International Journal of Physical Education, Sports and Health 2019; 6(6): 101-105

P-ISSN: 2394-1685
E-ISSN: 2394-1693
Impact Factor (ISRA): 5.38
IJPESH 2019; 6(6): 101-105

www.kheljournal.com
Received: 28-09-2019
Accepted: 30-10-2019

Jose A Martinez
Business Management Department, Technical University of Cartagena (Spain). Cartagena, Spain

Corresponding Author:
Jose A Martinez
Business Management Department, Technical University of Cartagena (Spain). Cartagena, Spain

A method to evaluate the most improved player in basketball based on a non-linear difficulty curve

Jose A Martinez

Abstract
This paper has proposed a method to evaluate the improvement of basketball players from one season to following, considering the difficulty of the achieved performance. After adjusting a non-linear function and employing the Lagrange interpolation, a difficulty curve is obtained. By integrating between the performance of the previous season and the performance of the current season, a final coefficient (difficulty area) is provided. This coefficient is empirically based on the probabilities of performance of players, and provides an alternative criterion to decide which players has been the most improved.

Keywords: Basketball, NBA, most improved paper, statistics, box-score, player productivity

Introduction
One of the awards that the NBA grants every season is the most improved player (MIP) prize. This prize is awarded to the player who has grown more in performance from one year to other. A panel of sportswriters give their votes; the election, consequently, is subjective. However, sportswriters use to ground the decision on some key statistics about the production of the player and the evolution from one year to another. The comparison of performance between two years use to be done through the raw comparison and the percentage comparison. The following example illustrates both situations. Consider Player 1, who has scored 6 points per game in the previous season and 16 points per game in the current one. Consider also Player 2 who has scored 10 points per game in the previous season and 20 points per game in the current one. And, finally, consider Player 3, who has scored 20 points in the previous season and 30 points in the current one. Comparison method computes the difference in raw data, while percentage method computes the percentage of growth, as Table 1 shows.

Table 1: Illustration of methods to evaluate the most improved player (performance is measured in points per game)

	Previous season	Current season	Raw difference	Growth percentage
Player 1	6	16	10	167%
Player 2	10	20	10	100%
Player 3	20	30	10	50%
Player 4	30	35	5	17%

The 3 first players have the same raw difference (10 points), but the growth percentage is different. In this scenario, considering growth percentage as a criterion to decide between them can be misleading. It is true that Player 1 grows 167% but the difficulty of growing in this range of points per game is lower than the difficulty of growing to the extent that the range of points increases. Therefore, it has more merit to grow 10 raw points from 10 to 20, and obviously it has even more merit to grow 10 raw points from 20 to 30.But, how to evaluate this “difficulty”? This point is particularly relevant if we consider Player 4, who has only growth by 5 raw points and a growth percentage of 17%. Although it seems that his evolution is not as good as Player 3 does, the merit could be higher because the extreme difficulty of jumping from 30 to 35 points per game. The aim of this research is to propose a method to handle these problems based on a difficulty curve.

This curve is empirically based on the probabilities of performance of players, and provides an alternative criterion to decide which player has been the most improved.

Method

Non-linear decay

We will employ the index Player Total Contribution (PTC), which has been presented in Martinez (2019a; 2019b; 2019c). PTC is an index to evaluate the performance (production) of basketball players, which is based only on box-score data, and it has been validated using several procedures (Martinez, 2019a).PTC = 1 PTS + 0.91 BLK + 0.58 DRB + 0.92 ORB + 0.86 STL + 0.48 AST + 0.23 FD - 0.91 MFG – 0.57 MFT – 0.86 TOV – 0.23 PFW here: PTS: points made; BLK: blocks made; DRB: defensive rebounds; ORB: offensive rebounds; STL: steals; AST: assists; FD: fouls drawn. MFG: missed field goals; MFT: missed free throws; TOV: turnovers; PF: personal fouls made. PTC can be easily computed by game (PTC/G) or by minutes played (PTC/MP), just dividing PTC by games or minutes, respectively. In this case, the fairest way to compare players with disparate minutes played is to choose PTC/MP.

Figure 1 shows the empirical distribution of PTC/MP for the 2018/19 NBA regular season.

