that represent players’ off-court commercial value are as follows: off-court income. Finally, we present the dependent variable: player salary for the subsequent 2023--2024 season.

Table 1: Basic description of indicators

Indicators	Minimum	Maximum	Average	Standard deviation
Game played	3.00	8.20×10^{1}	5.34×10^{1}	2.22×10^{1}
Playing time	2.40	3.74×10^{1}	2.19×10^{1}	9.53
Shooting average	1.58×10^{-1}	7.76×10^{-1}	4.69×10^{-1}	8.90×10^{-2}
Score	0.00	3.31×10^{1}	1.07×10^{1}	7.70
Rebound	0.00	1.25×10^{1}	3.96	2.55
Assist	0.00	1.07×10^{1}	2.47	2.20
Steal	0.00	1.80	6.75×10^{-1}	3.95×10^{-1}
Block	0.00	2.30	4.20×10^{-1}	4.04×10^{-1}
Turnover	0.00	4.10	1.28	9.31×10^{-1}
Absence	0.00	1.17×10^{2}	2.77	1.34×10^{1}
Winning percentage	2.07×10^{-1}	7.07×10^{-1}	4.99×10^{-1}	1.22×10^{-1}
Luxury tax	0.00	1.70×10^{8}	2.07×10^{7}	4.56×10^{7}
Business value	0.00	7.00×10^{7}	1.18×10^{6}	8.11×10^{6}
Salary	5.47×10^{5}	5.17×10^{7}	1.17×10^{7}	1.27×10^{7}

Owing to the significant dimensional differences among various indicator data, the data are first preprocessed by normalizing the maximum and minimum values of each indicator data. Maximum minimum normalization (also known as linear normalization) is a commonly used data normalization method that linearly transforms raw data into a specified range (usually [0, 1]). The mathematical formula is as follows:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}} \tag{1}$$

Among them is the raw data value X (such as players' playing time, scores, etc.).X_{min} is the minimum value of this feature (the minimum playing time, minimum score, etc., among all players).X_{max} is the maximum value of this feature (the maximum playing time, maximum score, etc., among all players).X' is the normalized value, ranging from [0,1].

Owing to the large number of indicators, the random forest algorithm was used in this study to select indicators with greater importance to the model. The random forest (RF), as an on parametric regression technique, has unique advantages in dealing with complex data relationships (Kam, 1995, Breiman, 2001). Compared with traditional classification algorithms, this model significantly reduces the risk of overfitting through the idea of ensemble learning, making the prediction results more universal. Moreover, it can effectively handle high-dimensional feature spaces and adapt to datasets with numerous variables. In addition, random forests synchronously implement feature importance assessment during the training process, providing a quantitative basis for variable selection. These characteristics make them excellent for complex relationship modeling and feature interpretation. This study uses a multiple linear regression model to analyze the impact of different factors on NBA player salaries (Sun, 2000, Cameron and Trivedi, 2005). The basic form of the multiple linear regression model is as follows:

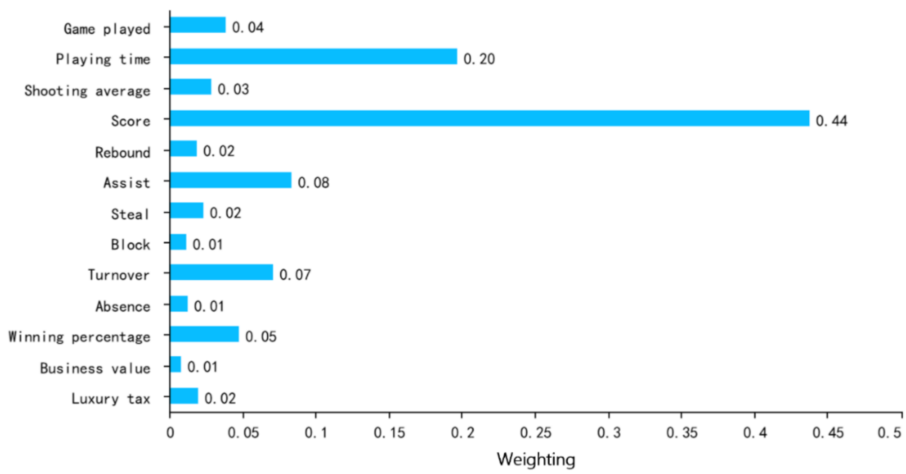
$$Salary = \beta_0 + \beta_1 \times Games\ Played + \beta_2 \times Minutes\ Played + \cdots + \beta_{13} \times Off - court\ Income + \epsilon \tag{2}$$

