International Journal of Physical Education, Sports and Health 2019; 6(4): 165-173

P-ISSN: 2394-1685
E-ISSN: 2394-1693
Impact Factor (ISRA): 5.38
IJPESH 2019; 6(4): 165-173

www.kheljournal.com
Received: 14-05-2019
Accepted: 18-06-2019

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

Correspondence
Jose A Martínez
Business Management Department, Technical University of Cartagena (Spain), Cartagena, Spain

An alternative method to compute win probabilities and to measure player productivity in basketball

Jose A Martínez

Abstract
This paper has proposed an alternative method to compute win probabilities which has been theoretically based, and that has been built upon the concept of estimated possessions. After taking into account the moment of time of each game action and the scoreboard differential, estimated possessions has been computed using a truncated Poisson regression model on a sample of 5,622 play-by-play observations. Once obtained the estimated possessions, the value of each action has been derived from the difference in theoretical probabilities of the potential value of each action reflected in a change in the score differential. Therefore, box-score statistics can be weighted using a context-dependent system of evaluation, and then computing a global index of productivity. As an empirical example, Player Total Contribution (PTC) was taken as an index summarizing the main box-score variables, and it has been showed how this index can change depending on the variations in the time and scoreboard for every play of the game. Consequently, two players with the same box-score performance could have really contributed very different to the winning probability of a team. Future research is needed to make this procedure more easily implemented.

Keywords: win probability, statistics, box-score, player productivity

1. Introduction

How to assess the value of player performance in basketball is one of the major objectives of basketball analytics, where a plethora of indexes have been proposed (see for example, www.basketball-reference.com). The vast majority of those metrics are static metrics in the sense that they do not take into account the moment of the game when the action was made.

This fact is relevant for basketball, because the sport psychology literature recognizes the existence of clutch behavior among elite athletes (Solomonov, Avugos & Bar-Eli, 2015; Swann *et al*., 2017) [19-20]. Those “clutch” players are considered “special” because they perform better than the average in the moments of the game were the shots have more value or difficulty. As Zuccolotto, Marica & Sandri (2018) [21] pointed out, those players have a “clutch skill”, i.e. a player’s ability to perform better in ‘‘clutch’’ or high-pressure situations The NBA site (www.stats.nba.com) and also several specialized websites (e.g. www.nbaminer.com, www.82games.com) provide clutch statistics defined as any shot taken in the last five minutes of the game in which the score differential is five points or less.

However, as Beuoy (2016) [5] indicated, a shot taken in the last minute when your team is up by five doesn't feel quite as "clutch" as a three pointer in the final 10 seconds with your team trailing by two. In addition, it seems also obvious that the psychological state of a player and the pressure of to make a shot in a tied game should be practically identical when time remaining is 5:01 that when is 4:59. Nevertheless, only in the second case the situation would be considered “clutch”.

This led Beouy (2013, 2015) [2-4] to develop a system of evaluating the value of each shot using win probabilities. Win probabilities is an empirical tool to evaluate the value of a shot based on the analysis of historical play-by-play data, where the probability of winning a game is predicted in function to the time remaining and the score difference between teams. In addition, Beouy (2015) [4] considered an estimate of the difference of team strength to provide a two version tool for knowing the value of each game action: (1) assuming teams are evenly matched; (2) controlling for difference in team strength. The value each action in the game is

quantified as the marginal gain or lose in win probability.
The use of win probabilities is not a novel concept in sports analytics (See Lock, 2016, for a brief review) [14], and also other researchers have proposed different models to obtain the impact of basketball players have in their teams' chances of winning. For example, Desphande & Jensen (2016) [9] implemented a bayesian regression model to estimate and individual player's impact, once controlling for the other players on the court. Their method had some parallelism with the Adjusted Plus-Minus (Rosembaum, 2004) [18], but instead of finding a measure of how effective a player is with regard to his contribution to the margin (home team points per possession minus away team points per possession), it obtains an estimation of each player's partial effect on his team chances of winning. A similar approach was also introduced by Lock & Nettleton (2014) [13]. Moreover, McFarlane (2019) [17], developed a probabilistic method to assess the tactical decisions coaches and players make at the end of NBA games. This method was based on empirical win probabilities computed for the last three minutes of the game.
Therefore, we should distinguish between the value of a shot made (and shot missed) and when a shot has to be considered clutch. In the first case, each player shot adds positive or negative value to the probability of winning a game, increasing its value to the extent that score difference is lowering and time remaining is reducing. In the second case, several high-pressure situations could be identified, not necessarily placed at the end of a game of a even match. Zuccolotto, Marica & Sandri (2018) [21] proposed four situations: (a) when the shot clock is going to expire; (b) when the score difference with respect to the opponent is small; (c) when the team as a whole has performed poorly during the match up to that particular moment in the game; (d) when the player has missed his previous shot. Consequently, some high-pressure situations do not have to necessarily match with high valuable shots (high wins added value).
The aim of this research is to propose a new method to compute the value of made and missed shots, and all the variables appearing in the box-score, which is not based on the direct calculation of empirical winning probabilities, but indirectly through estimated remaining possessions. The main advantage of our method is dealing with theoretical (not empirical) winning probabilities, so that the value of every shot is not directly conditioned by historical data. In addition, it allows an extension to quantify the value of each box-score statistic in function of its relationship with the value of a possession.
The remainder of this paper is organized as follows: (1) we explain step by step the rationale of our proposal, deriving theoretical probabilities from the time remaining to the end of a game, using a one possession-one second equivalence; (2) then we implement a Poisson regression model to estimate the remaining possessions for every second of a game; (3) after that, we match the estimated possessions with the remaining seconds, modifying the prior equivalence (one possession-one second) to the adjusted data; (4) we compute the win probabilities for this new equivalence; and (5) we expand our procedure to provide a full index of player productivity based of the importance of each action.

