International Journal of Physical Education, Sports and Health 2019; 6(3): 81-85

P-ISSN: 2394-1685
E-ISSN: 2394-1693
Impact Factor (ISRA): 5.38
IJPESH 2019; 6(3): 81-85

www.kheljournal.com
Received: 04-03-2019
Accepted: 06-04-2019

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

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

A more robust estimation and extension of factors determining production (FDP) of basketball players

Martínez, Jose A

Abstract
We have re-evaluated the research regarding how to measure the productivity of basketball players (FDP) through box-scores statistics. Using a much more comprehensive data set we have estimated the weights of the box-scores variables, and we have obtained a valid model, with a more reliable estimation of those weights, and a relevant improvement in the explained variance of the model. In addition, because of the availability of new information regarding turnovers assigned to teams, our new FDP metric is better from a theoretical viewpoint. Besides obtaining a more robust estimation of FDP, we have proposed a simple way to obtain a full index of productivity, considering points and blocks made. This new index (PTC) is an easily understandable and implementable metric that should help managers, media and fans to evaluate the performance of basketball players. Our contribution has important implications for managerial decision in professional basketball regarding signing players, and also simplifies understanding for media and fans.

Keywords: Sports economics, sports analytics, basketball, player performance, NBA

1. Introduction

In 2012, the author of [1] proposed a new method to evaluate the performance of basketball players through box-scores statistics. The method was called Factors Determining Production (FDP), because it separated points made from other game variables which may be easily registered. That separation was one of the main contributions of [1] in order to estimate consistent and valid weights for the simplest box-score variables.

The method employed data of 3,237 games of the 2007, 2008 and 2009 regular seasons of the National Basketball Association (NBA), and was validated using several procedures, including a performance test of the explained variance of the original model estimates with new data: 1,106 and 1,125 NBA games of the 2010 and 2011 regular season, and a sample of 485 games from the Spanish ACB League within the 1991-2010 period.

Therefore, an index of factors determining production (FDP) was written as:

FDP = .41 Defensive rebounds + .81 Offensive rebounds + .75 Steals minus Turnovers + .43 Assists - .82Missed field goals - .55Missed free throws – .23 Fouls

However, [1] did not estimate the parameters for the new test samples, but employed the estimated parameters obtained with the analysis of the first three seasons to test the predictive accuracy of the model for the new samples. Therefore, it is difficult to know the stability of those parameters with the new samples. In addition, original data did not consider play-off games, nor 10% of games placed in the tails of the distribution of scoring difference either.

The availability of a larger sample size of NBA data is an excellent opportunity to validate the FDP proposal, with a more robust estimation of the importance of the box-score variables to determine the margin of victory of basketball games. Therefore, the aim of this short communication is to study the validity of FDP original results and to provide a more robust assessment of the variables influencing game results under disparate scenarios of data.

Besides this more comprehensive evaluation of FDP, this research also proposes a new index of player performance, Player Total Contribution (PTC), which integrates all the box-score statistics in a single metric, i.e. it merges FDP with points made and blocks to provide a unique value representing the total productivity of a player.

2. Materials and methods

Box-scores from the 2007 to the 2018 NBA seasons were obtained from www.bigdataball.com, comprising a sample of 15,518 games, including regular season and play-offs. The re-evaluation of the FDP estimates was carried out using three disparate scenarios: (1) Replication of the framework proposed by [1], for the 2010 to the 2018 regular seasons, gathering observations in function of the margin of victory of the home team, and then deleting 5% of both tails of the distribution; (2) Re-analysis of all the data, merging the first 3 seasons with the remaining seasons, including play-offs games and the tails of the distribution; (3) The same as the prior analysis but considering a slightly (but important) modification of the turnover variable (total turnovers assigned to players) adding the total turnovers assigned to teams. This latter modification was not present in the data of the original publication of FDP.

The full data base is available as a supplementary from authors upon request, containing information regarding the season, date and an identifier code of each game, together with the home team margin (score difference between home and road teams). Box-score variables were: Missed field goals (MFG), Missed free throws (MFT), Defensive rebounds (DR), Offensive rebounds (OR), Assists (A), Fouls made (PF), Steals-Turnovers assigned to players (ST-TO1), and Steals-Turnovers assigned to players+teams (ST-TO2). Two additional variables indicating if matches were play-offs or regular season games, and if the case corresponded to the replication sample completed the data base.

