



Journal of Economics and Business

Sawant, Pratik V., Upadhyaya, Nikita S., and Berger, Paul D. (2019), Identifying Future Brand Ambassadors in The National Basketball Association (NBA) for Predicting Future NBA Superstars for Superior Marketing. In: *Journal of Economics and Business*, Vol.2, No.1, 127-136.

ISSN 2615-3726

DOI: 10.31014/aior.1992.02.01.73

The online version of this article can be found at:
<https://www.asianinstituteofresearch.org/>

Published by:
The Asian Institute of Research

The *Journal of Economics and Business* is an Open Access publication. It may be read, copied and distributed free of charge according to the conditions of the Creative Commons Attribution 4.0 International license.

The Asian Institute of Research *Journal of Economics and Business* is a peer-reviewed International Journal. The journal covers scholarly articles in the fields of Economics and Business, which includes, but not limited to, Business Economics (Micro and Macro), Finance, Management, Marketing, Business Law, Entrepreneurship, Behavioral and Health Economics, Government Taxation and Regulations, Financial Markets, International Economics, Investment, and Economic Development. As the journal is Open Access, it ensures high visibility and the increase of citations for all research articles published. The *Journal of Economics and Business* aims to facilitate scholarly work on recent theoretical and practical aspects of Economics and Business.



ASIAN INSTITUTE OF RESEARCH
Connecting Scholars Worldwide



Identifying Future Brand Ambassadors in The National Basketball Association (NBA) for Predicting Future NBA Superstars for Superior Marketing

Pratik V. Sawant¹, Nikita S. Upadhyaya², Paul D. Berger³

^{1,2,3} Bentley University, Waltham, MA, USA

Abstract

This paper seeks to determine which factors/variables are useful to predict the salary of a National Basketball Association (NBA) player. Our goal is motivated by the presumed connection between a player's salary and how well he is known nationally and perhaps internationally, and the connection between how well a player is known and his value as a brand ambassador/spokesperson for selected brands. We use stepwise regression as the key statistical tool for our analyses.

Keywords: National Basketball Association (NBA), Brand Ambassador/Spokesperson, Stepwise Regression Analysis, NBA Data

INTRODUCTION

Since the advent of cable television, popular sports have occupied a major part of the entertainment spectrum. Sportspeople have commanded a fan following, second only to major Hollywood Stars. Many of these sportspeople serve as brand ambassadors/spokespeople for major brands or franchises. When a sports person becomes a face of the brand, the image of the brand often directly benefits or "takes a hit" based on the performance of the player in the sport. As such, brands do well to perform due diligence before selecting their ambassadors.

The most popular sports stars are often also the most highly compensated. Many popular stars represent brands and are compensated handsomely for it. As such, player compensation by their league is a good indicator of player success. If player success is related to certain parameters, it will be beneficial for brands to look at these parameters as a precursor to the success of a player as a way to judge who might be an effective ambassador/spokesperson for the brand. This may be able to help brands to identify superstars (presumably, *superior ambassadors/spokespeople*) early in their career.

In this paper, we have used the National Basketball Association (NBA) as a model for such predictive modeling. We use data from the NBA, to predict the performance parameters that are the best indicators of future player performance, as measured by salary.

Our core objective is to study the performance statistics of players in the National Basketball Association and try to determine the performance parameters which are most likely to impact their salaries. Tying these parameters to salary implies, to an extent, that NBA general managers and owners are capable of determining which parameters are *deserving* of more vs. less of a role in salary determination. NBA athletes are the most highly paid sportsmen in the world, as measured by average annual salary per player (Gaines, 2015). While overall performance is generally credited for the high salary of a player, we are attempting to breakdown the overall performance into individual components, to see if there are specific identifiable components of performance that makes certain players more attractive to clubs and to what degree, again, as measured by salary.

Our approach is aimed toward helping brands to identify players who are more likely to become superstars by looking at certain variables in their performance which have a linear relationship with their compensation. Brands can then use these variables to identify and tap talent early on in their career. In theory, those who perform better on the court are more likely to be more effective ambassadors/spokespersons for the brand. While we do not believe that this is *always* true – after all – there are other aspects of players, such as personality, city played in, and others, which likely also speak to effectiveness as a brand ambassador/spokesperson for a brand, we believe that *everything else equal*, superior play on the court leads to being a superior ambassador/spokesperson.