2. Method
2.1 The possessions matrix
As a first approach, we can divide a 48 minutes' game into 2880 possessions (One for each second), obtaining a matrix of 2881x2881 possible options. Theoretically is possible to make a shot in one second (As we see sometimes in the end of matched games), so we will follow this unrestricted assumption to introduce our proposal.
Let's illustrate this with an example: Let's suppose that there are 16 seconds left until the match ends. Theoretically, it is possible that there are 8 possessions for each team (of 1 second each one), so we can build a matrix P_{ij} in the following way:

$$P_{ij} = \begin{pmatrix} P_{00} & P_{01} & P_{02} & P_{03} & P_{04} & P_{05} & P_{06} & P_{07} & P_{08} \\ P_{10} & P_{11} & P_{12} & P_{13} & P_{14} & P_{15} & P_{16} & P_{17} & P_{18} \\ P_{20} & P_{21} & P_{22} & P_{23} & P_{24} & P_{25} & P_{26} & P_{27} & P_{28} \\ P_{30} & P_{31} & P_{32} & P_{33} & P_{34} & P_{35} & P_{36} & P_{37} & P_{38} \\ P_{40} & P_{41} & P_{42} & P_{43} & P_{44} & P_{45} & P_{46} & P_{47} & P_{48} \\ P_{50} & P_{51} & P_{52} & P_{53} & P_{54} & P_{55} & P_{56} & P_{57} & P_{58} \\ P_{60} & P_{61} & P_{62} & P_{63} & P_{64} & P_{65} & P_{66} & P_{67} & P_{68} \\ P_{70} & P_{71} & P_{72} & P_{73} & P_{74} & P_{75} & P_{76} & P_{77} & P_{78} \\ P_{80} & P_{81} & P_{82} & P_{83} & P_{84} & P_{85} & P_{86} & P_{87} & P_{88} \end{pmatrix}$$

The first number i of each pair of numbers represents that the reference team takes advantage of a possession or not, and the second number j indicates that the opposite does it or not. Thus, for example, P_{00} means that no team has taken advantage of any of the 8 possessions they had, P_{52} means that the reference team has taken advantage of 5 of them and the opposite only of 2; P_{12} means that the reference team has taken advantage of 1 possession while the opponent has taken advantage of 2. Each possession has a maximum value of 3 points [1].
Therefore, we have a $9x9$ matrix, so we know that the number of k elements of the square matrix to be built would be: $k = \left(\frac{n}{2}+1\right)^2$, where n is the number of total possessions left.
In the P_{ij} matrix all possible options are represented. Therefore $i = 0,1,2 \ldots n$, and $j = 0,1,2 \ldots n$
If the scoreboard is in a tie, the reference team has the following probability $p_0^{(n)}$ of winning:

$$p_0^{(n)} = \frac{\frac{1}{2}\left[\left(\frac{n}{2}+1\right)^2 - \left(\frac{n}{2}+1\right)\right]}{\left(\frac{n}{2}+1\right)^2}$$

So, if there are 16 positions left, the probability of winning is:

$$p_0^{(16)} = \frac{\frac{1}{2}\left[\left(\frac{16}{2}+1\right)^2 - \left(\frac{16}{2}+1\right)\right]}{\left(\frac{16}{2}+1\right)^2} = \frac{0.5[81-9]}{81} = \frac{36}{81}$$

That corresponds to the bolded elements of the P_{ij} matrix:

---
[1] There are four points plays, when a 3-point is made and a foul is drawn (and the subsequent free-throw is made). In the 2017/18 NBA regular season there were 175 four points plays (www.nbaminer.com) over a total of 71,340 3-points attempted (0.25%) (www.basketball-reference.com), so the percentage is really small. In addition, four points plays are not an option per se for teams, because it depends of the will of the other team to make a foul, so is not directly under control for teams, and therefore we opted for not considering it as an option for quantifying the maximum possible score of a regular possession.

$$P_{ij} = \begin{pmatrix} P_{00} & P_{01} & P_{02} & P_{03} & P_{04} & P_{05} & P_{06} & P_{07} & P_{08} \\ \mathbf{P_{10}} & P_{11} & P_{12} & P_{13} & P_{14} & P_{15} & P_{16} & P_{17} & P_{18} \\ \mathbf{P_{20}} & \mathbf{P_{21}} & P_{22} & P_{23} & P_{24} & P_{25} & P_{26} & P_{27} & P_{28} \\ \mathbf{P_{30}} & \mathbf{P_{31}} & \mathbf{P_{32}} & P_{33} & P_{34} & P_{35} & P_{36} & P_{37} & P_{38} \\ \mathbf{P_{40}} & \mathbf{P_{41}} & \mathbf{P_{42}} & \mathbf{P_{43}} & P_{44} & P_{45} & P_{46} & P_{47} & P_{48} \\ \mathbf{P_{50}} & \mathbf{P_{51}} & \mathbf{P_{52}} & \mathbf{P_{53}} & \mathbf{P_{54}} & P_{55} & P_{56} & P_{57} & P_{58} \\ \mathbf{P_{60}} & \mathbf{P_{61}} & \mathbf{P_{62}} & \mathbf{P_{63}} & \mathbf{P_{64}} & \mathbf{P_{65}} & P_{66} & P_{67} & P_{68} \\ \mathbf{P_{70}} & \mathbf{P_{71}} & \mathbf{P_{72}} & \mathbf{P_{73}} & \mathbf{P_{74}} & \mathbf{P_{75}} & \mathbf{P_{76}} & P_{77} & P_{78} \\ \mathbf{P_{80}} & \mathbf{P_{81}} & \mathbf{P_{82}} & \mathbf{P_{83}} & \mathbf{P_{84}} & \mathbf{P_{85}} & \mathbf{P_{86}} & \mathbf{P_{87}} & P_{88} \end{pmatrix}$$