3. Results

As [1] did, all the models were estimated using the ordinary least squares (OLS) method via the Stata 13.0 software, studying miss-specification tests [8]. Models had the following specification (1):

$$y_t = \beta_0 + \sum_{k=1}^{p} \beta_{kt} x_{kt} + u_t \qquad (1)$$

Where y_t is the home team margin, β_k are the *p* different weights (coefficients) of the *p* disparate x_k covariates (box-scores variables), $t \in \mathbb{N}$ are the observations of the sample, and u_t is a random error normally distributed with zero mean, which represents the unmeasured variables (intangibles) and a pure random component. It is a model assumption that $cov(x, u) = 0$. Finally, probabilistic assumptions are normality, linearity, homoscedasticity, independence and t-invariance [8]. Results are shown in Table 1, and the Stata codes are provided in the Appendix.

Table 1: Results of the OLS estimation and tests of assumptions

	FDP (original) n=3,327	Scenario 1 Replication n=9,714	Scenario 2 Extension 1 n=15,518	Scenario 3 Extension 2 n=15,518
Covariates	*Coef.*	*Coef.*	*Coef.*	*Coef.*
Missed field goals	-.82**	-.80**	-.93**	-.91**
Missed free throws	-.55**	-.49**	-.57**	-.57**
Defensive rebounds	.41**	.49**	.53**	.58**
Offensive rebounds	.81**	.80**	.94**	.92**
Assists	.43**	.42**	.49**	.48**
Steals-Turnovers	.75**	.73**	.85**	.86**
Fouls made	-.23**	-.23**	-.23**	-.23**
Intercept	.57**	.52**	.24**	.19**
Model assessment				
R^2	.72**	.73**	.80**	.81**
Breusch-Pagan/Cook-Weisberg for homocedasticity:	10.93	.79	1.81	.93
Ramsey RESET test for no omitted variables, using powers of the independent variables:	1.49	3.20**	2.56**	2.59**
Run test for independency of residuals, considering the sign of unstandardized residuals:	.47	-1.22	*.42	-.16
Skewness/Kurtosis tests for normality of residuals:	21.66**	55.55**	21.63**	19.73**

**$p<0.05$

Replication of the FDP study yielded similar coefficients, with slightly modifications in some of them. Explaining variance was also equivalent (.73 vs .72). However, scenarios 2 and 3 provided models with better explained variance (.80 and .81). In addition, the reliability of the estimated coefficients was higher because of the larger sample size compared with the original study and the replication.

As scenario 3 had the more complete information regarding turnovers, this is the most theoretically recommended. Miss-specification tests reported in Table 1 supported the assumptions of homoscedasticity and independence, but not normality of residuals, probably because of the large sample size (high power). However, histogram of residuals showed that its distribution was approximately normal (Figure 1).

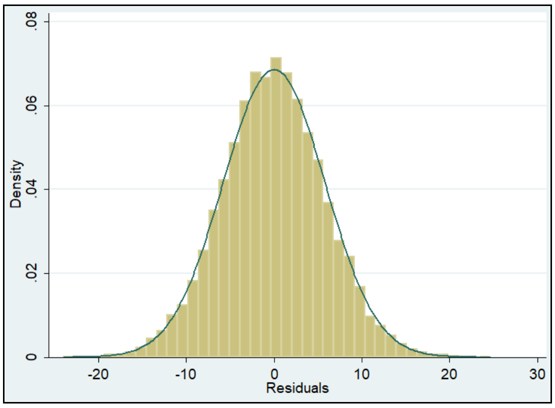

Fig 1: Histogram of the residuals with normal curve for scenario 3

Although the Ramsey RESET test provided significant results

(again probably caused by the sensibility to the large sample size), we believe this is not truly relevant. Moreover, we estimated an alternative model with the squared terms and all the possible interactions among all the covariates. The explained variances were practically the same: .8099 for the simplest model vs .8102 for the more complicated one, i.e. less than .04% of variation, what is a truly negligible change.
In order to reinforce the study of probabilistic assumptions we tested two additional models following the recommendations of [6] to test for linearity, homoscedasticity, independence and t-invariance (2) (3):

$$\hat{u}_t = \gamma_0 + \sum_{k=1}^{p} \gamma_{kt} x_{kt} + \delta_1 t + \delta_2 t^2 + \sum_{k=1}^{p} \lambda_{kt} x_{kt}^2 + \delta_3 \hat{u}_{t-1} + v_t \quad (2)$$

$$\hat{u}_t^2 = \gamma_0 + \delta_1 t + \sum_{k=1}^{p} \lambda_{kt} x_{kt}^2 + \delta_2 \hat{u}_{t-1}^2 + v_t \quad (3)$$

