

Factors determining production (FDP) in basketball

Jose A. Martinez*

Universidad Politecnica de Cartagena, Spain

Received: 21 February 2012

Revised: 20 March 2012

Accepted: 21 March 2012

Abstract

The aim of this research was to introduce a simple and easily computable metric to assess the performance of basketball players through non-scoring box-score statistics. This metric was called Factors Determining Production (FDP). FDP was created through separating points made from the remaining variables which may be quantitatively recorded. FDP was derived from the outcome of several games, it considers both teams' statistics, and it reflects the final result of a game with noticeable merit. This metric provides a simple linear weight formula which, together with the points made by each player, yields a comprehensible picture of how well a worker (player) performed. FDP has been validated through different statistical procedures and it overcomes Win Score from a theoretical viewpoint, because it departs production (points) from factors facilitating production.

Keywords: Sports economics, basketball, player performance

JEL Classification Codes: M21, M11

1. Introduction

To find best measure to value a basketball player is becoming like the search of the "Holy Grail" for basketball economics. The abundance of objective data about performance of workers (players) is no comparable with any other industry and with any other sports either. Therefore, in the last years, psychologists, economists, engineers and statisticians have proposed a plethora of methods to quantitatively measure the productivity/performance/value of players (Berri, Brook & Schmidt, 2006; Winston 2009) in order to obtain the most objective measure on the contribution of players to teams success, and consequently, to get a tool to make managerial decisions about these players with lesser risk.

*Corresponding author. E-mail: josean.martinez@upct.es.

Citation: Martinez, J.A. (2012) Factors determining production (FDP) in basketball, *Economics and Business Letters*, 1(1), 21-29.

Among the hundreds of methods proposed (see Martínez, 2010, for a comprehensive review), the works of Berri and colleagues are the only which have been published in academic journals within the field of economics and business. The “Wins Produced” and its simplification “Win Score” are the two most accepted ways to rate players by the academic community. The works of Berri (1999; 2008; 2012), Berri, Brook and Schmidt (2006) or Berri and Bradbury (2010) explain the rationale of this method, the important limitations of other famous metrics such as “Efficiency”, “PER”, “Plus-Minus” and “Adjusted Plus-Minus”, and why Wins Produced overcomes such metrics. Berri (1999; 2008) links box-score statistics with team wins through regression analysis. Therefore wins are determined by both offensive efficiency and defensive efficiency, where offensive efficiency is defined as points scored divided by possessions employed and defensive efficiency is defined as points surrendered divided by possessions acquired. Using NBA team data from 22 seasons, results indicate that 95% of the variation in winning percentage can be explained by offensive and defensive efficiency. In particular, Berri proposes a series of sequential analyses, based on a subset of structural equations to derive the value of each box-score variable from the offensive and defensive efficiency, i.e. the impact of points, missed shots, rebounds, assists, steals, turnovers, blocks and fouls on wins. Therefore, the final outcome of the analysis may be resumed in a simpler index “Win Score”, where the marginal effects of points, rebounds and steals are positive, and turnovers and field goals attempted are negative. These five variables have the same marginal effect on wins. Finally, assists and blocks have a positive effect and free throws attempted and fouls have a negative effect. These latter four variables contribute only 0.5 times the first five variables. To summarize:

$$\text{Win Score} = \text{Points} + \text{Rebounds} + \text{Steals} + 0.5 \text{ Assists} + 0.5 \text{ Blocks} - \text{Turnovers} - \text{Field Goals Attempted} - 0.5 \text{ Fouls} - 0.5 \text{ Free Throws Attempted}$$

This metric can be easily computed from the box-score of each game and it is accessible to analysts and fans. Win Score’s correlation with WinProduced is 0.99, when both variables are normalized per minute (Berri & Schmidt, 2010). Therefore, the major contribution of Berri (1999; 2008) is to propose an easily understandable metric, which may be manually computed for any person, and which exceptionally correlates with team wins¹.