If the reference team wins by a possession, then it has the following probability $p_1^{(n)}$ of winning:

$$p_1^{(n)} = \frac{\left[\frac{1}{2}\left[\left(\frac{n}{2}+1\right)^2 - \left(\frac{n}{2}+1\right)\right] + \left(\frac{n}{2}+1\right)\right]}{\left(\frac{n}{2}+1\right)^2}$$

$$p_1^{(n)} = \frac{\frac{1}{2}\left[\left(\frac{n}{2}+1\right)^2 - \left(\frac{n}{2}+1\right)\right]}{\left(\frac{n}{2}+1\right)^2} + \frac{\left(\frac{n}{2}+1\right)}{\left(\frac{n}{2}+1\right)^2} = p_0^{(n)} + \frac{\left(\frac{n}{2}+1\right)}{\left(\frac{n}{2}+1\right)^2}$$

Therefore, if there are 16 positions left, the probability of winning is:

$$p_1^{(16)} = \frac{36}{81} + \frac{9}{81} = \frac{45}{81}$$

That corresponds to the bolded elements of the P_{ij} matrix:

$$P_{ij} = \begin{pmatrix} \mathbf{P_{00}} & P_{01} & P_{02} & P_{03} & P_{04} & P_{05} & P_{06} & P_{07} & P_{08} \\ \mathbf{P_{10}} & \mathbf{P_{11}} & P_{12} & P_{13} & P_{14} & P_{15} & P_{16} & P_{17} & P_{18} \\ \mathbf{P_{20}} & \mathbf{P_{21}} & \mathbf{P_{22}} & P_{23} & P_{24} & P_{25} & P_{26} & P_{27} & P_{28} \\ \mathbf{P_{30}} & \mathbf{P_{31}} & \mathbf{P_{32}} & \mathbf{P_{33}} & P_{34} & P_{35} & P_{36} & P_{37} & P_{38} \\ \mathbf{P_{40}} & \mathbf{P_{41}} & \mathbf{P_{42}} & \mathbf{P_{43}} & \mathbf{P_{44}} & P_{45} & P_{46} & P_{47} & P_{48} \\ \mathbf{P_{50}} & \mathbf{P_{51}} & \mathbf{P_{52}} & \mathbf{P_{53}} & \mathbf{P_{54}} & \mathbf{P_{55}} & P_{56} & P_{57} & P_{58} \\ \mathbf{P_{60}} & \mathbf{P_{61}} & \mathbf{P_{62}} & \mathbf{P_{63}} & \mathbf{P_{64}} & \mathbf{P_{65}} & \mathbf{P_{66}} & P_{67} & P_{68} \\ \mathbf{P_{70}} & \mathbf{P_{71}} & \mathbf{P_{72}} & \mathbf{P_{73}} & \mathbf{P_{74}} & \mathbf{P_{75}} & \mathbf{P_{76}} & \mathbf{P_{77}} & P_{78} \\ \mathbf{P_{80}} & \mathbf{P_{81}} & \mathbf{P_{82}} & \mathbf{P_{83}} & \mathbf{P_{84}} & \mathbf{P_{85}} & \mathbf{P_{86}} & \mathbf{P_{87}} & \mathbf{P_{88}} \end{pmatrix}$$

If the reference team wins by more than one possession, then it has the following probability p_m of winning:

$$p_m^{(n)} = p_{m-1}^{(n)} + \frac{\left(\frac{n}{2}+1\right) - (m-1)}{\left(\frac{n}{2}+1\right)^2}$$

For example, if the team is winning by two possessions:

$$p_2^{(16)} = p_1^{(16)} + \frac{9-1}{81} = \frac{45}{81} + \frac{8}{81} = \frac{53}{81}$$

It quickly becomes apparent that when $m = 9$, then $p_m = 1$, i.e. when $m = (n+1)/2$, i.e. the other team does not have enough possessions to win the match, even if it scores in all of them, and the team that is winning does not score in any of them.

In this way, the formula for the probability of winning can be generalized as follows (1):

$$p_m^{(n)} = \frac{\frac{1}{2}\left[\left(\frac{n}{2}+1\right)^2 - \left(\frac{n}{2}+1\right)\right]}{\left(\frac{n}{2}+1\right)^2} \qquad if\ m = 0 \tag{1}$$

$$p_m^{(n)} = p_{m-1}^{(n)} + \frac{\left(\frac{n}{2}+1\right) - (m-1)}{\left(\frac{n}{2}+1\right)^2} \qquad if\ m = 1,2 \ldots n/2$$

$$p_m^{(n)} = 1 \qquad if\ m \geq (n+1)/2$$

To solve the inconvenience of odd possessions, the same procedure is followed, this time as an approximation. For example, if for 16 possessions $p_0^{(16)} = \frac{36}{81}$ and for 18 possessions $p_0^{(18)} = \frac{45}{100}$ then for 17 possessions:

$$p_0^{(17)} = \frac{1}{2}\left(\frac{36}{81} + \frac{45}{100}\right) = .447$$

2.2. Points differential in each possession

When a team is winning by one or more possessions of advantage, it can do so by 1, 2 or 3 points within each possession. Thus, if a team is winning by two advantage possessions then it may be winning by 4, 5 or 6 points. In this way, the opposing team needs at least two possessions to match or, in some cases, overcome the reference team.