Those auxiliary regressions tested for the assumptions of linearity ($\lambda_{kt} = 0$), independence ($\delta_3 = 0$), and t-invariance ($\delta_1 = 0, \delta_2 = 0$) for the first model, and homoscedasticity ($\lambda_{kt} = 0, \delta_2 = 0$), and t-invariance ($\delta_1 = 0$) for the second model.
Results indicated that all the assumptions were met for the first model, because F test for the joint significance of the coefficients of both models were non-significant; $F(17;15{,}499)=1.23$; p:.23. However, for the second model: $F(9;15{,}507)=5.71$; $p<.001$, which indicated some departures from the homoscedasticity and t-invariance assumptions. Given that the first model provided evidence supporting t-invariance, and given the Breusch-Pagan/Cook-Weisberg test supported homoscedasticity, we believe that, considering all the tests achieved, our target model (scenario 3) met the probabilistic assumptions reasonably well.
Therefore, a re-evaluated index of factors determining production (FDP) should be written as follows:

FDP = .58 Defensive rebounds + .92 Offensive rebounds + .86 Steals minus Turnovers + .48 Assists - .91 Missed field goals - .57 Missed free throws – .23 Fouls

3.1 Towards a total contribution index
In the original work of [1], the author distinguished between three level variables: points, FDP and blocks: “In the first level would be production, i.e. the points made by the two teams in a contest which reflect the final result of a game. In a second level would be FDP, i.e. quantifiable factors which help to explain why the first level variable varies. And then, a third level of analysis would be the factors who determine variation in FDP, being blocks the only quantifiable variable. It would be recommendable to be conservative and not to aggregate these three levels in a single metric”.
However, and in order to increase the utility and practical use of the FDP approach, we propose a way to merge FDP with the other two level variables in a single metric. The reasoning is straightforward: as a block made has the same meaning that forcing a missed shot to the rival team, then blocking a shot has the same statistical effect that a missed field goal, but with the opposite sign, i.e. it has a value of .91. Regarding points made, each point made contributes to the score differential, i.e. to the margin, so each point made has a value of 1.
Consequently, we could simplify all the box-scores contribution in a single metric by employing what we have a called PTC (Player Total Contribution) as follows:

PTC = 1 Points made + .91 Blocks made + FDP

Or equivalently:

PTC = 1 Points made +.91 Blocks made + 58 Defensive rebounds + .92 Offensive rebounds + .86 Steals minus Turnovers + .48 Assists - .91 Missed field goals - .57Missed free throws – .23 Fouls

Both FDP and PTC should be displayed per game or per minutes played (we recommend per minutes played) in order to obtain a comparable index among all basketball players.

3.2 Evaluating player production
We obtained from www.basketball-reference.com the box-score statistics of all players for the 2017-18 NBA regular season. After considering only players who played at least 500 minutes of more, we ranked the first 30 players by PTC/min. We compared this ranking with the MVP points obtained by the 13 players who got votes in the same season. Recall the MVP points were computed mainly from votes of a panel of media members and very small contribution from fans votes. Table 2 shows the results, and as it can be seen, the seven first players ranked by PTC/min were seven of the thirteen players who obtained points in the election of the MVP award.

Table 2: NBA players ranked by PTC/min for the 2017-18 regular season (at least 500 minutes played)

Ranking	Player	PTC/min	MVP ranking	MVP points
1	Anthony Davis	.792	3	445
2	Stephen Curry	.715	10	5
3	LeBron James	.710	2	738
4	James Harden	.703	1	965
5	Giannis Antetokounmpo	.702	6	75
6	Kevin Durant	.699	7	66
7	Joel Embiid	.689	12	4
8	Hassan Whiteside	.684		
9	Clint Capela	.674		
10	Enes Kanter	.673		
11	Karl-Anthony Towns	.667		
12	DeMarcus Cousins	.656		

13	Montrezl Harrell	.651		
14	Jonas Valanciunas	.646		
15	LaMarcus Aldridge	.641	9	6
16	Nikola Jokic	.636		
17	Andre Drummond	.633		
18	JaVale McGee	.628		
19	Kevin Love	.615		
20	Russell Westbrook	.611	5	76
21	Dwight Howard	.603		
22	Kyrie Irving	.600		
23	DeAndre Jordan	.597		
24	Greg Monroe	.581		
25	Damian Lillard	.577	4	207
26	Julius Randle	.577		
27	David West	.576		
28	Chris Paul	.566		
29	Rudy Gobert	.562		
30	Victor Oladipo	.559	13	2

Although this is not a criterion for validation itself because one of the objectives of basketball analytics is to overcome the possible subjective biases in the evaluation of player performance, it can be interpreted as a signal that results were not incongruent with common knowledge in basketball.

4. Discussion

In this short communications we have re-evaluated the research of [1] regarding how to measure the productivity of basketball players (FDP) through box-scores statistics. Using a much more comprehensive data set we have estimated the weights of the box-scores variables, and we have obtained a valid model, with a more reliable estimation of those weights, and a relevant improvement in the explained variance of the model.

In addition, because of the availability of new information regarding turnovers assigned to teams, our new FDP metric is better than the original FDP from a theoretical viewpoint.