LITERATURE REVIEW:

The NBA has been of interest to researchers from various fields ranging from labor research to economics. Hajime Katayama and Hudan Nuch of the University of Sydney, in their paper, “A game level analysis of salary dispersion and team performance in national basket ball association,” have studied the impact of “within team” salary dispersion on team performance (Katayama, 2009).

Michael Wallace’s (1982) study titled, “Labor Market Structure and Salary Determination among Professional Basketball Players,” concluded that structural variables, such as team and race, play an important role in earnings determination in the NBA and that player position is not significant. Also, out of all performance indicators, he found only scoring and rebounding to be statistically significant. For the purpose of this study, we are not considering (the specific) team as a variable (Wallace, 1988).

Andrew Fleenor of the University of Tennessee created a prediction model for salaries of players in NBA (Fleenor, 1999). However, his study entitled, “Predicting National Basketball Association (NBA) Player salaries,” considers data only for the 1997-98 season. Since the scope of our paper is to identify superstars based on performance characteristics displayed consistently, we have considered total career statistics of players.

Jerry Hausman (MIT) and Gregory Leonard (Cambridge Econometric Inc.), in their paper, “Superstars in NBA: Economic Value and Policy,” conducted an econometric analysis demonstrating that TV ratings are higher for games in which certain superstars are playing. They concluded that the presence of superstars leads to inefficient distribution of talent (Hausman & Leonard, 1997).

There are also a large number of other studies which explored a wide array of topics such as racial differences in salaries and the impact of reputation and status on salaries in the NBA.

RESEARCH METHODOLOGY:

For the purpose of this study, we have used 4 datasets available in the public domain (retrieved from Kaggle datasets) and consolidated them into a single dataset using “player name” as a common factor. The 4 datasets available to us were:

- 1) Performance Statistics of 3922 players, who have played in the NBA since 1950. This data set gave us the individual statistics for every player for every season on 31 parameters.
- 2) Player data of 3922 players was obtained from an NBA database. This included the Name, Height, Weight, College, Year of Birth, Birth City and Birth State

- 3) Player data of 4550 players was obtained from an independent data set. This data set included the name, start year, end year, position, height, weight, birth date and college
- 4) Player Salaries for 2017-2018 season were obtained from the NBA records. These data were for 573 players who played in that season.

A master file for analysis was created by consolidating the data from these 4 sources into a single file. For this analysis, only the data of 573 players who played in the 17'-18' season was considered (file 4). The data from files 1,2 and 3 were looked up from those files. A pivot was used to consolidate the statistics against each player. Out of 573 players, complete data was obtained for only 479 players. Thus, we have used a dataset with a total of 479 data points (i.e., n=479.)

The final master file for analysis, thus prepared, contained -

- All player related data such as team, salary, height, weight, college, birth year, birth city, state, start year, end year, no. of years playing and position.
- Parameters of player performance (refer to Table 1). As the parameters were separate for each year, they had to be consolidated / aggregated in the final file. The technique used for consolidation / aggregation varied for each parameter. For numerical values such as Games Started (GS) and Minutes Played (MP), a sum of the values was done. For percentage values such as True Shooting Percentage (TS%) and Point Attempt Rate (3Par3), an average of the values was taken.

Table 1 summarizes the consolidation / aggregation techniques used:

Acronym	Full form	Method of Aggregation
GS	Games Started	Sum
MP	Minutes Played	Sum
PER	Player Efficiency Rating	Average
TS%	True Shooting Percentage	Average
3Par3	Point Attempt Rate	Average
FTR	Free Throw Rate	Average
ORB%	Offensive Rebound Percentage	Average
DRB%	Defensive Rebound Percentage	Average
TRB%	Total Rebound Percentage	Average
AST%	Assist Percentage	Average
STL%	Steal Percentage	Average
BLK%	Block Percentage	Average
TOV%	Turnover Percentage	Average
USG%	Usage Percentage	Average
OWS	Offensive Win Shares	Sum
DWS	Defensive Win Shares	Sum
WS	Win Shares	Sum
WS/48	Win Shares per 48 mins	Sum
OBPM	Offensive Box Plus Minus	Average
DBPM	Defensive Box Plus Minus	Average
BPM	Box Plus Minus	Average
VORP	Value Over Replacement	Average
FG	Field Goals	Sum
FGA	Field Goal Attempts	Sum