However, and acknowledging the merit of this metric, it has some important concerns. First of all, it has not been derived directly from box-score data, but from aggregate data of wins, offensive and defensive efficiency. It would be desirable to find a metric which would reflect the result of a game with the best level of accuracy. For example, in the extreme case that the guard of a team loses the ball in every play (let’s suppose 90 plays for his team in a game), and the guard of the rival team always get an easy 2 points lay-out from stealing the ball, then the final result of the game would be 180-0, and the Win Score of the first player would be -90, and the Win Score of the second player would be 270 (180 points plus 90 steals). Therefore, the difference of Win Score between the two teams at the end of the game would be 360 (270 – (-90)), which is two times the real difference in the score. Moreover, if the player who steals the ball always assists other

¹ At 2011 December, David Berri actualized its formula and modified the value of rebounds (<http://wagesofwins.com/2011/12/11/wins-produced-comes-back-better-and-stronger/>). Therefore, defensive rebounds count 0.5 times offensive rebounds. This modification has been explained in his web site but it has not been published in any academic journal. Anyway the correlation of this new metric with the prior metric is 0.98. Consequently, this does not significantly modify my empirical analyses neither the theoretical criticisms about the Berri and colleagues work.

player to get the lay-out, then the Win Score of the winning team would be 315, because the value of the $0.5 \cdot 90$ assists is added, and then the difference of Win Score between the two teams would be 405, which is 2.25 times the real difference in the score. However, as well in the extreme case of a player gets the first two points of a game, and then the other team missed all their shots grabbing all the offensive rebounds, then the final score would be 2-0, and the difference in Win Score would be 2, which comes from the two points made of the first team, and the zero sum of the missed shots and grabbed rebounds of the second team. These extreme examples, commented for illustrative purposes only, show that, in some cases, Win Scores correctly reflects the final result of a game, but in other cases it clearly fails to reflect the reality of a game. Although Win Score, (and other measures based on linear weights) are free-scale measures, i.e. units of measures are arbitrary, the difference of the Win Score made by two teams in a game should approximate the final margin of victory, i.e. the difference between the points made by two teams. This latter idea is the basis of the Adjusted Plus/Minus metric (Winston, 2009). However, to properly use this metric players have to be rated after playing a great amount of games, so Adjusted Plus/Minus is not available to get insights of player performance from a single game. In addition, Berri (2012) details other criticism of this metric, based in the large standard errors of the estimates and the lack of consistency: Berri and Bradbury (2010) report that only 7% of a player's adjusted plus-minus value in the current season can be explained by what the player did in the prior season.

Secondly, and the most important, Win Score mixes "apples and oranges", because it puts at the same level points and the remaining statistics. And this is not recommendable. Points are the outcome of the process of play; they are the production of a team. Rebounds, assists, missed shots, etc. they are factors contributing to enhance or decrease the score difference between two teams in a contest. They are factors facilitating the outcome, but they are not an outcome *per se*. This point deserves more explanation:

Consider, for example, two sellers competing to sell more products to a subset of customers. Sales are the desired outcome, and seller A will obtain a prize if he defeats seller B, i.e. if A sells more products than B to customers. Which are the factors determining variation in the difference of sales between A and B? These factors can be quantifiable (number of the contacts made with the client, price of the products offered, etc.) and intangibles (dealing skills, ability to persuade others, luck, etc). But it would be nonsense to speak about if A defeats B in terms of sales, because sales are precisely what they are competing for. Consequently, following this simple analogy, points (outcome) and the remaining box-score statistics (factors affecting outcome) should not be put at the same level. So which are the determinants of the result of a game? Obviously the non-scoring variables and the intangibles (leadership of players, sacrifice, team chemistry, etc.) because the result of a game is the difference between the points made by two teams. In addition, a random component must be added, because randomness plays an important role in the outcome of sporting games (Berry, 2006). Therefore, a straightforward regression model can be proposed, as it will be subsequently explained.

Therefore, the aim of this research is to introduce a new method to value basketball players using simple and easily understandable box-score statistics, through separating points made from the remaining variables which may be quantitatively recorded. This metric is derived from the outcome of several games, it considers team and opponent statistics, and it reflects the final result of a game with noticeable merit. It provides a simple linear weight formula which, together with the points made by each player,

yields a comprehensible picture of how well a worker (player) of a basketball team performed.

