

1. Introduction

Every professional athlete will say “I have myself fully devoted to my club”. But is that true? For many basketball players in the history of NBA (National Basketball Association), their professionalism is a long-debating issue. There are always sayings that some players only play at high level in the seasons prior to their free agency eligibility, after then their placements and salaries will be finalized. But once they sign the new contract, their performance return to the previous level. By constructing this project, we aim to uncover the mystery of such “contract year phenomenon”.

This project firstly examines in general the existence of statistically significant difference between contract year performance and non-contract year performance. Then we will construct a “slippery index”, an indicator attempting to quantify the contract year phenomenon for each individual player. And we will build a predictive model to evaluate whether we can use such index to predict the contract salary.

2. Expletory Data Analysis

The dataset we used was downloaded from <https://www.basketball-reference.com/>, a centralized database for NBA player statistics, including their yearly performance and salary. We randomly picked 100 players (50 from the East and 50 from the West) having complete league experience over 5 seasons with at least one contract year season. The performance statistics is measured by 23 features (Table 1):

(1) Minutes played per game
(2) Field goals per game
(3) Field goal attempts per game
(4) Field goals percentage

(5) 3_Points field goals per game
(6) 3_Points field goal attempts per game
(7) 3_Points field goal percentage
(8) 2_point field goals per game
(9) 2_point field goal attempts per game
(10) 2_point field goal percentage
(11) Effective field goal percentage
(12) Free throws per game
(13) Free throws attempts per game
(14) Free throws percentage
(15) Offensive rebounds per game
(16) Defensive rebounds per game
(17) Total rebounds per game
(18) Assists per game
(19) Steals per game
(20) Blocks per game
(21) Turnovers per game
(22) Personal fouls per game
(23) Points per game

Table 1: player performance statistics features

Sanity check and data cleaning:

We firstly examined the dataset generally. Totally we had 100 players’ data with 989 season sample, average each player 9.89 seasons. The data occurred several minor data missing at features “3_Points field goal percentage” and “Free throws percentage” (Figure 1). After sanity check, the missing was due to the absence of 3 points field goal and free throw attempt, thus the percentage ratios were undefined in the dataset. We filled the missing values with 0.

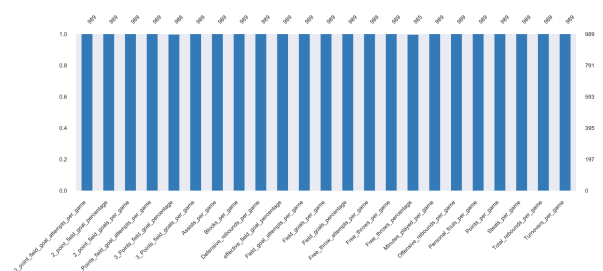


Figure 1: data missing sanity check
Feature distribution and correlation

By visualization, we examined the general distribution of some important features including “minutes played per game” (Figure 2), “points per data” (Figure 3) and “total rebounds per game” (Figure 4).