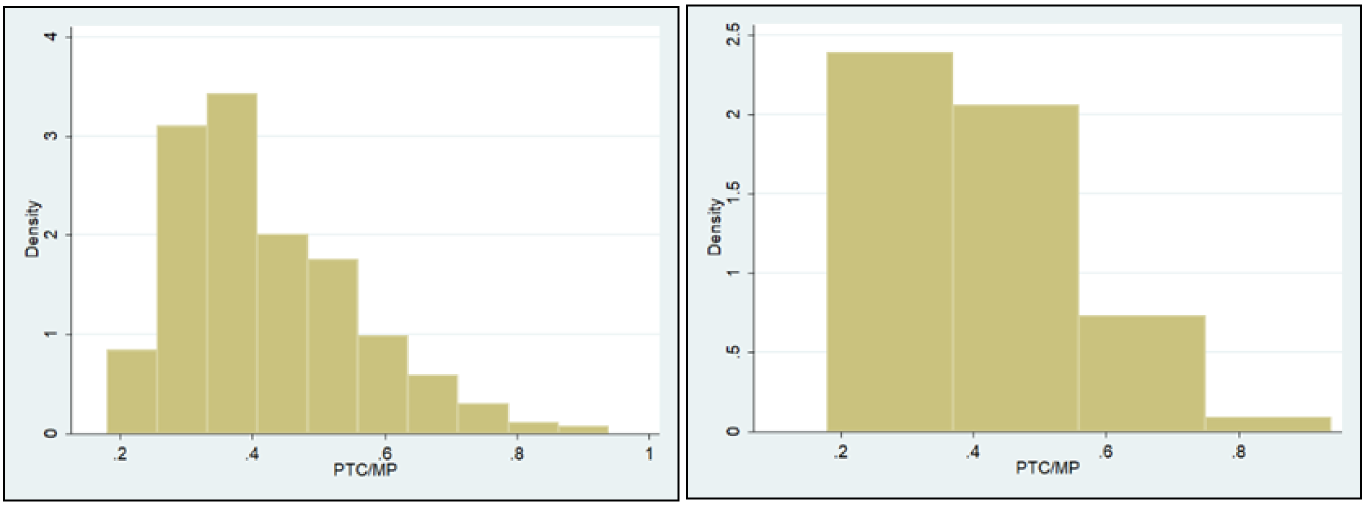

Fig 1: Histogram of the distribution of PTC/MP for the 2018/19 NBA regular season: (a) using 10 bins; (b) using 4 bins

As Figure 1a shows, there is a point (0.4) where the frequency of PTC/MP starts to decline in a non-linear way. This represents the difficulty of increasing performance, because the probability of achieving a high performance has an exponential decay. However, for the lowest PTC/MP data, the difficulty is very low, although the first two bins of the histogram show lower height than the third. It is easy to understand that the reason is not an increase in difficulty, but only an effect of approaching to the median performance. Therefore, as a Figure 1b shows, we may consider a non-linear decay which represents a difficulty coefficient.

Partition of two functions

We propose to aggregate data in the following way (Table 2).

Table 2: Distribution of PTC/MP for the 2018/19 NBA regular season and probabilities

Partitions of PTC/MP	Percentage of data in each partition	Cumulative probability
0.4	0.55956	0.55956
0.5	0.19668	0.75623
0.6	0.13296	0.88920
0.7	0.07479	0.96399
0.8	0.02493	0.98892
0.9	0.00831	0.99723
1	0.00277	1.00000

The aim is to fit a function $f_2(x)$ representing data. As Figure 1 has shown the distribution is similar to an exponential decay, so we would seek models with the following specification:

$$f_2(x) = y_i = e^{\beta_0 - \beta_1 x_i + u_i} \qquad (1)$$

Where y_i is the percentage of data in each i partition, β_1 is the weight of the x_i PTC/MP partition, and u_i is a random error. To estimate equation (1) via OLS method, we can re-specify the model as follows (2)

$$\ln(y_i) = \beta_0 - \beta_1 x_i + u_i \qquad (2)$$

The difficulty curve

Once obtained the two function, the next step is to compute the difficulty coefficient (DC) as simply one minus the value of the fitted curves. Therefore, the less probable performances (high PTC/MP) would have a higher DC, always in a (0, 1) range. Therefore, DC, for each i player, would be computed as follows (3):

$$\widehat{DC_i} = 1 - \hat{y}_i \qquad (3)$$

The “hat” remembers us that is an estimation.