FG%	Field Goal Percentage	Average
3P3	Point Field Goals	Sum
3P3A	Point Field Goals Attempts	Sum
3P%3	Point Field Goal Percentage	Average
2P2	Point Field Goals	Sum
2PA2	Point Field Goals Attempts	Sum
2P%2	Point Field Goal Percentage	Average
eFG%	Effective Field Goal Percentage	Average
FT	Free Throws	Sum
FTA	Free Throw Attempts	Sum
FT%	Free Throw Percentage	Average
ORB	Offensive Rebounds	Sum
ORB	Offensive Rebounds	Sum
DRB	Defensive Rebounds	Sum
TRB	Total Rebounds	Sum
AST	Assists	Sum
STL	Steals	Sum
BLK	Blocks	Sum
TOV	Turnovers	Sum
PFP	Personal Fouls	Sum
PTS	Points	Sum

Table 1: Consolidation / Aggregation techniques

Qualitative Variables and Recoding

We had 47 independent variables, of which, 46 were quantitative and 1 was qualitative / categorical. The categorical variable was player position. There are 7 (n=7) player positions (C, C-F, F, F-C, F-G, G, G-F). Guard Forward (G-F) was taken as the base/dummy category and 6 (i.e., [n-1]) dummy variables were created. (C = Center, F = Forward, G = Guard; many players can play/two positions, which is why we have C-F, F-C, F-G and G-F, where the first letter is his majority position.)

Dependent and Independent Variables:

Our dependent variable (Y) is Salaries of Players (2017-2018.) We have 52 quantitative independent variables, of which 6 are dummy variables, as noted in Table 2.

X	Variable	Full name
X1	Height	Height
X2	Weight LBS	Weight
X3	No. of years played	No. of years played
X4	Games Started	Games Started
X5	MP	Minutes Played
X6	PER	Player Efficiency Rating
X7	TS%	True Shooting Percentage
X8	3Par	Point Attempt Rate
X9	FTr	Free Throw Rate

X10	ORB%	Offensive Rebound Percentage
X11	DRB%	Defensive Rebound Percentage
X12	TRB%	Total Rebound Percentage
X13	AST%	Assist Percentage
X14	STL%	Steal Percentage
X15	BLK%	Block Percentage
X16	TOV%	Turnover Percentage
X17	USG%	Usage Percentage
X18	OWS	Offensive Win Shares
X19	DWS	Defensive Win Shares
X20	WS	Win Shares
X21	WS/48	Win Shares per 48 mins
X22	OBPM	Offensive Box Plus Minus
X23	DBPM	Defensive Box Plus Minus
X24	BPM	Box Plus Minus
X25	VORP	Value Over Replacement
X26	FG	Field Goals
X27	FGA	Field Goal Attempts
X28	FG%	Field Goal Percentage
X29	3P	Point Field Goals
X30	3PA	Point Field Goals Attempts
X31	3P%	Point Field Goal Percentage
X32	2P	Point Field Goals
X33	2PA	Point Field Goals Attempts
X34	2P%	Point Field Goal Percentage
X35	eFG%	Effective Field Goal Percentage
X36	FT	Free Throws
X37	FTA	Free Throw Attempts
X38	FT%	Free Throw Percentage
X39	ORB	Offensive Rebounds
X40	DRB	Defensive Rebounds
X41	TRB	Total Rebounds
X42	AST	Assists
X43	BLK	Blocks
X44	TOV	Turnovers
X45	PF	Personal Fouls
X46	PTS	Points
X47	C	Dummy
X48	CF	Dummy
X49	F	Dummy
X50	FC	Dummy
X51	FG1	Dummy
X52	G	Dummy

Table 2: Based on previous research, and our knowledge of the game, we have the following hypotheses:

H_0 : Position, No. of years played, player efficiency rating, true shooting percentage and field goals are not related to salary.

