A Novel Approach in Evaluating Efficiency of Basketball Players

UDC: 005.336.1:796.323.2.071.2
DOI: 10.7595/management.fon.2013.0012

1. Introduction

All the way to the nineties of the 20th century, the sports economics and management was regarded as a hobby for a minority of economists and managers who were primarily involved in other areas of research and business. Modern sports, however, value the significance of the management and economy as impact of sports on the world economy. This has coincided with an equivalent rise in the volume of economic literature devoted to the study of sport and because of that more and more attention is paid to the analysis of sports teams and athletes.

It is demonstrated that studying the relationships between sport management and sport results can be empirically analyzed and verified with the usual economic and econometric methodology. That relationship is called sports analytics. Sports analytics is comprised primarily of statistical analysis (t-test, \textit{t}^2 test, ANOVA, descriptive statistics), analysis of efficiency and more recently sports data mining. Usually, events on the field, like number of goals, passes or assists, are being analyzed in order to improve team results and identify weaknesses of opponents. However, with the growth in popularity and the amount of capital invested in sport, sports analytics becomes increasingly oriented to the events off the pitch. Efficiency analysis in sports began with the work by Scully (1974) on baseball and Zak et al. (1979) on basketball. After success in quantifying the relationship between sporting inputs and sporting success by the aforementioned authors, efficiency analyses founded their application not only in basketball (Lee & Worthington 2012, Moreno & Lozano 2012, Hill & Jolly 2012), but in many other sports like football (Ribeiro & Lima 2012, Fernandez et al. 2012), baseball (Jane 2012, Regan 2012) or chess (Jeremic & Radojicic 2010).

Inspired by these and other works, the purpose of this paper is to provide a comprehensive assessment of the National Basketball Association (NBA) players’ efficiency. Fortunately, research in sports analytics and sports economics has recently embraced statistical and mathematical methods for the assessment of sporting efficiency. Those new methods are a very important development as these theoretical and empirical relationships are useful for management decision-making processes such as hiring, amnesty, play positions, minutes, play combining and salaries. According to those methods, one can successfully manage team or player performance, as well as club financial stability. Based on the results a club may even decide to...
amnesty contract of some players, which are not efficient, in order to get better financial and sport results. The remainder of the paper is structured as follow. Section 2 explains methodology. Section 2 is divided into DEA and DBA. Section 3 explains findings and analysis and is divided to ranking of NBA players and comparison of DEA and DBA to other NBA efficiency methods. Section 4 concludes the paper.

2. Methodology

The NBA players who play at the guard position (shooting guard or point guard) and had good results in the season 2011/12, are used for efficiency analysis. For this purpose, DEA and DBA are used.

Data envelopment analysis

The Data envelopment analysis (DEA) is an increasingly popular non-parametric method for relative efficiency evaluation. It allows performance measurement of the decision making unit (DMU) in comparison to achievement of the other units in the observing set (NBA players), that operate in similar circumstances, produce the same outputs consuming the same inputs (homogeneity property). DEA has been used for performance evaluation in wide spread areas in the last 30 years, from non-profit sector like education, power plants and hospitals evaluation (Jeremic et al. 2011a, Jeremic et al. 2012a, Savic et al. 2012, Sueyoshi & Mika 2013) to profit sector like banks, hotels and casinos evaluation (Tsang & Chen 2012, Savic et al. 2013). DEA was introduced by Charnes et al. (1978). In order to make difference among efficient DMUs and allow their ranking, super-efficiency measuring models were used proposed by Andersen & Petersen (1993). Suppose that $\text{DMU}_j (j = 1, \ldots, n)$ uses inputs $x_{ij} (i = 1, \ldots, m)$ to produce outputs $y_{ij} (r = 1, \ldots, s)$. The input-oriented weighted version of Andersen-Petersen's super-efficiency DEA model is the following:

$$\begin{align*}
\text{max} h_k = & \sum_{r=1}^{s} \mu_r y_{rk} \\
\text{s.t.} & \sum_{i=1}^{m} v_i x_{ij} = 1 \\
& \sum_{r=1}^{s} \mu_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, \quad j = 1, \ldots, n, \quad j \neq k \\
& \mu_r \geq s, \quad r = 1, \ldots, s \\
& v_i \geq 0, \quad i = 1, \ldots, m
\end{align*}$$

(1)

The optimal values of efficiency scores $h_k$ are obtained by solving the linear model, $k$-times (once for each DMU in order to compare it with other DMUs). Efficiency score $h_k$ is greater or equal to 1 for all efficient units and smaller than 1 for inefficient units. In this way, ranking of units, according to their efficiency, is enabled. (Ray 2004)

The NBA players’ data consist of eight indicators, from which two are input factors and six are output factors. The inputs, for all players used in the analysis, are gross salary and minutes spend on the court. Outputs used in the analysis are the number of points, the number of assists, the number of rebounds, the number of steals, the number of turnovers and the number of blocked shots which the player made during the regular season 2011/12. All data can be found on (National Basketball Association 2013, ESPN 2013).