Integrating the DC curve to obtain the difficulty area (DA)

The final step is to compute the integral between the two required values of the player, to obtain the difficulty area (DA). The first one is the performance for the previous season, i.e.PTC/MP_{t-1}, and the second one is the performance of the current season, i.e. PTC/MP_t

Therefore, if $PTC/MP_{t-1} < 0.4$ and $PTC/MP_t < 0.4$, the integral would be (4):

$$DA = \int_{PTC/MP_{t-1}}^{PTC/MP_t} 1 - f_1(x)\ dx \quad (4)$$

If $PTC/MP_{t-1} < 0.4$ and$PTC/MP_t \geq 0.4$, the integral would be (5):

$$DA = \int_{PTC/MP_{t-1}}^{0.4} 1 - f_1(x)\, dx + \int_{0.4}^{PTC/MP_t} 1 - f_2(x)\, dx \quad (5)$$

Finally, if $PTC/MP_{t-1} \geq 0.4$ and$PTC/MP_t \geq 0.4$, the integral would be (6):

$$DA = \int_{PTC/MP_{t-1}}^{PTC/MP_t} 1 - f_2(x)\ dx \quad (6)$$

The area under their respective curves would be the final DC value that we need to compute, i.e. the new raw difference in performance now adjusted by the coefficient of difficulty.

Results

Results of the OLS estimation is showed in Table 3.

Table 3: Results of the OLS estimation

	Coefficient
PTC/MP	-8.55**
Constant	2.95**
R^2	.97**

**p<0.05

Therefore, the function for the range [0.4, 1] is (7):

$$f_2(x) = y = e^{2.95-8.55x} \quad (7)$$

Figure 2 shows the raw data and the prediction after fitting the function.

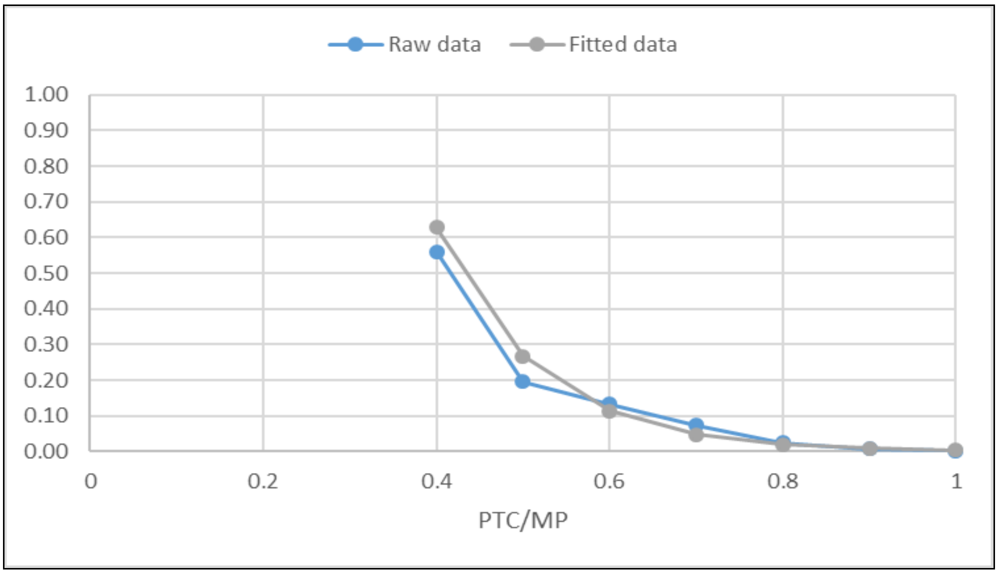

Fig 2: Raw and fitted data for the range [0.4,1]

After applying the Lagrange interpolation, we obtain the function for the range [0, 0.4):

$$f_1(x) = y = 1 - 0.9325x$$

Now, we may plot the two functions (Figure 3)