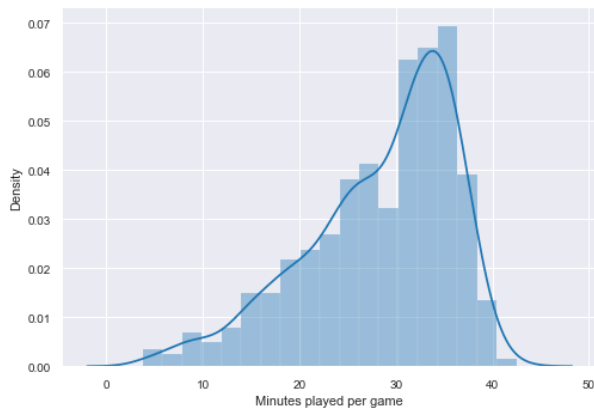


Figure 2: minutes played per game distribution

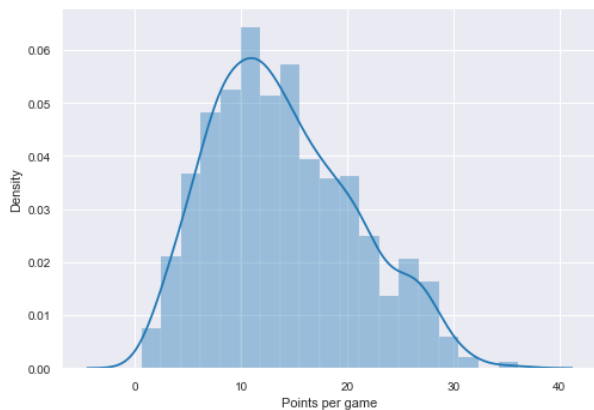


Figure 3: points per game distribution

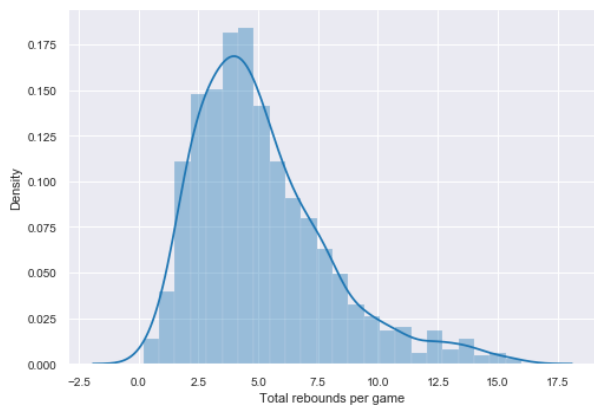


Figure 4: total rebounds per game

We see that the distributions were generally

normal with skewness. And we plot the correlation matrix for all performance features (Figure 5).

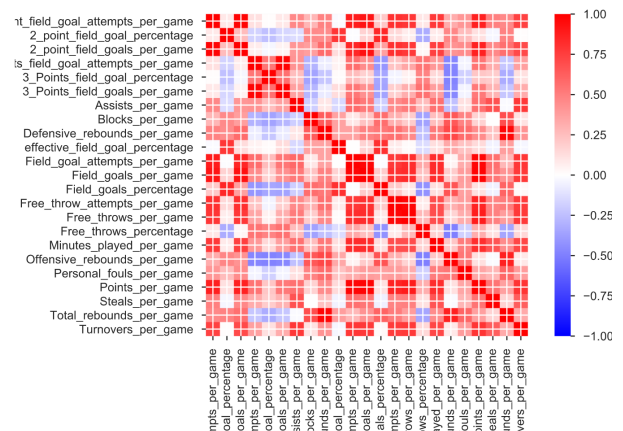


Figure 5: correlation matrix for all performance features

From the correlation matrix, we see that most performance statistics are highly correlated. Except for “percentage” statistics, as they are negatively correlated with “attempts”. Thus in the “slippery index” construction, we will exclude all scoring attempts features as their information has been embedded in the percentage and scores.

3. Non-/Contract Year Performance

The first question we researched was whether there existed difference between players’ contract year performance and non-contract year performance. Contract year season was generally defined as the seasons prior to players’ free agency eligibility. We firstly want to see when contract year will usually occur in the players career. Thus we plot a histogram showing the happening of contract year seasons relative to entire career length (0 means the contract year happened at the player’s first season; 1 means the contract year happened at the player’s last season) (Figure 6)