2. Method

Box-scores of the 2007, 2008 and 2009 NBA regular seasons were collected from www.nbastuffer.com. The initial sample was composed of 3690 games. I ordered observations by the margin of victory of the home team, and I carried out a filtering process by deleting 5% of both tails of the distribution following Wilcox (2010). The final sample was composed of 3327 games, and margin ranged from -18 to +24 points. Several variables were recorded: The margin of the home team, i.e. the difference of the points made by the two teams (Y_i), and the difference between the two teams of the remaining non-scoring statistics (X_{ki}): missed field goals, missed free throws, defensive rebounds, offensive rebounds, assists, fouls, and the difference between steals and turnovers. Therefore, a simple econometric model was proposed: $Y_i = \beta_0 + \beta_k X_{ki} + e_i$, where β_k are the effects of the k non-scoring statistics (factors determining production) on the dependent variable (production) for each i game observed, and e is a random error normally distributed with zero mean, which represents the unmeasured variables (intangibles) and a pure random component. It is assumed that $cov(X, e) = 0$.

3. Results

Model was estimated via OLS. Results are showed in Table 1. Spanos's (2007) approach was employed in order to achieve an inductive process of learning from data, using a statistical model and testing its assumptions regarding error. The idea behind this approach is that to secure the reliability of any inductive inference one should validate the underlying inductive premises by probing for all possible errors. Model should account with data regularities, obtaining a white noise error term, i.e. a random component without systematic contamination. Therefore, reliability of the model is primary addressed by testing the model assumptions using miss-specification tests.

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

Variable	Coefficient
Missed field goals	-.82**
Missed free throws	-.55**
Defensive rebounds	.41**
Offensive rebounds	.81**
Assists	.43**
Steals-Turnovers	.75**
Fouls	-.23**
Intercept (β_0)	.57**
R ²	.72**

^a Breusch-Pagan / Cook-Weisberg for homocedasticity: 10.93

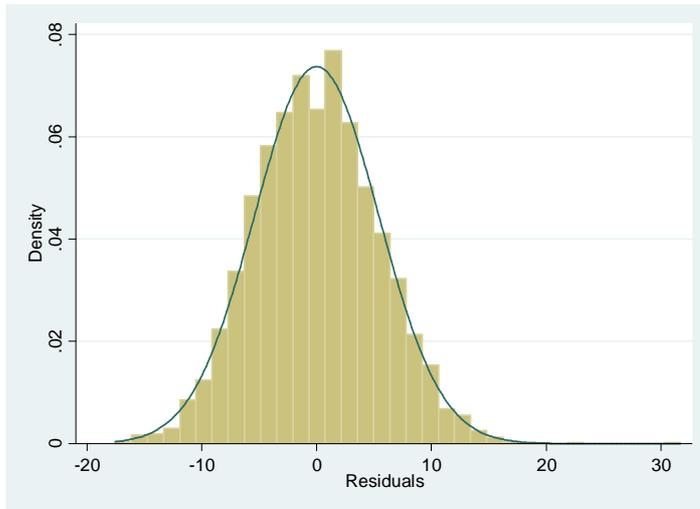
^b Ramsey RESET test for no omitted variables, using powers of the independent variables: 1.49

^c Run test for independency of residuals, considering the sign of unstandardized residuals: .47

^d Skewness/Kurtosis tests for normality of residuals: 21.66**

**p<0.05

Figure 1. Histogram of the residuals with normal curve



All miss-specification tests support the assumptions, except normality of residuals, because of the large sample size (high power). However, histogram of residuals shows that distribution is approximately normal (Figure 1). Model explains 72% of the variation in the dependent variable. This is a considerable explained variance, taking into account that no quantifiable factors such intangibles and the inherent random component also determine the margin of victory. Unstandardized regression coefficients are also marginal effects, because all variables are measured in comparable and similar units. Therefore, interpretation of coefficients indicate that when the difference of missed shots between two teams increases by one unit, i.e. an additional shot missed, then margin of victory decrease by 0.82 points. Moreover, when the difference of defensive rebounds increases by 1 rebound, then margin also increases by 0.41 points. A similar interpretation can be made for the remaining variables.

Consequently, an easily computable formula can be proposed rounding the estimates of the parameters (Berri, 2008, also made a similar procedure). Therefore, an index of factors determining production (FDP) can be written as:

$$FDP = 0.41 \text{ Defensive rebounds} + 0.81 \text{ Offensive rebounds} + 0.75 \text{ Steals minus Turnovers} + 0.43 \text{ Assists} - 0.82 \text{ Missed field goals} - 0.55 \text{ Missed free throws} - 0.23 \text{ Fouls}$$

Validation

This new proposed metric, FDP, has been theoretically justified and empirically sustained through regression analysis and miss-specification tests. However, it also should fulfill two conditions Berri (2012) recommends: it has to be linked with wins and it has to be consistent.

