Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(6):2400-2405

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Principal component factor analysis-based NBA player comprehensive ability evaluation research

Wei Yin

Department of Sport, Shandong Sport University, Jinan, Shandong, China

---

ABSTRACT

A basketball player's ability not only is directly related to how many scores he get in the field, but also has correlation whether he can help team to win, number of shooting release in the field, number of fault, number of rebound and number of foul .In order to discuss player ability evaluation model, the paper mainly applies multiple analysis's principal component analysis and factor analysis, with the help of SPSS software to analyze data, starts from measuring players' technical level's score, assist, field-goal percentage and others total 10 indicators, and gets each indicator and common factor expression. Use factor analysis to make evaluation analysis on season 2011 to 2012 eight NBA teams' active service players' comprehensive abilities, gets players' abilities comprehensive indicator model, calculates every player comprehensive score. Make quadratic nonlinear regression on players' obtained salary and personal ability, use MATLAB software to fit the two functional relationships. Make comparative analysis of calculated due value and actual obtained value, get estimated values errors, and then put forward relative reasonable explanation.

Key words: factor analysis, ability indicator, regression analysis, basketball competition, evaluation model

---

INTRODUCTION

Kobe, Stoudemire, Dirk Nowitzki and other players are brilliant starts in NBA league, and it is nothing wrong that they can obtain several ten million annual salary at every turn .But in a statistics made by economics professor David Pele from Southern Utah University recent days, he got that Kobe, Stoudemire, Nowitzki and others actually belonged to presentation of overpaid [1-5]. They earnings and performance cannot be in direct proportion.

From competition result, it cannot reflect players values, is impossible to evaluate players' ability value. In recent years, with TrueSkill model being put forward, introduced the concept of player ability evaluation, through learning players' ability value, it makes prediction on confrontation two parties scores status, player ability value learning process adopted Bayes deduction method, what TrueSkill model used was Expectation Propagation[1, 2] algorithm, verified by experiment, its prediction accuracy was 64.42%.But it has trained players ability value, it only has a player ability value variable, and all attack and defense conditions in field were shared and relied on the variable [6-11]. Xue Hui by comprehensive analyzing NBA players each item ability, he established a kind of comprehensive technical indicator that could reflect players efficiency, and established income and ability regression model, explored players income and their abilities relationships[3]. Wang Cang-You applied RSR rank-sum ratio, normal distribution principle and other statistical method to analyze season 2009 to 2012 totally 97 CBA foreign players basic information and competition abilities [4-8]. The paper mainly applies multiple analyses' principle component analysis and factor analysis, with the help of SPSS software to analyze data, and gets players' ability comprehensive indicator model.

FACTOR ANALYSIS MODEL ESTABLISHMENTS

For player ability and score, rebound, assist, block shot, steal, fault and others ten items personal data, the paper adopts factor analysis to analyze them. Considering NBA has numerous teams, and every team staff composition has no big difference, so the paper selects ten players located eight teams to analyze, in the following, it takes Nets as an example, solves players' comprehensive ability indicator [12, 13]. Factor analysis steps in $SPSS$ are like following:

In order to define factor analysis applicability, we adopt KMO and spherical Bartlett test. KMO tests whether players' indicators partial correlation is smaller or not, Bartlett spherical test is judging whether correlation matrix is unit matrix or not, it can refer to Table 1.

Table 1: KMO and Bartlett test

Sampling enough measure Kaiser-Meyer-Olkin measurement		.797
Bartlett sphericity test	approximate Chi-square	266.476
	df	45
	Sig.	.000

By Bartlett test, it is clear that player indicators have stronger correlation, and KMO statistical amount is 0.797 that is above 0.7, which shows each indicator's information overlapping level is higher.

By Table 2 showed common factor variance, it is clear: each common factor that is extracted nearly is above 80%, therefore the extracted common factors explanatory ability on each variable is stronger. That means each indicator that is extracted has higher evaluation degree on player comprehensive ability.

Table 2: Common factor variance

	Initial	Extract
Score	1.000	.943
Rebound	1.000	.978
Assist	1.000	.911
Steal	1.000	.883
Block shot	1.000	.937
Field-goal percentage	1.000	.816
Free throw percentage	1.000	.795
Number of faults	1.000	.953
Games played	1.000	.763
Playing time（ minute）	1.000	.969
Extraction method: Principal component analysis.		

By following Table 3, it is clear that for output result, only the former three feature roots are above 1, the former three factors' variance contribution rate is 89.481%, therefore it selects the former three factors is enough to describe players' comprehensive ability level.