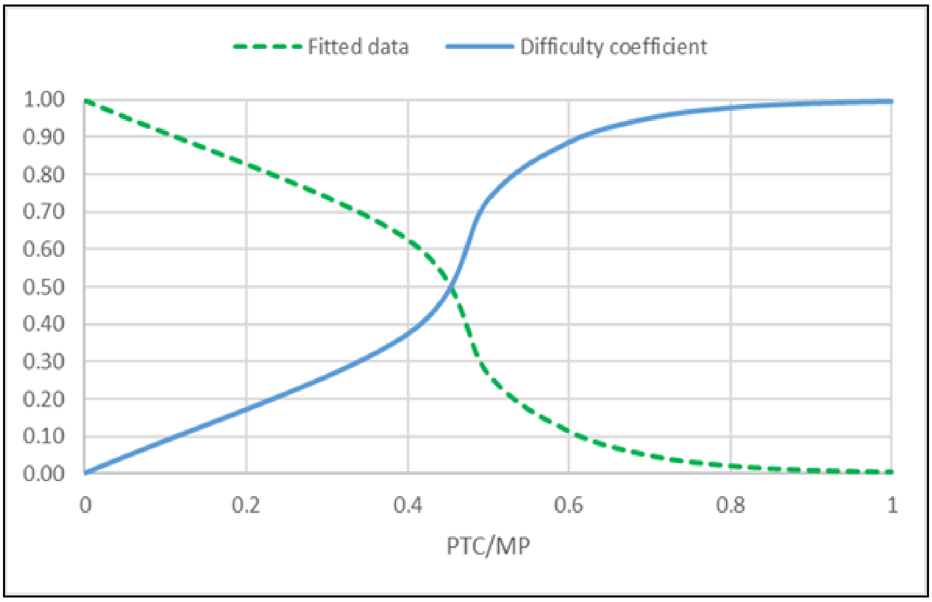

Fig 3: Fitted data and Difficulty Curve (DC)

Example of application

We could apply our method to a similar situation depicted in Table 1, but this time using PTC/MP (Table 4). A first sight could lead some experts to choose Player 1, because it has a raw difference equal to Player 2 and Player 3 and higher than Player 4, and it has the highest growth percentage.

Table 4: Illustration of de DC method to evaluate the most improved player (performance is measured PTC/MP)

	Previous season	Current season	Raw difference	Growth percentage	DA
Player 1	0.2	0.6	0.4	200%	0.196
Player 2	0.4	0.8	0.4	100%	0.329
Player 3	0.5	0.9	0.4	80%	0.370
Player 4	0.7	0.95	0.25	36%	0.245

However, as we have explained, we should take into account the difficulty coefficient to obtain the difficulty area (DA) for Each player. The last column of Table 4 shows the final values, once computed the integrals depicted in the methods section. Player 3 would be the most improved player, although he has a lower growth percentage that Players 1 and 2. In addition, it is clearly showed that the improvement of Player 4 is better than the improvement of Player 1; although Player 1 has higher raw difference and much higher raw percentage, once considering the difficulty of improving the performance in the right tail of the distribution of data, we may say that Player 4 has more merit than Player 1.Graphically we could see the usefulness of this proposed method computing two equivalents DA for two disparate players with apparently different performance (Table 5 and Figure 4).

Table 5: Two equivalent improvements (performance is measured PTC/MP)

	Previous season	Current season	Raw difference	Growth percentage	DA
Player 4	0.7	0.95	0.25	36%	0.245
Player 5	0.4	0.719	0.32	80%	0.245