Most suitable for the given problem is the input-oriented Andersen-Petersen variable return to scale DEA model, because increasing an input does not result in an identical increase in output. The input-oriented model is used because management can only affect inputs i.e. management of sport team can consider lower gross salary for next year or limit players’ minutes while they cannot affect the number of points scored. One of the most important factors, especially when variable return to scale is used, by Lovell & Rouse (2003), is to control weight restrictions. Measurement of super efficiency in variable return to scale can cause unnatural solutions or even cause that the model has no solution. To avoid that, safety regions of I and II type were used.
Distance based analysis

Quite frequently, the ranking of specific marks is done in such a way that it can seriously affect the process of taking exams, entering competitions, UN participation, medicine selection, and many other areas (Jeremic & Radojicic 2010). I-distance is a metric distance in an n-dimensional space. It was originally proposed and defined by B. Ivanovic, and has appeared in various publications since 1963 (Ivanovic 1977). Ivanovic originally devised this method to rank countries according to their level of development on the basis of several indicators; many socio-economic development indicators had been considered and the problem was how to use all of them in order to calculate a single synthetic indicator which would thereafter represent the rank (Radojicic et al. 2012).

For a selected set of variables \( X_T = (X_1, X_2, ..., X_k) \) chosen to characterize the entities (Jeremic et al. 2011a, Bulajic et al. 2012), the I-distance between the two entities \( e_r = (x_{1r}, x_{2r}, ..., x_{kr}) \) and \( e_s = (x_{1s}, x_{2s}, ..., x_{ks}) \) is defined as

\[
D(r, s) = \sum_{i=1}^{k} \left| \frac{d_i(r,s)}{\sigma_i} \prod_{j=1}^{i-1} (1 - r_{ji,12,i-1}) \right|
\]  

(2)

where \( d_i(r,s) \) is the distance \( d_{ij}(r,s) \) between the values of variable \( X_i \) for \( e_r \) and \( e_s \), e.g. the discriminate effect,

\[
d_i(r,s) = x_{ir} - x_{is}, \quad t \in \{1, ..., k\}
\]  

(3)

\( \delta_i \) the standard deviation of \( X_i \), and \( r_{ji,12,i-1} \) is a coefficient of the partial correlation between \( X_i \) and \( X_{j_i,j<j_i} \). (Jeremic et al. 2012b)

The construction of the I-distance is iterative; it is calculated through the following steps:

- Calculate the value of the discriminate effect of the variable (the most significant variable, that which provides the largest amount of information on the phenomena that are to be ranked)
- Add the value of the discriminate effect of which is not covered by
- Add the value of the discriminate effect of which is not covered by and
- Repeat the procedure for all variables. (Radojicic & Jeremic 2012)

In order to rank the entities, it is necessary to have one entity fixed as a referent in the observing set using the I-distance methodology (Jeremic et al. 2011b). The entity with the minimal value for each indicator or a fictive minimum, maximum, or average value entity may all be utilized as the referent entity, since the ranking of the entities in the set is based on the calculated distance from the referent entity (Jeremic et al. 2011c, Bulajic et al. 2012). In that, the I-distance method shall be applied to several input indicators so as to calculate their I − distance\_input values. The same approach shall be applied to Output indicators and the I − distance\_output values will be calculated for these as well. The obtained values will be brought to a 0-1 level by implementing an \( \text{L}_\infty \) norm. The efficiency of the DMU will be calculated as the:

\[
DBA_{eff} = \frac{I \text{- distance}_{output}}{I \text{- distance}_{input}}
\]  

(4)

Any DMU with an efficiency ratio of at least 1 (DBA ≥ 1) is to be considered as efficient (Jeremic et al. 2012a).