Besides obtaining a more robust estimation of FDP, we have proposed a simple way to get a full index of productivity, considering points and blocks made. This new index (PTC) is an easily understandable and implementable metric that should help managers, media and fans to evaluate the performance of basketball players.

As [1] pointed out, in order to facilitate managerial decisions, players should be compared by position, and also by the PTC/min of an average league player. Futher research could re-evaluate or improve this metric if new relevant statistics are routenily included in the box-score data sets. For example, fouls drawn could be an interesting target variable for that purpose. Nevertheless, our box-score data set (from www.bigdataball.com) did not have information regarding this variable, only about fouls made.

Fouls drawns, however, could be incorporated to the PTC index in the same form that blocks made was considered. As almost every foul made means a fould drawn by the opponent team (except for technical fouls), then an option would be to consider the same weight but with the opposite sign. Therfore, the final PTC proposal when fouls drawn are routinely available in the box-score of every game (e.g. Euroleague, ACB Spanish League) is the following:

PTC = 1 Points made +.91 Blocks made + 58 Defensive rebounds + .92 Offensive rebounds + .86 Steals minus Turnovers + .48 Assists - .91 Missed field goals - .57Missed free throws – .23 Fouls made + .23 Fouls drawn.

As every quantitative metric, it has a limited scope, because it is not able to consider all the intangible elements of the player performance. We agree with [9] in thar there is no holy grail in basketball metrics and metrics serve mostly to complement traditional scouting. However, due to its building process, its is explain more than 80% of the variability in score differential, which reinforcces its utility as a sound proxy of productivity. This is the reason why, in this case, we are not in agreement with [10] regarding that about 80% of what happens in a basketball game is not reflected in the box-score. It is true that to obtain a full picture of a game we have to go beyond box-scores, but our proposal have shown that precisely we may explain 81% of variation of score differential with only box-score statistics, which indicates that, well analyzed, box-scores could be highly useful.

Nevertheless, a major limitation of indexes based on box-score variables is the lack of context to evaluate every action. There are players that perform better than the average in the moments of the game where pressure is high [11], and also it is acknowledged that the value of every action depends of the specific win probability assigned to it [3]. Authors such as [3] have proposed a dynamic approach to consider all those contextual factors, which are based on empirical probabilities of huge amounts of historial data. We offer, however, an alternative to such meritorius approach, quantifying the value of the box-score variables a valid proxy for estimating their productivity in a game. Further research could try to integrate each component of the PTC index, i.e. the box-score variables, in a dynamic approach weithing again all those variables in function of the win probabilities linked to any action. The required method would be different to [3] because there would not be neccesary to empirically analyze several million lines of play-by-play historical data, but computing theoretical win probabilities in function of the scoreboard and the time left to the end of a game. Future studies could also propose a way to consider the difficulty of each shot based on the distance to the rim, following the proposal of [12] in a dynamic metric based on win probabilities.

Finally, we encourage researchers to propose improvements of our approach using the raw data base.

5. Conclusion

We have provided a valid and reliable estimation of the factors determining production (FDP) of basketball players, improving the original proposal of [1], and justifying the creation of a new index, player total contribution (PTC), which joins FDP to points and blocks made, with the option to also include fouls drawn. Both FDP and PTC are theoretically and empirically well grounded, offering PTC an alternative to the current methods of summarizing box-score statistics. Given the importance of basketball analytics in the modern business of professional basketball, and acknowledging the need to offer a single and easily understandable valid metric for all the stakeholders of the business (including teams, media and fans), our research can be a valuable reference for advancing.

6. Acknowledgments

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.

7. References

1. Martinez JA. Factors determining production (FDP) in basketball. Economics and Business Letters. 2012; 1(1):21-9.
2. Rosenbaum D. (Internet). Measuring How NBA Players Help Their Teams Win; Available from: http://www.82games.com/comm30.htm
3. Beouy M. (Internet). NBA win probability calculator; Available from: https://www.inpredictable.com/2013/11/nba-win-probability-calculator.html
4. 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.
5. 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.
6. Kubatko J. (Internet). Calculating PER. Available from: https://www.basketball-reference.com/about/per.html
7. Berri DJ, Bradbury JC. Working in the land of metricians. Journal of Sports Economics. 2010; 11(1):29-47.
8. Spanos A. Mis-specification testing in retrospect. Journal of Economic Surveys. 2018; 32:541-77.
9. Oliver D. (Internet). How numbers have changed the NBA. Available from: http://www.espn.com/nba/story/_/id/9980160/nba-how-analytics-movement-evolved-nba
10. Winston WL. *Mathletics*. New Yersey: Princeton University Press, 2009.
11. 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):569-89.
12. Piette J, Anand S, Zhan K. Scoring and shooting abilities of NBA players. Journal of Quantitative Analysis in Sports. 2010; 6(1):1.