The fulfillment of the first condition is almost automatically derived from the construction of FDP, because FDP links some box-score statistics with the margin of victory, and this latter variable indicates a team wins when $\text{margin} > 0$ and a team losses when $\text{margin} < 0$. However, I considered the winning percentage of the NBA teams at the end of the regular season and the difference in FDP obtained by these teams. Data from www.basketball-reference.com about team and opponent statistics from the last 5 seasons (from 2007 to 2011) was employed. Pearson correlations between the winning percentage and the difference in FDP across the five seasons considered ranged from 0.93 to 0.95, which obviously is a very noticeable association.

Regarding consistency, the per-minute FDP and Win Score of some players were computed, in order to achieve a comparison. Berri (2012) indicates that Win Score is highly consistent over time for players, and it is much more consistent than other advanced metrics, such as Adjusted Plus/Minus. Fifteen players of the Lakers', Celtics'' and Mavs' roster of the 2011 season were considered. All these players have a long career in the NBA, with at least 10 years of experience. Taking standard deviation as a measure of consistency or dispersion from the mean, and normalizing both measures in a (0,1) scale, then results show that FDP is consistent (Table 2). Note that standard deviations are quite similar between both variables. Therefore, FDP is quite similar to Win Score with regard to consistency.

Table 2. Consistency of FDP vs. Win Score (standard deviation normalized*)

Player	FDP/min	Win Score/min
Kobe Bryant	0.17	0.17
Pau Gasol	0.22	0.10
Lamar Odom	0.31	0.23
Derek Fisher	0.33	0.26
Ron Artest	0.29	0.26
Kevin Garnett	0.33	0.31
Ray Allen	0.31	0.26
Jermaine O'Neill	0.26	0.25
Shaquille O'Neill	0.22	0.21
Paul Pierce	0.27	0.27
Jason Kidd	0.24	0.26
Shawn Marion	0.26	0.27
Dirk Nowitzki	0.23	0.27
Peja Stojakovic	0.31	0.29
Jason Terry	0.27	0.32

* Procedure proposed by Cohen et al (1999)

Finally, FDP for the box-scores of the following two seasons in the NBA (2010 and 2011) and for a sample of box-scores of the Spanish ACB League from 1991 to 2010 was calculated. Observations after deleting outliers were 1106, 1125 and 485, respectively. Prediction of the model in these three samples was computed, using the fixed parameters estimated following the suggestions of Schmueli, Patel and Bruce (2007). The aim was to assess the performance of the model with new data. Explained variance for each model was 0.73, 0.74 and 0.75, which was akin to the 0.72 obtained for the original model.

4. Concluding remarks

This paper has introduced a simple and easily computable metric to assess the performance of basketball players through non-scoring box-score statistics: Factors Determining Production (FDP). This metric has been validated through several statistical procedures and it overcomes Win Score from a theoretical viewpoint, because it departs production (points) from factors facilitating production.

FDP has been built from individual box-scores data, taking statistics of both teams into account, and it explains a considerable variance (72%) of the final margin of victory. The remaining unexplained variance would be attributed to intangibles, i.e. factors such as leadership, concentration, defensive attitude, etc. These factors are highly complicated to quantify and obviously contributes to the final result of a game, and also to the winning percentage of a team. For this reason, FDP is also a more realistic metric than Win Score, because the rationale of this latter metric is based on the fact that box-scores explain about 95% of wins. The missing 5% would not be intangibles, because winning teams tend to accumulate more points (and other stats) than the team needs to win the game. These excess stats tend to even out over the course of the season, and hence the 5% error. Consequently, Berri (2008) essentially explains the success of a team only from nine quantifiable variables, which is difficult to accept for the vast majority of the basketball community (Martínez & Martínez, 2011).

Further research should go deep into how to value blocks. This study does not consider such statistics for building FDP, because blocks made by a team are counted as missed shots for the other team. Therefore, blocks would be a “third level” variable: 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. Appendix shows the per-minute three level metrics for the All-Star players in 2011 (ranked by FDP). Consequently, managers should search for players with higher levels of these metrics, considering that if a player is a very good scorer but he has a low FDP, then other players of the team should compensate this probable disequilibrium. In addition, in order to facilitate managerial decisions, players should be compared by position, and also by the FDP of an average player of the league.