Table 3: Explanatory total variance

Component	Initial feature value			Extract squares sum and input			Rotate squares sum and input		
	Total	Variance %	Accumulation %	Total	Variance %	Accumulation %	Total	Variance %	Accumulation %
1	6.419	64.186	64.186	6.419	64.186	64.186	4.171	41.705	41.705
2	1.362	13.620	77.806	1.362	13.620	77.806	3.177	31.768	73.474
3	1.167	11.675	89.481	1.167	11.675	89.481	1.601	16.007	89.481
4	.492	4.918	94.399						
5	.328	3.283	97.682						
6	.111	1.115	98.796						
7	.089	.887	99.683						
8	.015	.148	99.831						
9	.011	.109	99.940						
10	.006	.060	100.000						
Extraction method: Principal component analysis.									

Scree plot also further indicates each factor importance degree that can refer to Figure 1. It is clear the former three factors scattering points locates in steep hill, and the later seven factors scattering points become the platform while all feature roots are less than 1, therefore only need to consider former three factors at most.

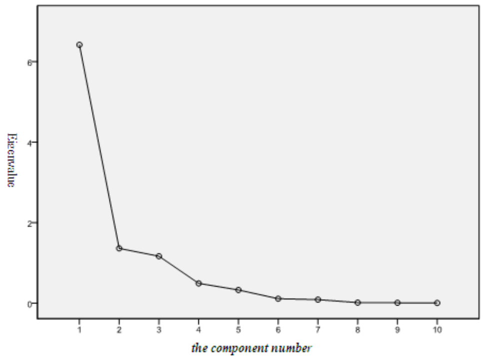

Figure 1: Scree plot

As following Table 4, it shows each factor to each player indicator variable impact.

Table 4: Component matrix

	Component		
	1	2	3
Playing time (minute)	.979	-.058	-.083
Score	.947	-.161	.143
Steal	.927	-.153	-.012
Number of faults	.925	-.241	.197
Games played	.867	-.009	-.105
Rebound	.817	.252	-.497
Block shot	.765	.312	-.505
Assist	.732	-.456	.408
Field-goal percentage	.298	.795	.309
Free throw percentage	.466	.501	.572
Extraction method: Principal component analysis.			
a. Already extracted three components.			

Player each indicator ability model is as following:

$$ZX_1 = 0.947F_1 - 0.161F_2 + 0.143F_3 + \varepsilon_1$$
$$ZX_2 = 0.817F_1 + 0.252F_2 - 0.497F_3 + \varepsilon_2$$
$$ZX_3 = 0.732F_1 - 0.456F_2 + 0.408F_3 + \varepsilon_3$$
$$ZX_4 = 0.927F_1 - 0.153F_2 - 0.012F_3 + \varepsilon_4$$
$$ZX_5 = 0.765F_1 + 0.312F_2 - 0.505F_3 + \varepsilon_5$$
$$ZX_6 = 0.298F_1 + 0.795F_2 + 0.309F_3 + \varepsilon_6$$
$$ZX_7 = 0.466F_1 - 0.501F_2 + 0.572F_3 + \varepsilon_7$$
$$ZX_8 = 0.925F_1 - 0.241F_2 + 0.197F_3 + \varepsilon_8$$
$$ZX_9 = 0.867F_1 - 0.009F_2 - 0.105F_3 + \varepsilon_9$$
$$ZX_{10} = 0.979F_1 - 0.058F_2 - 0.083F_3 + \varepsilon_{10}$$

Among them: ZXi represents the i indicator individual ability contribution; F_i represents the i common factor; ε_i represents the i extrinsic factor.

In expression, for each indicator variable after standardization, ε_i represents special factor, is the other factor affects the variable except for the three common factors. Originally, it designs ten indicators to show players' comprehensive ability level, but after factor analysis, only needs three factors then can describe player comprehensive ability level influence status.

In the paper, it adopts variance maximum orthogonal rotation method to make factor rotation, after proceeding with maximum variance rotation, factor load matrix after rotation is as Table 5 show.

Table 5: Rotational component matrix

	Component		
	1	2	3
Assist	.953	.011	.043
Number of faults	.899	.351	.144
Score	.852	.429	.182
Steal	.769	.531	.098
Playing time（ minute）	.730	.644	.146
Games played	.611	.607	.146
Rebound	.278	.941	.122
Block shot	.204	.934	.153
Field-goal percentage	-.061	.213	.876
Free throw percentage	.325	.033	.829
Extraction method: Principal component analysis. Rotation method: Orthogonal rotation method with Kaiser standardization.			
a. Rotation makes convergences after five times iteration.			

By Table 5, it is clear that the first common factor has larger loading in $X_{1.}$、X_3、X_4、X_8、X_9、X_{10}, it mainly reflects player attack ability from score, assist, steal, fault, games played and playing time these aspects, which can be named as attack factors. The second common factor has larger loading X_2、X_5, it reflects player defense ability from rebound and block shot aspects, therefore named them as defense factors. The third common factor has larger loading in X_6、X_7, it shows as field-goal percentage and free throw percentage, therefore named them as stable factors. It roughly conforms to practical status, each common factor significance is relative reasonable.