H_1 : Position, No. of years played, player efficiency rating, true shooting percentage and field goals **are** related to salary.

We expect to reject H_0 overwhelmingly.

REGRESSION ANALYSIS & DISCUSSION OF RESULTS:

Our data are used to conduct stepwise multiple linear-regression analysis using 2017-2018 Salary as the dependent variable to determine how it is related to the independent variables. The variables that are part of our final stepwise model are Value over Replacement (VORP), Games Started, Defensive Win shares (DWS), Offensive Rebounds (ORB), Offensive Rebounds Percentage (ORB%), Player Efficiency Rating (PER), Center Position, and Two Point Average (2PA). The R-square of this model is 0.47; this means that 47% of the variability in salary is estimated to be explained by the variables included in this model. The details of the model summary are shown below in Figure 1 (in essence, the last step of the stepwise-regression analysis.)

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.636 ^a	0.404	0.403	5795278.
2	.640 ^b	0.410	0.408	5773369.
3	.646 ^c	0.417	0.413	5746623.
4	.662 ^d	0.438	0.433	5647230.
5	.669 ^e	0.448	0.442	5602822.
6	.668 ^f	0.446	0.441	5607263.
7	.672 ^g	0.452	0.446	5581319.
8	.676 ^h	0.457	0.450	5561574.
9	.681 ⁱ	0.464	0.456	5531695.
10	.685 ^j	0.470	0.461	5508773.
a. Predictors: (Constant), VORP				
b. Predictors: (Constant), VORP, AST				
c. Predictors: (Constant), VORP, AST, Games Started				
d. Predictors: (Constant), VORP, AST, Games Started, DWS				
e. Predictors: (Constant), VORP, AST, Games Started, DWS, ORB				
f. Predictors: (Constant), VORP, Games Started, DWS, ORB				
g. Predictors: (Constant), VORP, Games Started, DWS, ORB, ORB%				
h. Predictors: (Constant), VORP, Games Started, DWS, ORB, ORB%, PER				
i. Predictors: (Constant), VORP, Games Started, DWS, ORB, ORB%, PER, C				
j. Predictors: (Constant), VORP, Games Started, DWS, ORB, ORB%, PER, C, 2 2PA				

Figure 1

It should be noted that Figure 1 contains 8 independent variables, even though there are 10 steps. One variable, AST, entered the model early on, but was then deleted, as other variables entered the model and, while adding predictive ability, eventually rendered AST redundant, and, consequently, AST was deleted as not adding any incremental value. This phenomenon (of a variable entering the model and subsequently being deleted from the model) is not an infrequent one in the stepwise-regression process.

The coefficients in the model are depicted in Figure 2:

Coefficients ^a								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
10	(Constant)	3118879.3	833589.7		3.7	0.0		
	VORP	4274633.8	367026.2	0.7	11.6	0.0	0.4	2.8
	Games Started	9694.8	3010.5	0.4	3.2	0.0	0.1	10.8
	DWS	-335740.8	62194.5	-0.6	-5.4	0.0	0.1	10.0
	ORB	5432.8	970.2	0.4	5.6	0.0	0.2	5.1
	ORB%	-446181.9	104755.3	-0.2	-4.3	0.0	0.4	2.8
	PER	253843.5	80063.8	0.2	3.2	0.0	0.4	2.5
	C	2231316.8	951821.7	0.1	2.3	0.0	0.7	1.4
	2PA	-521.6	235.0	-0.2	-2.2	0.0	0.1	10.5

a. Dependent Variable: Salary in 2017-2018 Season

Figure 2

Multicollinearity Analysis

With this output, the VIF value was found to be higher than 10 for 3 variables - Games Started, DWS and 2PA. As such, these variables have a possible multicollinearity issue. To address this issue, we will include just one of these 3 variables and run the regression again. We conducted the stepwise multiple linear regression excluding the following variables: DWS and Games Started (GS).