where β_0 is the intercept term; β_1, β_2, ...,β_{13} are the regression coefficients of the respective variables; and ϵ is the error term.

3. Experimental Results

Using the number of games, playing time, shooting accuracy, scoring, rebounds, assists, steals, blocks, turnovers, absenteeism, team win rate, commercial value, and team luxury tax as independent variables and salary as the dependent variable, a random forest model was conducted with a total of 288 samples included in the analysis. The training set had a sample size of 230, and the testing set had a sample size of 58.

Figure 1: Indicator feature weight map (Data form: Kaggle, basketball reference, Sportrac, and NBA official websites)



The feature weight reflects the importance of each indicator's contribution to the model, and its sum is 1. Figure 1 shows that the weight of the score is 43.82%, which is the highest among all the indicators and plays a key role in model construction. It can also represent the player's performance on the offensive end to a certain extent. The proportion of playing time is 19.74%, which plays an important role in model construction and represents the attendance of players. The proportion of assists reached 8.36%.

The above three indicators account for 71.92% of the total proportion. Among the remaining 10 indicators, the indicators related to on-court performance include turnover, shooting accuracy, steals, rebounds, and blocks, accounting for 7.12%, 2.86%, 2.31%, 1.87%, and 1.21%, respectively. The proportion of absenteeism time among the indicators related to player injuries is 1.30%, which does not play a significant role in model construction. In terms of team indicators, the team win rate and the proportion of team luxury taxes are 4.75% and 1.99%, respectively. The proportion of commercial value is only 0.84%, which has the least effect on model construction. These findings indicate that a player's performance on the court is the most important, with their offensive performance (scoring, assists, etc.) being more important than their defensive performance (rebounding, steals, etc.). The team situation comes second; the injury situation and commercial value of players are far less important than the first two. On the basis of the results obtained from the random forest model, indicators with weights greater than or equal to 2% (score, playing time, assisting, errors, team win rate, number of games, shooting accuracy, and steals) are selected as independent variables, and player salary is selected as the dependent variable. A multiple linear regression model was used to explore the impacts of different indicators on player salaries.

Table 2: Results of the Linear Regression

	Nonstandardized coefficients		Standardization coefficient	t	p	Collinearity diagnosis	
	B	Standard error	*Beta*			VIF	Tolerance
Constant	-0.141	0.038	-	-3.670	0.000**	-	-
Playing time	0.099	0.089	0.111	1.107	0.269	8.914	0.112
Shooting average	0.104	0.062	0.060	1.685	0.093	1.131	0.884
Score	0.694	0.096	0.673	7.237	0.000**	7.652	0.131
Assist	0.098	0.089	0.080	1.103	0.271	4.693	0.213
Turnover	0.018	0.099	0.017	0.183	0.855	7.317	0.137
Winning percentage	0.158	0.035	0.159	4.529	0.000**	1.092	0.916
Game played	-0.147	0.042	-0.171	-3.507	0.001**	2.100	0.476
Steal	0.024	0.061	0.021	0.390	0.697	2.634	0.380
R^2	0.685						
Adjust R^2	0.676						
F	$F(8,279)=75.856$, $p=0.000$						
D-W value	1.878						