Factor score: Common factor score coefficient function cannot be got by factor load matrix through matrix transformation method, but only can be solved by adopted estimation method; the paper adopts regression method, express common factors into each variable linear form. Factor score coefficient matrix is as Table 6 show.

Table 6: Component score coefficient matrix

	Component		
	1	2	3
Score	.222	-.035	.010
Rebound	-.176	.444	-.059
Assist	.403	-.288	-.043
Steal	.162	.061	-.060
Block shot	-.207	.458	-.030
Field-goal percentage	-.159	.023	.622
Free throw percentage	.066	-.192	.582
Number of faults	.269	-.089	-.011
Games played	.068	.146	-.022
Playing time (mi nute)	.107	.132	-.037
Extraction method: Principal component analysis. Rotation method: Orthogonal rotation method with Kaiser standardization. Constitute the score.			

It can directly write down each common factor score model:

$$F_1 = 0.222ZX_1 - 0.176ZX_2 + 0.403ZX_3 + 0.162ZX_4 - 0.207ZX_5$$
$$-0.159ZX_6 + 0.066ZX_7 + 0.266ZX_8 + 0.068ZX_9 + 0.107ZX_{10}$$
$$F_2 = -0.035ZX_1 + 0.444ZX_2 \text{ - } 0.288ZX_3 + 0.061ZX_4 + 0.458ZX_5$$
$$+0.023ZX_6 - 0.192ZX_7 \text{ - } 0.089ZX_8 + 0.146ZX_9 + 0.132ZX_{10}$$
$$F_3 = 0.010\mathrm{ZX}_1\text{-}0.059\mathrm{ZX}_2\text{-}0.043\mathrm{ZX}_3 - 0.060\mathrm{ZX}_4 - 0.030\mathrm{ZX}_5$$
$$+0.622\mathrm{ZX}_6 + 0.582\mathrm{ZX}_7 - 0.011\mathrm{ZX}_8\text{-}0.022\mathrm{ZX}_9 - 0.037\mathrm{ZX}_{10}$$

SPSS has already put forward three common factors' scores, saved them in fac_1~fac_3, according to each factor corresponding variance contribution rate as weights, calculate following comprehensive statistics:

$$F = \frac{\lambda_1}{\lambda_1+\lambda_2+\lambda_3}F_1 + \frac{\lambda_2}{\lambda_1+\lambda_2+\lambda_3}F_2 + \frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3}F_3$$

$= 0.718F_1 + 0.152F_2 + 0.130F_3$

In SPSS, use program to calculate comprehensive factor score model:

Nets: $Comp\ score = 0.718 * fac1\_1 + 0.152 * fac2\_1 + 0.130 * fac3\_1$

According to above principles, similarly, we can solve following seven teams' comprehensive factors score model:

Mavericks: $Comp\quad score = 0.625 * fac1\_2 + 0.255 * fac2\_2 + 0.120 * fac3\_2$
Wizards: $Comp\quad score = 0.705 * fac1\_3 + 0.177 * fac2\_3 + 0.118 * fac3\_3$
Lakers: $Comp\quad score = 0.710 * fac1\_4 + 0.173 * fac2\_4 + 0.117 * fac3\_4$
Knicks: $Comp\quad score = 0.651 * fac1\_5 + 0.231 * fac2\_5 + 0.118 * fac3\_5$
Bobcats: $Comp\quad score = 0.738 * fac1\_6 + 0.262 * fac2\_6$
Hornets: $Comp\quad score = 0.738 * fac1\_7 + 0.262 * fac2\_7$

By above model, it can respectively calculate each team every player comprehensive score.

PLAYER ABILITY AND PLAYER OBTAINED SALARY RELATIONSHIP MODEL
By analysis, it is clear that player obtained salary high-low is closely related to player himself comprehensive ability, by analyzing mastered data, we establish salary and comprehensive ability regress model. Similarly, we take nets as an example, use MATLAB function to make quadratic fitting and get Figure 2.