The way to consider the difference of points in each possession is simple. This differential would be determined by the difference in probability between these two situations. Following the example of the 16 possessions, we have the following differentials:

$$d_{1,0}^{(16)} = p_1^{(16)} - p_0^{(16)} = \frac{45}{81} - \frac{36}{81} = \frac{9}{81} = .1111$$

$$d_{2,1}^{(16)} = p_2^{(16)} - p_1^{(16)} = \frac{53}{81} - \frac{45}{81} = \frac{8}{81} = .0987$$

So the general expression is (2):

$$d_{m,m-1}^{(n)} = p_m^{(n)} - p_{m-1}^{(n)} \tag{2}$$

And now we have to divide that differential into 3 parts according to whether that difference is 1, 2 or 3 points in each possession.

$$d_{1,0\_(1)}^{(16)} = 0.1111\,\frac{1}{3} = .0370$$

$$d_{1,0\_(2)}^{(16)} = 0.1111\,\frac{2}{3} = .0740$$

$$d_{1,0\_(3)}^{(16)} = 0.1111\,\frac{3}{3} = .1111$$

Therefore, when the team wins by 3 points within each possession differential, the maximum value is reached (probability of winning), while if it is winning by 1 or 2 points in that possession the probability of winning decreases.

2.3 The non-symmetry of the probabilities in the case that the reference team is losing

When the reference team is losing we have to consider that there is asymmetry in probabilities.

To exemplify this fact, it is enough to go back to the commented case that there are 16 possessions left for the end, and the reference team is now losing by a possession (instead of winning by a possession).
If the reference team is losing by 1 possession, then it has the following probability $p_{-1}^{(n)}$ of winning:

$$p_{-1}^{(n)} = \frac{\frac{1}{2}\left[\left(\frac{n}{2}+1\right)^2 - \left(\frac{n}{2}+1\right)\right]}{\left(\frac{n}{2}+1\right)^2} - \frac{\left(\frac{n}{2}+1\right)-1}{\left(\frac{n}{2}+1\right)^2} = p_0 - \frac{\left(\frac{n}{2}+1\right)-1}{\left(\frac{n}{2}+1\right)^2}$$

So, if there are 16 positions left, we have:

$$p_{-1}^{(16)} = \frac{36}{81} - \frac{8}{81} = \frac{28}{81}$$

And therefore, the general expression is:

$$p_{-m}^{(n)} = p_{-m+1}^{(n)} - \frac{\left(\frac{n}{2}+1\right)-m}{\left(\frac{n}{2}+1\right)^2}$$

For example, if the reference team is losing by 2 possessions:

$$p_{-2}^{(16)} = p_{-1}^{(16)} - \frac{9-2}{81} = \frac{28}{81} - \frac{7}{81} = \frac{21}{81}$$

In this way, the formula for the probability of winning when the reference team is losing can be generalized as follows (3):

$$p_{-m}^{(n)} = p_{-m+1}^{(n)} - \frac{\left(\frac{n}{2}+1\right)-m}{\left(\frac{n}{2}+1\right)^2} \quad if\ m = 1,2 \ldots n/2 \qquad (3)$$

$$p_{-m}^{(n)} = 0 \qquad si\ m > n/2$$

The way to consider the difference of points in each possession when the reference team is losing is similar than when is winning. This differential would be determined by the difference in probability between these two situations. Following the example of the 16 possessions, we have the following differentials:

$$d_{-1,0}^{(16)} = p_0^{(16)} - p_{-1}^{(16)} = \frac{36}{81} - \frac{28}{81} = \frac{8}{81} = .0987$$

$$d_{-2,-1}^{(16)} = p_{-1}^{(16)} - p_{-2}^{(16)} = \frac{28}{81} - \frac{21}{81} = \frac{7}{81} = .0864$$

So the general expression is (4):

$$d_{-m,-m-1}^{(n)} = p_{-m+1}^{(n)} - p_{-m}^{(n)} \qquad (4)$$

And now we have to divide that differential into 3 parts according to whether that difference is -1, -2 or -3 points in each possession.

$$d_{-1,0\_(-1)}^{(16)} = 0.0987\,\frac{1}{3} = .0329$$

$$d_{-1,0\_(-2)}^{(16)} = 0.0987\,\frac{2}{3} = .0658$$

$$d_{-1,0\_(-3)}^{(16)} = 00987\,\frac{3}{3} = .0987$$

Clearly, it is appreciated the non-symmetry of differential probabilities. When 16 possessions remain, if the game was tied and the reference team has made 1 point, the value of such point is:

$$d_{1,0\_(1)}^{(16)} = 0.1111\,\frac{1}{3} = .0370$$

But if the reference team was 1 point below the other team and make 1 point to tie the game, when 16 possessions remains the value of such point would be:

$$d_{-1,0\_(-1)}^{(16)} = 0.0987\,\frac{1}{3} = .0329$$

2.4 The value of a missed shot and the value of the remaining box-score statistics

We can consider two different approaches to compute the value of a missed shot. The first one is directly related with the value of making a shot. The average points per possession in the NBA is 1.06, from the 1973-74 to the 2017-18 seasons (www.basketball-reference.com). Therefore, we could make equal every point made with the cost of a non-scoring possession, following a simple equivalence relationship:

$$value = \frac{1.06\ point}{1\ possession} = -\frac{1.06\ miss}{1\ possession}$$