Note: Dependent variable = salary for the following year

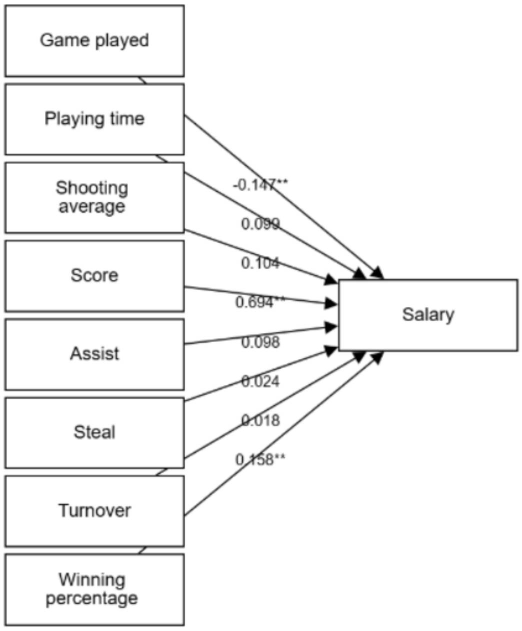
** p<0.05 ** p<0.01*

As shown in Table 2, the regression coefficients of the indicators related to player performance, such as points, shooting percentage, assists, steals, and turnovers, are 0.694, 0.104, 0.098, 0.024, and 0.018, respectively. The regression coefficients of the number of games and playing time related to player attendance were -0.147 and 0.099, respectively. For team-related indicators, the regression coefficient of the team win rate is 0.158, and the model figure is shown in Figure 2.

The final model formula is as follows:

$$\begin{aligned} Salary\ for\ the\ following\ year &= -0.141 + 0.099 \times Playing\ time + 0.104 \times Shooting\ average + 0.694 \times Score \\ &+ 0.098 \times Assist + 0.018 \times Turnover + 0.158 \times Winning\ percentage \\ &- 0.147 \times Game\ played \\ &+ 0.024 \times Steal \end{aligned} \tag{3}$$

Figure 2: Multiple linear regression model diagram (Data form: Kaggle, basketball reference, Sportrac, and official NBA websites)



Summary analysis shows that scores and team win rates have a significant positive effect on player salaries, whereas the number of games played has a negative effect on player salaries. Other indicators do not have a significant effect on player salaries.

4. Experimental Analysis

On the basis of multidimensional data of NBA players in the 2022--2023 season, combined with random forest and multiple linear regression models, this study explores the impact of different factors on players' salaries for the following year. The main conclusions are as follows:

4.1 Core Influencing Factors: On Field Offensive Performance and Team Competitiveness

Through the ranking of feature importance via the random forest model, it was found that the performance of players in the field has the most significant impact on salary, especially in offensive data, where the feature weights of scores and assists are much higher than those of other indicators. In addition, playing time, as a direct reflection of the actual intensity of a player's participation in the game, also has a positive impact on salary. This result was further validated via multiple linear regression analysis: standardized regression coefficients revealed that both scores and team win rates had a significant positive impact on salary. Specifically, controlling for other variables, for every 1 standard deviation increase in score, the average salary increase for players in the following year is approximately 0.694 units; for every one standard deviation unit increase in the team's winning rate, the average salary increase is approximately 0.158 units. From this, it can be concluded that the individual attacking efficiency of players and the overall competitive level of their teams are the core factors determining salary levels.

4.2 Complex Role of Attendance

The appearance time has a high weight in the random forest model, which can demonstrate its importance; however, in the multiple linear regression model, this indicator did not pass the significance test and may be related to multicollinearity between variables. Notably, the number of matches has a significant negative effect on salary, meaning that the more players participate in matches, the lower their salary may actually be. This seemingly contradictory phenomenon may be caused by two factors: on the one hand, frequent appearances may increase the risk of injury or physical exhaustion, which is not conducive to the long-term development of players; on the other hand, teams may limit the income ceiling of rotating players through salary structures. This discovery indicates that simply increasing playing time may not necessarily improve salary bargaining power, and players need to find a balance between attendance and efficiency.