The new output of our stepwise regression is shown below in Figure 3:

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.636 ^a	0.404	0.403	5795278.96
2	.640 ^b	0.410	0.408	5773369.03
3	.645 ^c	0.417	0.413	5748028.05
4	.650 ^d	0.422	0.417	5727288.32
5	.654 ^e	0.427	0.421	5706840.13
6	.661 ^f	0.437	0.430	5665467.96
7	.666 ^g	0.444	0.436	5635240.97
8	.671 ^h	0.450	0.441	5609815.77
9	.682 ⁱ	0.465	0.455	5540361.06
10	.688 ^j	0.474	0.462	5500492.08
a. Predictors: (Constant), VORP				
b. Predictors: (Constant), VORP, AST				
c. Predictors: (Constant), VORP, AST, 3P				
d. Predictors: (Constant), VORP, AST, 3P, 3Par				
e. Predictors: (Constant), VORP, AST, 3P, 3Par, OBPM				

f. Predictors: (Constant), VORP, AST, 3P, 3PAr, OBPM, ORB%
g. Predictors: (Constant), VORP, AST, 3P, 3PAr, OBPM, ORB%, C
h. Predictors: (Constant), VORP, AST, 3P, 3PAr, OBPM, ORB%, C, ORB
i. Predictors: (Constant), VORP, AST, 3P, 3PAr, OBPM, ORB%, C, ORB, WS/48
j. Predictors: (Constant), VORP, AST, 3P, 3PAr, OBPM, ORB%, C, ORB, WS/48, BLK

Figure 3

The variables in this model includes Value Over Replacement (VORP), Assists (AST), Point Field Goals Attempts(3PA), Point Attempt Rate (3PAr), Offensive Box Plus Minus (OBPM), Offensive Rebound Percentage (ORB%), Center Position C, Offensive Rebounds (ORB), Winshares per 48 mins (WS/48), and Blocks (BLKs). The R-square of this model has slightly improved to 0.474. This means that we estimate that 47.4% of the variability in the Salary is explained by the variables included in this model.

The coefficients in the model are depicted in Figure 4:

Coefficients ^a								
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
10	(Constant)	9069609.779	1205728.651		7.52	0.000		
	VORP	4115288.497	342327.192	0.631	12.02	0.000	0.409	2.448
	AST	-1193.123	328.407	-0.233	-3.63	0.000	0.275	3.641
	3P	2975.381	940.738	0.181	3.163	0.002	0.344	2.906
	3Par	-7004321.347	2145036.236	-0.176	-3.27	0.001	0.387	2.582
	OBPM	545674.469	131518.162	0.210	4.15	0.000	0.440	2.275
	ORB%	-377105.097	107641.257	-0.203	-3.50	0.001	0.334	2.991
	C	2531639.073	976798.188	0.104	2.59	0.010	0.696	1.437
	ORB	4660.124	953.811	0.365	4.89	0.000	0.202	4.948
	WS/48	-2528740.408	839471.712	-0.235	-3.01	0.003	0.184	5.433
BLK	-4337.377	1550.695	-0.190	-2.80	0.005	0.245	4.086	

a. Dependent Variable: Salary in 2017-2018 Season

Figure 4

The Stepwise Multiple Linear Regression equation is as follows:

$$Y\text{-predicted} = 9069609.779 + 4115288.497 \cdot \text{VORP} - 1193.123 \cdot \text{AST} + 2975.381 \cdot \text{3P} - 7004321.347 \cdot \text{3Par} + 545674.469 \cdot \text{OBPM} - 377105.097 \cdot \text{ORB\%} + 2531639.073 \cdot \text{C} + 4660.124 \cdot \text{ORB} - 2528740.408 \cdot \text{WS/48} - 4337.377 \cdot \text{BLK}$$

Interpretation of the model

To interpret the intercept of the model, if all variables in the model were zero for an NBA player, that NBA player would be paid a base salary of about \$9 million annually. We are not “married” to this number, since often, an intercept is not a very accurate value, especially when there are no data values “near” the premise of every single variable having a zero value.