3. Findings and analysis

In order to obtain the following results, software EMS 1.3 (Scheel 2000) and IBM SPSS Statistics for academic purposes is used. For the analysis we have chosen 26 players. All players play at the guard position (shooting guard or point guard) and had good results in the season 2011/12. (National Basketball Association 2013, ESPN 2013)
NBA players ranking

Huge capital was invested in the NBA in the past years. Therefore there was a need to develop ways to measure the effectiveness of players. Until then, there was only one way to measure efficiency (NBA EFF). It is calculated by adding desirable and subtracting undesirable actions during the game.

It is clear that there is no bias toward players’ attributes like minutes on the court and, because of that, it is possible to happen that a player who has played the entire game and has zero efficiency has the same value as a player who has played a few seconds and has shown zero efficiency. Therefore, new methods for measuring the efficiency were developed. One of them is a player efficiency rating (PER). It is invented by John Holinger, a sports journalist. Efficiency is measured by weighting of each “point”, depending on the time and situation, and efficiency obtained in that way is divided by playing minutes. This formula takes on many variables including points, assists, blocked shots, fouls, free throws, shots made, missed shots, rebounds, steals and turnovers to quantify the player performance as regards their pace throughout the game and the average performance level of the league. In this way the management and viewers get a lot more information and can compare different players. Another measure of efficiency was developed because some players “chase” efficiency, while the other players are “supporting” players i.e. they allow other players to score points. In order to catch that phenomenon the plus/minus rating (+/-) was invented. This type of efficiency is evaluated by calculating the number of points the team makes with that player on the field minus the number of points the opposing team made. This calculation is done for each team player while they are on court. (Schumaker et al. 2010)

According to the PER rating, the league average is 15, so every player above that value can be considered efficient. As seen in Table 1, according to PER, only three players are inefficient – Ray Allen, O.J. Mayo and Ben Gordon. The NBA EFF rating does not have a strict line in relation to which we can say whether a player is efficient or not. It can be said only that LeBron James has the highest score while Ben Gordon has the lowest. For plus/minus rating it is simple – if a player has a positive value he is efficient, otherwise he is inefficient. The more plus/minus value he has, the more efficient he is.

Based on the DEA it can be seen that ten players are efficient, while other 14 are not. The player with the highest efficiency score is John Wall with a score of 115.30%. The second place belongs to Russell Westbrook with 114.26%. O.J. Mayo, James Harden and Mo Williams are close to the border of efficiency, while Jason Terry, Kobe Bryant and Joe Johnson are extremely inefficient. According to the DBA, thirteen players are considered efficient because they have the score above 1. Unlike the DEA, the DBA considers O.J Mayo, Mo Williams and Andre Iguodala (who only passed the border of efficiency by 0.014 points) to be efficient. The highest score has been achieved by Derrick Rose – 4.326 and hence, according to DBA, he should be considered the most efficient player.

Table 1: Players ranking

<table>
<thead>
<tr>
<th>Rank</th>
<th>DMU</th>
<th>DEA Score</th>
<th>DBA Score</th>
<th>PER</th>
<th>NBA EFF</th>
<th>+/-</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>John Wall</td>
<td>115.30%</td>
<td>1.131</td>
<td>17.77</td>
<td>18.2</td>
<td>-256</td>
</tr>
<tr>
<td>2.</td>
<td>Russell Westbrook</td>
<td>114.26%</td>
<td>1.187</td>
<td>23</td>
<td>20.5</td>
<td>366</td>
</tr>
<tr>
<td>3.</td>
<td>Dwayne Wade</td>
<td>108.95%</td>
<td>1.445</td>
<td>26.37</td>
<td>22</td>
<td>344</td>
</tr>
<tr>
<td>4.</td>
<td>Derrick Rose</td>
<td>107.23%</td>
<td>4.326</td>
<td>23.1</td>
<td>20.4</td>
<td>289</td>
</tr>
<tr>
<td>5.</td>
<td>Rajon Rondo</td>
<td>104.72%</td>
<td>2.151</td>
<td>17.55</td>
<td>19.3</td>
<td>196</td>
</tr>
<tr>
<td>6.</td>
<td>Ray Allen</td>
<td>104.24%</td>
<td>2.996</td>
<td>14.83</td>
<td>13.5</td>
<td>98</td>
</tr>
<tr>
<td>8.</td>
<td>Kevin Durant</td>
<td>101.78%</td>
<td>1.237</td>
<td>26.26</td>
<td>27.3</td>
<td>369</td>
</tr>
<tr>
<td>9.</td>
<td>LeBron James</td>
<td>101.55%</td>
<td>1.367</td>
<td>30.8</td>
<td>29.9</td>
<td>473</td>
</tr>
<tr>
<td>10.</td>
<td>Chris Paul</td>
<td>100.47%</td>
<td>1.400</td>
<td>27.09</td>
<td>24.5</td>
<td>326</td>
</tr>
<tr>
<td>11.</td>
<td>O.J. Mayo</td>
<td>99.49%</td>
<td>1.708</td>
<td>14.76</td>
<td>10.7</td>
<td>34</td>
</tr>
<tr>
<td>12.</td>
<td>James Harden</td>
<td>98.52%</td>
<td>0.985</td>
<td>21.13</td>
<td>17.5</td>
<td>367</td>
</tr>
</tbody>
</table>
Comparison of DEA and DBA with other methods