4.3 Weak correlation between the team system and commercial value

The impact of team luxury taxes and off-field commercial value on salary is not significant. The luxury tax, as a tool for regulating the salary ceiling of the League, serves as an economic penalty at the team level rather than individual players. Although commercial endorsement value can theoretically enhance players' bargaining power, its weight in current data is only 0.84%, indicating that NBA player salaries are still driven by competitive performance at the core and that the premium effect of commercial added value has not yet been fully manifested.

4.4 Discussion

According to Table 1, the model R^2 value is 0.685, and all the above indicators can explain 68.5% of the changes in player salaries, although there are still some unclear reasons. When the F test was conducted on the model, it was found that the model passed the F test (F=75.856, p=0.000<0.05), which also indicates that at least one of the above indicators will have an impact on salary, indicating that the model is meaningful. In addition, testing for multicollinearity in the model revealed that there were VIF values greater than 5 but less than 10, indicating that there may be some collinearity issues with the indicators.

5. Conclusions

This study aims to explore the impact of different factors on NBA players' salaries. The random forest model and multiple linear regression model were used for analysis. After the data of each indicator are processed, the random forest model is first used to screen the indicators with high contribution values to the model, and then the remaining indicators are analyzed via the multiple linear regression model. To further obtain the influence of different factor indicators on players' salaries. Ultimately, the core determining factors for players' salaries are their offensive performance in the field and the competitiveness of the team. The impact of players' attendance on salaries is complex. The league system and commercial value have little effect on players' salaries. The current limitations of this study are reflected in the collinearity problem among the indicators, and the complex interaction mechanisms among some indicators have not yet been clarified. Further exploration is still needed to explore the mutual influence among the indicators and the influence mechanism of higher-order complex indicators on salary.

References

Breiman, L. (2001). Random forests. *Machine learning,* vol. 45**,** no. 1, pp. 5-32.

Cameron, A. C. and Trivedi, P. K. (2005). *Microeconometrics: Methods and Applications,* Cambridge, UK: Cambridge University Press.

Chen, L. Z. (2025). The efficiency evaluation of NBA players' performance on their salary creation. *Zhejiang Sport Science,* vol. 47**,** no. 03, pp. 34-45.

Chen, S., Cheng, C. and Chai, Y. (2015). A Study on the Correlation between Salary and Team Performance and Player Performance. *Sports***,** no. 19, pp. 33-35.

Gan, W. (2022). *Research on the Relationship between NBA Salary Cap System and NBA Clubs and Players' Salaries.* Master's Thesis, Beijing Sport University.

Kam, H. T. (1995). Published. Random decision forests. Proceedings of 3rd International Conference on Document Analysis and Recognition, 1995 Piscataway, NJ. IEEE, pp. 278-282.

Lyons, R., Jr., Jackson, E., Jr. and Livingston, A. (2015). Determinants of NBA player salaries. *The Sport Journal,* vol. 18, p. unpaginated ref. 28.

Sun, D. D. (2000). Selection of the linear regression model according to the parameter estimation. *Wuhan University Journal of Natural Sciences,* vol. 5**,** no. 4, pp. 400-405.

Xu, Z. and Xu, Z. (2026). Enhanced prediction of NBA players' salaries using hybrid ensemble models and multi-objective optimization techniques based on the 2022–23 season dataset. *Expert Systems with Applications,* vol. 295, p. 128849.

Zhang, M. (2015). A study on luxury tax in American professional sports. *Journal of Sports Culture***,** no. 2, pp. 142-145.

Zhao, J. and Li, C. (2018). American professional basketball league salary system research. *Journal of Sports Adult Education***,** no. 1, pp. 61-66.

Funding

This research received no external funding.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgment

This paper is an output of the science project.

Copyrights