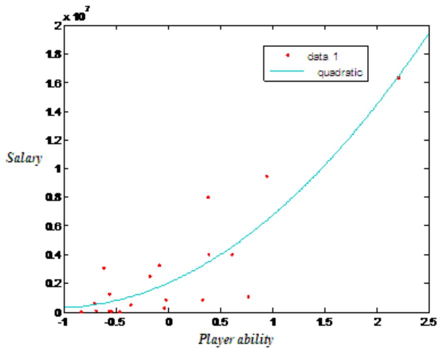

Figure 2: Salary and comprehensive ability fitting curve

That: $f(x) = p1 * x^2 + p2 * x + p3$

Among them:

p1 = 1.499e+006 confidence interval is (2.044e+005, 2.794e+006)
p2 = 3.201e+006 confidence interval is (1.361e+006, 5.041e+006)
p3 = 2.026e+006 confidence interval is (8.46e+005, 3.207e+006)

R-square: 0.7829 Adjusted R-square: 0.7574

Due to in NBA field, many players are hard to avoid trouble of injury and diseases, which affects their playing time, score, rebound and other abilities, it also directly causes their comprehensive abilities to be lower, however it will not affect their salary in this season, so we fit and get that function fitting degree as 78.3% is reasonable. According to that, we can get every player deserved salary. Below Table 7 lists ten players' actual salary and deserved salary, and make comparison of the two:

Table 7: Ten players' actual salary and deserved salary

Player	Player rank in list	Actual consulted salary	Deserved salary according to model calculation	Additional salary by calculating	Additional salary in rank	Calculation and actual additional parts differences
Rashard Lewis	1	21136631	1458542.851	19678088.149	21167231	1489143
Kobe Bryant	2	25244493	9646284.171	15598208.829	19693258	4095049
Antawn Jamison	3	15076715	6257564.571	8819150.429	17402350	8583200
Amare Stoudemire	4	18217705	4231757.002	13985947.998	14918309	932361
Chris Karman	5	14030000	8313528.642	5716471.358	14613480	8897009
Corey Maggette	7	10262069	4546489.131	5715579.869	12862248	7146668
Dirk Nowitzki	8	19092873	4426000.000	14666873.000	12851295	-1815578
Deron Williams	9	16359805	12461790.076	3898014.924	12784867	8886852
Tyrus Thomas	10	7305785	1941601.922	5364183.078	12459225	7095042

Note: Play No.6 is not taken into consideration here because he is free player. In order to more clearly show the two relationships, we use EXCEL to draw, as following Figure 3.

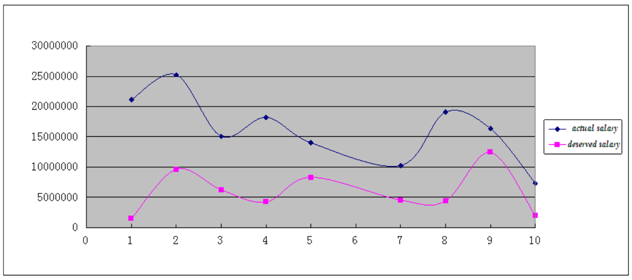

Figure 3: Player actual salary and deserved salary curve graph

According to above Figure 3, we can clearly see that former ranking players' actual salary and deserved salary gap is larger, while the two gaps gradually reduces with the later ones, which shows the overpaid gets more serious while ranks in the former.

CONCLUSION

The paper adopts factor analysis, better integrates player each ability variable, especially considers many effects impacting, selects nets data as center point, and gets verification by other teams. Finally it also takes errors analysis, result relative conforms to practice. But the shortcoming in the paper is that it only selects regular seasons, player' rebound ability and some teams eliminate partial players.

REFERENCES

[1] CHEN Jian, YAO Song-ping. *Journal of Shanghai Physical Education Institute*, **2009**, 33(5).
[2] Wang Luning et al. *China Sport Science*, **1999**, 19(3), 37-40.
[3] MAO Jie, SHAN Shuguang. *Journal of Wuhan Institute of Physical Education*, **2012**, 46(2), 70-73, 87.
[4] ZHANG Lei. *China Sport Science and Technology*, **2006**, 42(1), 50-52.
[5] FU Fan-fei, YU Zhen-feng, ZHANG Quan-ning. *Shandong Sports Science & Technology*, **2006**, 28(2), 24-25.
[6] Wang Luning et al. *China Sport Science*, **1999**, 19(3), 37-40.
[7] LI Ji-hui. *Journal of Shenyang Sport University*, **2006**, 25(2), 63-66.
[8]YANG Yue-qing, RAO Han-fu, LEN Ji-lan. *Journal of Hubei Sports Science*, **2003**, 22(2), 204-205.
[9] Xiaomin Zhang. *Journal of Chemical and Pharmaceutical Research*, **2013**, 5(12), 8-14.
[10] Wang Bo; Zhao Yulin. *Journal of Chemical and Pharmaceutical Research*, **2013**, 5(12), 21-26.
[11] Mingming Guo. *Journal of Chemical and Pharmaceutical Research*, **2013**, 5(12), 64-69.
[12] Bing Zhang; Zhang S.; Lu G.. *Journal of Chemical and Pharmaceutical Research*, **2013**, 5(9), 256-262.
[13] Bing Zhang. *Journal of Chemical and Pharmaceutical Research*, **2014**, 5(2), 649-659.