As each value depends of probabilities, there would be no problem to make equal points made and shots missed. Therefore, each field goal missed would have the same value of increasing the probability of winning by increasing one point the score differential, but with the opposite sign. Each missed free throw, however, would count differently, because, in a general way two free throw missed are equivalent to a non-scoring possessions. As the prevalence of a sequence of two free throws are largely higher than one free throw (technical foul or and-1), and three free throws (shooting foul from the 3 point-line), we may consider that each free throw made counts 1 points but each free throw missed counts 0.5 negative points, i.e. the half. This procedure, however, has the shortcoming of not considering, for example, that missing a shot is not so costly if an offensive rebound is grabbed.
The second one is related to computing the cost of missing a shot considering the probability of recovering the possession. A percentage of missed shots is grabbed by the same team (as offensive rebounds), so that true value of missing a shot would be lower than 1. In fact, the negative value of missing a shot, losing the ball or making a foul, and the positive value of making an assist, rebounding, stealing and blocking depends on several factors, i.e. disparate probabilities. For example, as 82games.com (2019) shows, offensive rebounding depends on the shot type (Jumper, close up, dunks and free throws), so the cost of missing a shot would also depend on the probability of grabbing an offensive rebound, which also depends on the shot type. In addition, the probability of making a shot after an offensive rebound depends on the type of play the team makes; a right back up with a follow-up shot, or restart the offense.
Consequently, it would be practically impossible to consider all those factors to compute the positive or negative contribution to the win probability after each action. An

alternative method is simply to rely on the weights of the box-score variables relative to making a point. In fact, if those weights are correctly computed, they have to approximately consider all the complex factors mentioned. For example, it is still prevalent in many professional competitions to consider a weight of 1 for positive actions (rebounds, steals, blocks, assists), and 1 for negative actions (missed field goals, missed free throws, turnovers, fouls made). Obviously, these kind of “efficiency” metrics are theoretically and empirically flawed, and other more sophisticated metrics have been proposed, which compute the weights of the box-score variables more correctly. For example, Win Score (Berri, 1999; 2008; 2012) [6, 7, 8] assign half weights to assists, blocks and fouls with respect to points.

In this research we are going to employ a more recent proposal (Martínez, 2019) [16], Player Total Contribution (PTC), which is also based on box-score data, and it has been validated using several procedures (Martínez, 2012; 2019) [15-16]. Weights of PTC are the following (5) [2]:

$$PTC = 1PTS + .91BLK + .58DRB + .92ORB + .86STL -. \quad (5)$$
$$86TOV + .48AST - .91MFG - .57MFT - .23PF$$

Where: PTS: points made; BLK: blocks made; DRB: defensive rebounds; ORB: offensive rebounds; STL: steals; TOV: turnovers; AST: assists; MFG: missed field goals; MFT: missed free throws; PF: personal fouls made

2.5 Estimating remaining possessions for each second of a game

Despite its potential, the method we are depicting has a main problem related to the value assigned to a shot action (Success or failure) when there is few time left to finish the game. Since the possessions are divided in seconds, the number of possessions remaining until the end of the game is overestimated, which is why wrong or unrealistic values of the odds are given.

For example, when 16 possessions remain, if the reference team gets a triple that leads to a 9 to 12 point difference in the scoreboard, then the value of that shot would be .07407. However, the real value of that action should be lesser, even extremely close to .000, because with 16 seconds of match and 9 points of advantage the match is practically sentenced and there is no material possibility of tie except miracle. Therefore, we must look for a procedure to weight those probabilities based on estimated possessions left, especially when the game is close to the final minutes.

To get this aim, we have employed a truncated Poisson regression approach (Hardin & Hilbe, 2012) [11] to estimate the number of possessions left. The dependent variable was the possessions left after every shot made, shot missed or turnover, over a sample of NBA regular season games. We employed the time when each action was made, the scoreboard difference, the interaction between time and scoreboard difference, and the natural logarithm of time as covariates.

A truncated approach is needed because in the count data values of zero cannot occur. In the case of basketball, zero possessions left can only occur if a team shot in the last moment of the game and the other team has no time to achieve any action. As clock is stopped tenth by tenth in the last seconds of the game, in practice this only happens in “buzzer beater” shots within the last second of games. As our objective was to predict remaining possessions, it has no sense to consider such rare cases of zero possessions left.

2.6 Data

Play-by-play data of the first 36 games of the 2008-09 NBA regular season was obtained from www.82games.com. The set of all game actions was 11,919 lines of code, which were converted into a Microsoft Excel spreadsheet.

All the data base had to be manually revised and all the possessions and score differentials were coded manually. We decided to employ this method because algorithms for detecting possessions and score failed because some lines of code were no correctly ordered. For example, the order of a sequence of two free throws was not always from the first to the second but from the second to the first, so that the score differential could change depending of the chosen criterion for the same number of remaining possessions.

Consequently, the final data set to be analyzed comprised 5,622 actions which constituted a possession (shots made, shots missed – included the final free throw of a sequence of free throws- and turnovers). After each of those actions a new possession was considered. For every action the remaining possessions to the end of the game was computed, together with the absolute value of score differential and the seconds left to the end.

3. Results

3.1 Estimating possessions left

A truncated Poisson regression was implemented using Stata 13.0. Following Hardin & Hilber (2012) [11], in order to analyze the possible presence of overdispersion, a truncated negative binomial regression was also computed. Results are showed in Table 1.

Table 1: Models for estimating possessions

	Truncated Poisson n=5,622	**Truncated negative binomial n=5,622**
Covariates	*Coef.*	*Coef.*
Seconds left	.0001074**	.0001074**
Score differential	.0000151	.0000151
Seconds left * Score differential	2.29e-06 **	2.29e-06 **
Ln (Seconds left)	.8342448**	.8342448**
Constant	-1.582962**	-1.582962**
Alpha		5.61e-13
Model assessment		
Pseudo R^2	.8623**	.4009**
AIC	37636.74	37636.74
BIC	37669.91	37669.91
LR test of alpha=0		.000

**p<.001

No over dispersion was detected (LR test for the alpha parameter=.000, *p*=1.000). Results indicated that score differential had no effect on the remaining possession, once controlled for the seconds left. However, its interaction with seconds was significant, together with the natural logarithm of seconds left. Taking together, results indicated that possessions increases with seconds left (an obvious result), but also with the multiplicative effect of score differential and the additional non-linear term of seconds left, which weights differently the last seconds of the game. Figure 1 shows the good fit of the model to predict possessions.