From our model, we can predict that the salary of the player depends on Value Over Replacement (VORP), Assists (AST), Point Field Goals Attempts (3PA), Point Attempt Rate (3PAr), Offensive Box Plus Minus (OBPM),

Offensive Rebound Percentage (ORB%), Offensive Rebounds (ORB), Winshares per 48 mins (WS/48), and Blocks (BLKs) among all the quantitative variables. Center (C) is also a variable that is predictive of salary, given all other variables are held constant.

Player compensation increases with an increase in value over replacement. This makes sense because VORP measures the marginal utility of a player to his team, compared to a replacement level player (Woolner 1997). Player compensation increases with increase in point field goals. This was expected since the field goal is one of the most popular metrics of measurement of player success. Player compensation increases with Offensive Box Plus Minus and Offensive Rebound. This also makes sense because Offensive Box plus minus is a metric for evaluating a basketball player's quality and contribution to the team while an offensive rebound gives the offensive team another opportunity to score right away. (The majority of rebounds are defensive, since the defending team is generally in a better position to recover the ball; the shooter is generally not positioned well to get a rebound of his own missed shot, so that, at minimum, there are 5 defensive players, and only 4 offensive players, positioned to get the rebound.) As such, a player with ability to get offensive rebounds is valuable.

Player compensation decreases with increase in assists, point attempt rate, offensive rebound percentage, Winshares per 48 minutes, and blocks; these are the variables with a negative coefficient in the final model. None of these negative coefficients make very clear sense. In some cases, one might make a "partial argument" (e.g., assists are distributed quite liberally, and *perhaps*, is less reflective of true contribution), but overall, we admit that we are not prepared to support a logical explanation for these negative coefficients. Since this is a stepwise regression with no "very high" VIF values, none of the variables should be highly correlated with other variables. If they were, not all the variables would "survive" in the final model – redundancy would suggest/mandate some deletions. Thus, we cannot "blame" multicollinearity for the negative values.

The qualitative variable, Center Position (C) is also statistically significant, which essentially means that (we estimate that) a player who plays the center position receives approximately \$2.5 million more in salary annually compared to players who play in the Guard-Forward position (the "dummy" category), when all other variables in the model are held constant.

CONCLUSION

Our study concludes that brands can effectively identify future superstars by looking at certain performance parameters. A player who has a high VORP, scores a "larger" number of field goals, has a higher box plus minus, and has more offensive rebounds, fewer assists, has a lower point attempt rate, winshares and blocks is most likely to be successful, salary-wise. As mentioned earlier in this paper, this study can help advertising agencies and brands to identify promising prospects from a pipeline of future players.

LIMITATIONS AND SCOPE OF FUTURE RESEARCH:

In an ideal situation, the VIF values of variables should have been in the range of 1 to 2. However, the nature of the data is such that many variables are related to each other. As such, we have relaxed this requirement and excluded only those variables that are very strongly correlated with each other ($r > 0.8$.)

We have excluded certain qualitative variables such as college/school of a player, and city and/or country of birth of each player. The cap on salaries of players in the form of cap of maximum allowable expenditure on player salaries by a team is another factor which may distort the market forces at work. Certain adjustments made for those factors could change the model output.

REFERENCES

- Berri, D.J. (1999), Who is the Most Valuable? Measuring Players' Production of Wins in the National Basketball Association
- Fleenor, A.T. (1999), Predicting National Basketball Association (NBA) Player Salaries, Tennessee Research and Creative Exchange, Spring 1999
- Gaines,C.(2015), The NBA is the Highest paying Sports League in the World, Business Insider
- Hausman, A.J., Leonard,G.(1997), Superstars in National Basketball Association, Economic Value and Policy, Journal of Labor Economics, Vol.15, No.4
- Katayama, H., Nuch, H.(2011), A Game Level Analysis of Salary Dispersion and Team Performance in NBA, Applied Economics, 43, 1193-1207
- Matthew, S.D., (1997), Racial Differences in National Basketball Association Players' Salaries - A New Look, American Economist, Pg.84, 41,2, ProQuest Central
- Wallace, M. (1988), Labour Market Structure and Salary Determination among Professional Basketball Players, Work and Occupations, Vol 15, No. 3, 294-312.