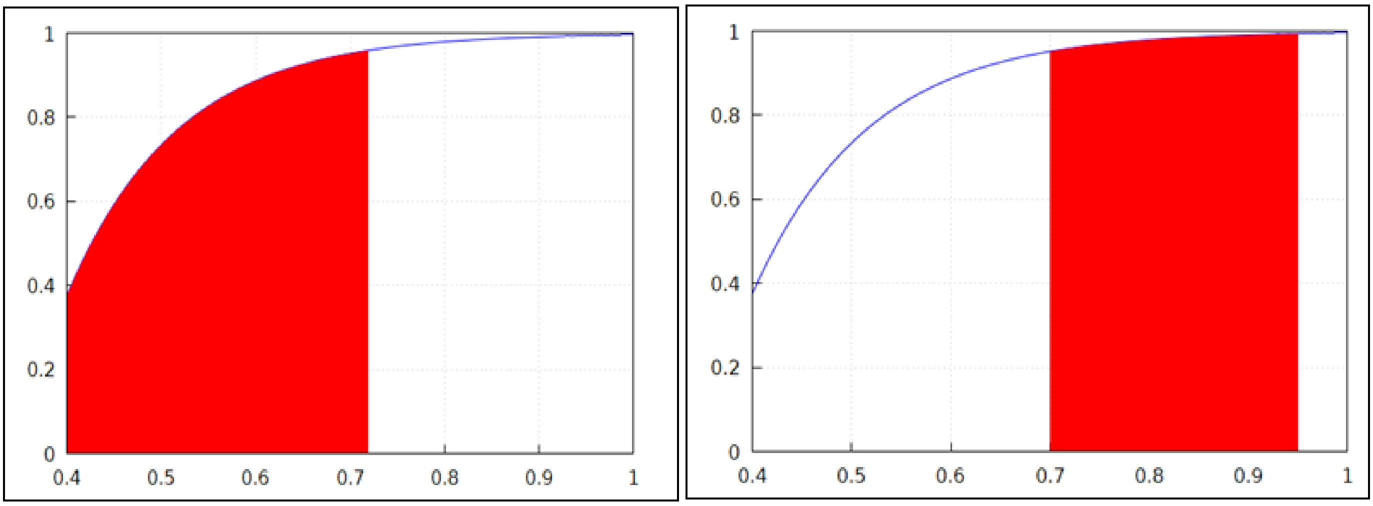

Fig 4: Two equivalents areas under the difficulty curve; (a) Player 4; (b) Player 5

Although Player 5 is better than Player 4 in raw difference and in growth percentage, when considering the difficulty curve, we do notice that both improvements are equal.

Concluding remarks

This paper has introduced a way to evaluate the improvement of basketball players from one season to the next one, considering the difficulty of performance. Therefore, this study contributes to the basketball analytics field by advancing in the knowledge and treatment of data. Basketball analytics is a "hot" research stream that is continuously changing the way this sport is evaluated (e.g. Beouy, 2013; Berri & Bradbury, 2010; Deshpande & Jensen, 2016; Winston, 2009) [1, 2, 3, 7]. Further research could improve the proposed method handling more data, in order to obtain a more robust histogram, which is the initial step of the method. Once of the advantage of our proposal is that the obtained functions can be re-calculated year by year, once considering the PTC/MP of each season. Changes of the functions are not expected to be very significant, but probably it will slightly improve the final estimates.

Funding: This study is the result of the activity carried out under the program Groups of Excellence of the region of Murcia, the Fundación Séneca, Science and Technology Agency of the region of Murcia project 19884/GERM/15.

Conflicts of Interest: The author have no conflict of interest to declare.

References

1. Beouy M. (Internet). NBA win probability calculator; Available from:https://www.inpredictable.com/2013/11/nba-win-probability-calculator.html
2. Berri DJ, Bradbury JC. Working in the land of morticians. Journal of Sports Economics, 2010; 11(1):29-47.
3. Deshpande SK, Jensen S. Estimating an NBA player's impact on his team's chances of winning. Journal of Quantitative Analysis in Sports, 2016; 12(2):51-72.
4. Martinez JA. A method to compute standard errors in per-minute performance metrics in basketball. International Journal of Physical Education, Sports and Health. 2019; 6(4):45-51.
5. Martinez JA. A more robust estimation and extension of factors determining production (FDP) of basketball players. International Journal of Physical Education, Sports and Health, 2019; 6(3):81-85.
6. Martinez JA. An alternative method to compute win probabilities and to measure player productivity in basketball. International Journal of Physical Education, Sports and Health. 2019; 6(4):165-173.
7. Winston WL. Mathletics. New Yersey: Princeton University Press, 2009.