[2] There is a version of PTC including fouls drawn, with a weight of .23, but fouls drawn are not usually reported in the raw box-score in the NBA.

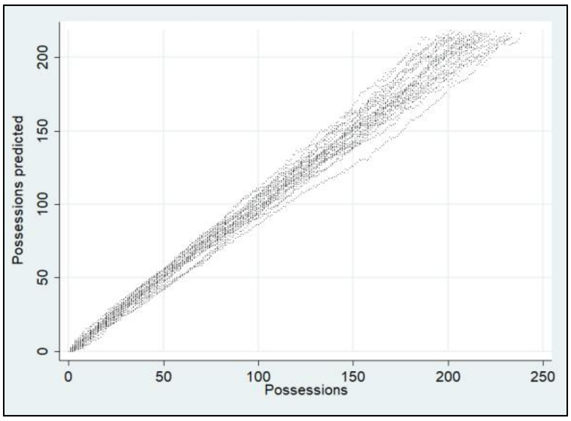

Fig 1: Predicted possessions vs real possessions

3.2 Computing the player total contribution considering win probabilities

We employed the weights of the covariates displayed in Table 1 to predict possessions left. In order to test the value of each box-score action, we worked with the first game of the Dallas Mavericks' player Luka Doncic, in the 2018/19 NBA regular season. Box-score can be obtained from https://www.basketball-reference.com/boxscores/201810170PHO.html, and play-by-play data from https://www.basketball-reference.com/boxscores/pbp/201810170PHO.html. The game was at Phoenix, and the final result was 121-100, the first win for the Suns' team. Table 2 displays the box-score for Doncic.

Table 2: Box-score for Luka Doncic

Player	MP	FG	FGA	FG%	3P	3PA	3P%	FT	FTA	FT%	ORB	DRB	TRB	AST	STL	BLK	TOV	PF	PTS
Luka Doncic	31:47	5	16	.313	0	5	.000	0	2	.000	2	6	8	4	0	0	4	3	10

MP: minutes played: FG; Field goals made; FGA: field goals attempted; FG%: field goals percentage; 3P: Three points made; 3PA: three points attempted; 3P%: three points percentage; FT: free throw made; FTA: free throws attempted; FT%: free throw percentage; ORB: offensive rebounds; DRB: defensive rebounds; TRB: total rebounds; AST: assists; STL: steals; BLK: blocks made; TOV: turnovers; PF: personal fouls made; PTS: points made.

The Player Total Contribution (PTC) was 1.960, and PTC/min was .062. However, PTC, as other box-score metrics, is a static metric which does not reflect the dynamic of a game, and the win probabilities associated with each action.

With the aim of applying our proposed method, we registered every action of Doncic using the play-by-play data of the game, and then we estimated the remaining possessions (Table 3).

Table 3: Play-by-play for Luka Doncic and estimated possessions

Quarter	Minutes	Seconds	Seconds to the end	Action	Reference team scoreboard (Dallas)	Other team scoreboard (Suns)	Estimated possessions
1	11	23	2843	DRB	0	0	212
1	11	16	2836	AST	2	0	214
1	11	4	2824	PF	2	2	210
1	10	39	2799	MFG	2	3	210
1	9	49	2749	2 PTS	4	5	205
1	9	21	2721	MFG	4	9	208
1	8	52	2692	2 PTS	6	11	206
1	7	56	2636	DRB	6	13	203
1	7	50	2630	AST	8	13	200
1	7	6	2586	AST	12	13	192
1	4	48	2448	2 PTS	16	20	184
1	4	11	2411	PF	16	23	184
2	4	1	1681	MFG	40	49	125
2	3	22	1642	MFG	40	49	122
2	2	8	1568	TOV	42	51	116
2	1	30	1530	PF	44	54	114
3	11	12	1392	DRB	46	59	104
3	10	53	1373	MFT	46	59	103
3	10	53	1373	MFT	46	59	103
3	9	56	1316	TOV	48	62	99
3	7	35	1175	DRB	54	66	88
3	6	50	1130	MFG	56	68	84
3	5	40	1060	ORB	57	72	80
3	5	39	1059	2 PTS	59	72	79
3	3	37	937	DRB	64	76	70
3	3	32	932	2 PTS	66	76	70
3	2	58	898	MFG	66	76	67
3	2	30	870	TOV	66	76	65
3	2	13	853	MFG	66	78	64
3	2	10	850	ORB	66	78	64
3	2	8	848	AST	69	78	64
4	11	9	669	MFG	75	85	51
4	8	46	526	MFG	82	90	41
4	3	56	236	DRB	97	102	20
4	2	36	156	TOV	97	108	14

4	1	33	93	MFG	100	114	9
4	0	42	42	MFG	100	119	5

MFG: missed field goals; MFT: missed free throws

Once obtained the estimated possessions, we computed the value of each action using the difference of win probabilities. For assists and points made we had to consider the score of the reference team before the action and the score produced by the action, but for the remaining box-score variables we had to consider the score after the action and the potential score obtained adding 1 point to the reference team. Table 4 shows the value of each action:

Table 4: Value of each game action

Estimated possessions	PTS	MFG	MFT	DRB	ORB	AST	STL	TOV	BLK	PF
212				.0031300						
214						.0030864				
210		.0031150								.0031447
205	.0063791									
208		.0031141								
206	.0062562									
203				.0031719						
200						.0032350				
192						.0034010				
184	.0070143									.0034686
125		.0050016								
122		.0051162								
116								.0053624		
114										.0053508
104				.0056960						
103			.0114899							
99								.0059476		
88				.0067490						
84		.0070308								
80					.0071386					
79	.0144303									
70	.0164609			.0082305						
67		.0085429								
65								.0087636		
64		.0088766			.0088766	.0088766				
51		.0106824								
41		.0133454								
20				.0247934						
14								.0208333		
9		0,0046296								
5		0								
Sum	.0505408	.0694548	.0114899	.0517706	.0160153	.018599	0	.0409069	0	.011964
PTC Weights	1	-.91	-.57	.58	.92	.48	.86	-.86	.91	-.23
Weighted sum	.050540	-.06320	-.00654	.030027	.014734	.008927	0	-.03518	0	-.00275

PTC wp = -.003456
PTC wp/min= -.0001087

As Table 4 shows, PTC_wp is negative, what indicates that Doncic globally "hurted" his team, in the sense that his performance did not increase the probability of winning the game, but decreased it.

3.3 Simulating the variation of productivity

The method proposed to contextualize box-score statistics considering the dynamics of the game assigns a value of probability in function of the time remaining and the scoreboard. Therefore, and in order to exemplify the power of this method, we could simulate what would have been the PTC_wp of Luka Doncic under other circumstances of the game.

To achieve this simulation, we only changed the features of 5 of the 37 actions that Luka Doncic achieved. Specifically, we changed the time where he made his first two field goals, and the scoreboard of the 2 other field goals made. Changes are showed in Table 5, where his first 3 field goals where simulated to be at the end of the first quarter, and the other 2 field goals where simulated to be made being his team 10 points better (the score differential was then pretty even). Obviously, under those new circumstances estimated possessions also varied together with the assigned value of each of those 5 actions. As Table 5 shows, all those 5 actions would be now more valuable and the new computed PTC_wp would be positive (.003958) instead of negative (-.003456).

Table 5: Simulation of the variation of PTC_wp

Real performance						
Seconds to the end	**Action**	**Reference team scoreboard (Dallas)**	**Other team scoreboard (Suns)**	**Estimated possessions left**	**Assigned value**	**PTC_wp**
2749	2 POINTS	4	5	205	.0063791	-.003456
2692	2 POINTS	6	11	206	.0062562	
2448	2 POINTS	16	20	184	.0070143	
1059	2 POINTS	59	72	79	.0144303	
932	2 POINTS	66	76	70	.0164609	
Simulated performance						
Seconds to the end	**Action**	**Reference team scoreboard (Dallas)**	**Other team scoreboard (Suns)**	**Estimated possessions left**	**Assigned value**	**PTC_wp**
2269	2 POINTS	4	5	166	.0078420	.003958
2212	2 POINTS	6	11	165	.0077452	
2208	2 POINTS	16	20	164	.0078386	
1059	2 POINTS	69	72	77	.0160254	
932	2 POINTS	76	76	68	.0185034	

Note: Changes made in the initial conditions are shadowed

4. Discussion

We have proposed an alternative method to compute win probabilities in professional basketball based on theoretical probabilities. Instead of using million lines of historical data to compute empirical probabilities of every action under different set of game situations, we have derived a way to calculate empirical probabilities based on estimated possessions left.

This method allows calculating a context-dependent index of global productivity of players using the box-score variables. We have employed PTC (Martínez, 2019) [16] as the selected index, but other researchers could decide to use other, such as, for example, Game Score (Hollinger, 2005) [12] or Win Score (Berri, 2008; 2012) [7-8].

The value of each action is computed as the difference in win probability between the divergence in the scoreboard potentially produced by any of such actions. As other authors acknowledged before (e.g. Beouy, 2013; Deshpande & Jensen, 2016) [2, 9], to consider win probabilities changes substantially the interpretation of the box-score, providing a more real picture of the impact of each player to the performance of his team.

Our proposal does not underestimate the excellent contribution of other authors who based their proposal on empirical probabilities. Conversely, our method aims to be simply a new alternative to all of them. In fact, it has some similitudes with the Expected Win Probability Added (eWPA), developed by Beouy (2014) [3], which considers box-score stats to compute the added win probability expected.

One of the strength of our method is not to rely on the huge data necessary to empirically estimate win probabilities with a certain degree of reliability, because empirical models needs to consider all the combinations of situations. For example, and as only a rough estimation, if for every second of a 48 minutes' game (2880 seconds), 10 different actions can be registered (the box-score variables), then we have 28,880 possibilities. In addition, if we consider only the second half of the game where the score margin for a team could vary, for example, from -30 to +30 (61 possibilities), then the full possible scenarios would be 0.5*2880*10*61, i.e. 878,400 possibilities. Therefore, we would have almost 1 million of different possibilities only with the second half of a basketball game. Obviously we would need to add the disparate possibilities for the first half of game, and to expand the range of score differential from, for example -50 to +50. Consequently, we would have more than 1 million of disparate possibilities.

Again this reasoning does not pretend to devaluate methods based on empirical probabilities, just pretend to stress the difficulties of obtaining a reliable estimation of every possibility, even by handling several millions of data.

The analysis of clutch behavior is gaining attention in professional basketball. However, by only focusing in the last minutes of the game analysts could lose some valuable information regarding the value of each shot made during all the game, which also contributes to add probability to win the match. Therefore, we think the fairest way to evaluate player performance is to use any type of win probability model, although we agree with Zuccolotto, Marica & Sandri (2018) [21] that high pressure situations do not have to be necessary associated with high valuable shots from a win probability approach.