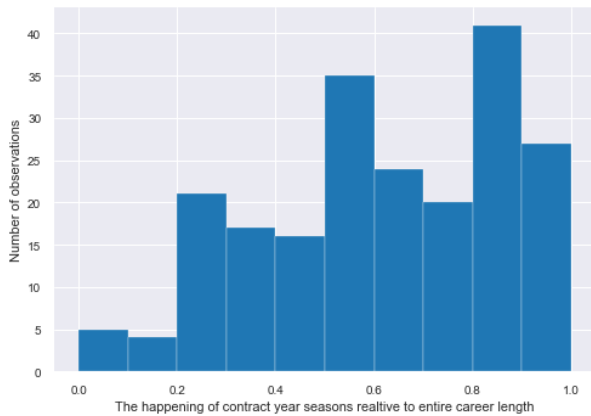


Figure 6: Contract year happening time relative to career length

From the histogram we can see that more contract years happened in the twilight of the players' career. Thus, a simple comparison between contract year and non-contract year will introduce age bias. Also, at the end of a player's career, it could be often that the contract was extended on yearly based. To overcome these challenges, we focused on consecutive contract year and post-contract year comparison, as it mitigated the age bias. Also whether it presents performance reversion after the contract year should be the key to justify the existence of "contract year phenomenon".

We established our formal definition of contract/post-contract consecutive years model:

(1) contract year samples composed all contract year season which the next season was not a contract year;

(2) post-contract year samples composed all non-contract year seasons which the previous season was in contract year.

Totally we had 157 pairs of such contract/post-contract year seasons. Then we took the average of them (Table 2):

Average performance	Contract	Post-
---------------------	----------	-------

statistics		contract
(1) Minutes played per game	28.2879	27.8
(2) Field goals per game	5.189809	5.13121
(3) Field goal attempts per game	11.13057	11.04459
(4) Field goals percentage	0.468529	0.464529
(5) 3_Points field goals per game	1.251592	1.271338
(6) 3_Points field goal attempts per game	3.417834	3.53121
(7) 3_Points field goal percentage	0.320102	0.31779
(8) 2_point field goals per game	3.94586	3.861146
(9) 2_point field goal attempts per game	7.710191	7.510191
(10) 2_point field goal percentage	0.509764	0.510669
(11) Effective field goal percentage	0.524096	0.522134
(12) Free throws per game	2.592357	2.57707
(13) Free throws attempts per game	3.320382	3.282166
(14) Free throws percentage	0.774962	0.768796
(15) Offensive rebounds per game	1.235032	1.13758
(16) Defensive rebounds per game	4.089809	4.068153
(17) Total rebounds per game	5.313376	5.205096
(18) Assists per game	3.216561	3.296178
(19) Steals per game	0.954777	0.929299
(20) Blocks per game	0.575159	0.569427
(21) Turnovers per game	1.740127	1.761783
(22) Personal fouls per game	2.191083	2.1
(23) Points per game	14.22293	14.09045

Table 2: Average post-/contract year performance

From the table, we could see that 18 over 23 performance statistics, contract year season took a lead over post-contract year season.

Take the “points per game” these critical statistics as an example. We examined the distribution difference between contract year and post-contract year (Figure 7). And we could observe a clear shift to lower points per game after the contract year (blue line was the contract year; orange line was the post-contract year).

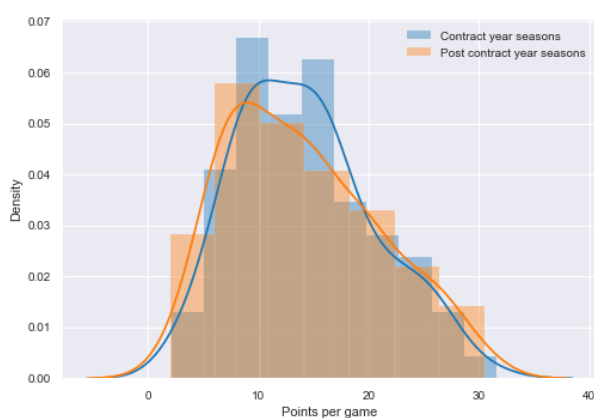


Figure 7: points per game distribution for two groups

From the above analysis, we made a preliminary conclusion that contract year phenomenon exists.

4. Future Work

In our future work, there are two primary focuses:

(1) The above contract year/post contract year comparison was not very statistically robust. We will do further statistical test to conclude or reject the existence of statistical important difference.

(2) As we introduced at the beginning, we are going to break down the divergence between

contract year and post-contract year to each individual player, and attempt to construct a “slippery index” to quantify the contract year behavior for each player, which would be used to examine the relationship between players’ salary and contract year behaviors.