Acknowledgements

The author thanks editor and an anonymous reviewer for their helpful suggestions. In addition, thanks to Serhat Ugur, from www.nbastuffer.com, for his support in the development of this paper. Finally, thanks to David Berri for his assistance in solving some concerns regarding his work.

References

- Berri, D.J, L. Brook, S.L. and Schmidt, M.B. (2006) *The Wages of Wins: Taking measure of the many myths in modern sport*, Stanford University Press.
- Berri, D.J. (1999) Who is most valuable? Measuring the player’s production of wins in the National Basketball Association, *Managerial and Decision Economics*, 20(8), 411-427.
- Berri, D.J. (2008) 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: 1-40.
- Berri, D.J. (2012) Measuring performance in the National Basketball Association. In Stephen Shmanske, S. and Kahane, L. (Eds): *The Handbook of Sports Economics*. Oxford University
- Berri, D.J. and Bradbury, J.C. (2010) Working in the land of metricians, *Journal of Sports Economics*, 11(1), 29-47.

- Berri, D.J. and Schmidt, M.B. (2010) *Stumbling on Wins: Two economists explore the pitfalls on the road to victory in professional sports*, Financial Times Press (Princeton, N.J.)
- Berry, S.M. (2006) Statistical fallacies in sports, *Chance*, 19(4).
- Cohen, P., Cohen, J., Aiken, L., and West, S. (1999) The problem of units and the circumstance for POMP, *Multivariate Behavioral Research*, 34(3), 315-346.
- Martínez, J. A. (2010) Una revisión de los sistemas de valoración de jugadores de baloncesto (I). Descripción de los métodos existentes, *Revista Internacional de Derecho y Gestión del Deporte*, 10, 37-77.
- Martínez, J.A. and Martínez, L. (2011) A stakeholder assessment of basketball player evaluation metrics, *Journal of Human Sport & Exercise*, 6(1), 153-183.
- Shmueli, G, Patel, N. R. and Bruce, P. (2007) *Data mining for business intelligence: Concepts, techniques and applications in Microsoft Office Excel with XLMiner*. Wiley-Interscience.
- Spanos, A. (2007) Curve fitting, the reliability of inductive inference, and the error- statistical approach, *Philosophy of Science*, 74, 1046–1066.
- Wilcox, R. R. (2010) *Fundamentals of Modern Statistical Methods: Substantially Improving Power and Accuracy*, Springer, 2nd Ed.
- Winston, W. (2009) *Mathletics: How gamblers, managers, and sports enthusiasts use mathematics in baseball, basketball, and football*, Princeton University Press.

Appendix. Per-minute three level metrics for the All-Star players in 2011 (ranked by FDP)

Player	Points	FDP	Blocks
Rajon Rondo	0.2858	0.0286	0.0044
Kevin Love	0.5648	0.0242	0.0103
Al Horford	0.4359	0.0134	0.0296
Chris Paul	0.4391	0.0050	0.0017
Kevin Garnett	0.4767	-0.0075	0.0248
Pau Gasol	0.5065	-0.0076	0.0427
Tim Duncan	0.4735	-0.0147	0.0676
Dwight Howard	0.6083	-0.0406	0.0634
Chris Bosh	0.5131	-0.0718	0.0175
Blake Griffin	0.5937	-0.0722	0.0145
Deron Williams	0.3958	-0.0757	0.0066
Ray Allen	0.4574	-0.0894	0.0059
Paul Pierce	0.5443	-0.0994	0.0187
LeBron James	0.6887	-0.1070	0.0160
Manu Ginobili	0.5747	-0.1117	0.0116
Dirk Nowitzki	0.6714	-0.1178	0.0184
Russell Westbrook	0.6283	-0.1287	0.0105
Dwyane Wade	0.6884	-0.1404	0.0309
Joe Johnson	0.5133	-0.1457	0.0027
Amare Stoudemire	0.6867	-0.1541	0.0523
Derrick Rose	0.6688	-0.1683	0.0168
Kevin Durant	0.7104	-0.1719	0.0250
Carmelo Anthony	0.7274	-0.1781	0.0153
Kobe Bryant	0.7475	-0.2041	0.0043

Note: Some of the best scorers such as Bryant, Anthony or Durant have low ratings of FDP. These are “paradoxical players”, because they contribute to team success making more points, but they also contribute to team failure because they have a large negative FDP.