Limitations and further research

The proposed method has several limitations that must be considered. First of all, we also have relied on empirical data to estimate remaining possessions. Although win probabilities have been theoretically derived, we need to know the estimated possession left for every second of a game, and this have to be empirically based. We employed 36 games to do that (5,622 observations), and although we obtained a high pseudo R-square of our model (.86), obviously a more robust estimation would have been reached with a larger sample size. Further research could re-do the analyses increasing the sample size. We employed a manual coding to register all the 5,622 observations, so it was a high costly demanding procedure. However, we considered it necessary in order to obtain a reliable registration of all the data, because as explained in the methods section, play-by-play data are not always ordered in the same way, and, depending on the employed source, errors of automatic coding can be important.

Second, our proposal needs a way to be really implemented, i.e. to obtain PTC wp for every player in every game in an automatic form. Therefore, a computing program has to be developed to deal more easily with all the calculations needed to obtain the value of each action. Theoretical probabilities are easily available after applying our method, but the challenge would be to link every action with estimated possessions, and then to automatically compute the difference in winning probabilities for each estimated possession, considering that points and assists have to be programmed differently from the other box-score statistics, as we explained previously. Consequently, a necessary subsequent step for the use of our proposal is to build a program to make easier calculations.

And third, further modifications could be implemented regarding adding new box-score variables. For example, in the two most important professional leagues outside NBA (Euroleague, and the Spanish ACB League), fouls drawn are also displayed in box-scores, and they are considered to compute a global efficiency index of player productivity. Fouls drawn are, of course, available in the play-by-play data, so it could be also considered. Regarding the PTC measure we have employed, fouls drawn would have the same weight that fouls made, but with the opposite sign, i.e. its weight would be .23 of 1 point made.

5. Conclusion

This paper has proposed an alternative method to compute win probabilities which has been theoretically based, and that has been built upon the concept of estimated possessions. After taking into account the moment of time of each game action and the scoreboard differential, estimated possessions has been computed using a truncated Poisson regression model. Once obtained the estimated possessions, the value of each action has been derived from the difference in theoretical probabilities of the potential value of each action reflected in a change in the scoreboard differential. Therefore, box-score statistics can be weighted using a context-dependent system of evaluation, and then computing a global index of productivity. As an empirical example, Player Total Contribution (PTC) was taken as an index summarizing box-score variables, and it has been shown how this index can change depending on the variations in the time and scoreboard for every play of the game. Consequently, two players with the same box-score performance could have really contributed very different to the winning probability of a team. Future research is needed to make this procedure more easily implemented.

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.

6. References

1. 82games.com a study of the value of offensive rebounds, 2019. Retrieved from: http://www.82games.com/comm13.htm
2. Beouy M. NBA win probability calculator, 2013. Retrieved from: https://www.inpedictable.com/2013/11/nba-win-probability-calculator.html
3. Beouy M. New Feature: NBA win probability added totals by player, 2014. Retrieved from: https://www.inpredictable.com/2014/01/nba-player-win-probability-added.html
4. Beouy M. Updated NBA win probability calculator, 2015. Retrieved from: https://www.inpredictable.com/2015/02/updated-nba-win-probability-calculator.html
5. Beouy M. Heroes, goats and garbage men: New clutch shooting reports, 2016. Retrieved from: https://www.inpredictable.com/2016/01/heroes-goats-and-garbage-men-new-clutch.html
6. Berri DJ. Who is most valuable? Measuring the player’s production of wins in the national basketball association. Managerial and decision economics. 1999; 20(8):411-427.
7. Berri DJ. A simple measure of Worker productivity in the National Basketball Association. In Humphreys, B and Howard, D. (Eds): The Business of Sport; Westport, Conn. Praeger, 2008, 1-40.
8. Berri DJ. Measuring performance in the National Basketball Association. In Stephen Shmanske, S. and Kahane, L. (Eds): The Handbook of Sports Economics. Oxford University, 2012
9. 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.
10. Ganguly S, Frank N. The problem with win probability. MIT Sloan sports analytics conference, 2018.
11. Hardin JW, Hilber JM. Generalized linear models and extensions. Third Edition. Texas: Stata Press, 2012.
12. Hollinger J. Pro basketball forecast: 2005-2006. Potomac Books Inc, 2005.
13. Lock D, Nettleton D. Using random forests to estimate win probability before each play of an NFL game. Journal of quantitative analysis in sports. 2014; 10(2):197-205.
14. Lock D. Statistical methods in sports with a focus on win probability and performance evaluation. Graduate Theses and Dissertations. 15926. Iowa State University, 2016.
15. Martínez JA. Factors determining production (FDP) in basketball. Economics and business letters. 2012; 1(1):21-29.
16. Martínez 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.
17. McFarlane P. Evaluating NBA end-of-game decision-making. Journal of Sports Analytics. 2019; 5:17-22.
18. Rosenbaum D. Measuring how NBA players help their teams win, 2004. Retrieved from: http://www.82games.com/comm30.htm
19. Solomonov Y, Avugos S, Bar-Eli M. Do clutch players win the game? Testing the validity of the clutch player's reputation in basketball. Psychology of sport and exercise. 2015; 16(3):130-138.
20. Swann C, Crust L, Jackman P, Vella A, Allen SA, Keegan MSR. Performing under pressure: Exploring the psychological state underlying clutch performance in sport. Journal of sports sciences. 2017; 35(23):2272-2280.
21. Zuccolotto P, Manisera M, Sandri M. Big data analytics for modeling scoring probability in basketball: The effect of shooting under high-pressure conditions. International journal of sports science & coaching. 2018; 13(4):569-589.