It is important for the management that the results obtained by the DEA or the DBA derives values compared with the classical methods of measuring efficiency, which we mentioned above. In order to test that the Spearman correlation was carried out.

Table 2 shows that the DEA and the DBA have a positive correlation compared to the other three methods but that correlation is not significant. Between each other, however, they are significantly highly correlated. On the other hand, rankings obtained by plus-minus rating, PER and classic player efficiency rating show a significantly high correlation between each other. The main reason for such a result is that those three methods do not include the player salary in their way of calculating efficiency.

If those methods are modified, in such a way that salary is included, then those methods can be compared directly with DEA. It is done by dividing annual salary with rating points. That can be interpreted as follows: How much the player earns for each of his rating points. That can be interpreted as follows: The plus/minus rating is excluded from the analysis because the values obtained when dividing salary with plus-minus ratings cannot be interpreted properly. The rankings can be seen in Table 3.
The player who made the highest improvement on scale is Dwayne Wade. Players who have good rank in other scales but by the DEA and the DBA methods are significantly lower are Tyreke Evans and James Harden. It is interesting to see that when these methods include the annual salary in their efficiency measurement, Kobe Bryant also has a very low ranking. After checking the Spearman correlation between the ranks, the following results were obtained:

### Table 4: Spearman correlation between DEA, DBA, modified PER rating and modified efficiency rating

<table>
<thead>
<tr>
<th>Spearman's rho</th>
<th>DEA</th>
<th>DBA</th>
<th>Salary/PER</th>
<th>Salary/EFF NBA</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Correlation Coefficient</strong></td>
<td>.100</td>
<td>.853**</td>
<td>.726**</td>
<td>.743**</td>
</tr>
<tr>
<td><strong>Sig. (2-tailed)</strong></td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
<td>.000</td>
</tr>
</tbody>
</table>

**. Correlation is significant at the 0.01 level (2-tailed).

As seen in Table 4, it can be concluded that the DEA and the DBA are significantly highly correlated with both the PER and efficiency rating when they include salary in their ranking calculation. Correlation between the DEA and the DBA is even higher, and it means that these two methods give us similar results. This tells that the DEA and the DBA can be good methods to measure an NBA player’s efficiency because they include one more dimension which is very important in an economic sense - salary. It can be now concluded that those two methods can change the way of looking at the player’s efficiency.
This paper employs the DEA and the DBA to evaluate the efficiency of NBA players from the economy and management aspects in 2011/2012 regular season. Using the data available from NBA.com and available source of NBA salaries we had an opportunity to examine the efficiency in a way different different from that in which classical efficiency methods do it, performing the ranking by Anderson-Peterson model in the DEA and the DBA by performing distance-based efficiency analysis. In that way salary was captured in calculation of efficiency, which is very important for making decisions on, among other things, hiring, amnesty, play positions and salaries.

Further analysis revealed that the all of NBA players are located in the range of efficiency from 70% to 116%. Ten players exceed the efficiency of 100%. In fact, seven players exceed the efficiency of 100% and show true efficiency, while other three are efficient but they are not a role model for any player. The player with the highest efficiency score is John Wall with the score of 115.30%. The second place is occupied by Russell Westbrook with 114.26%. O.J. Mayo, James Harden and Mo Williams are close to the border of efficiency, while Jason Terry and Joe Johnson are extremely inefficient. An interesting fact is that Kobe Bryant, a famous basketball superstar, is also highly inefficient, compared with other players and their performances and salaries.

It is important to note that both the DEA and the DBA came to very similar results. Correlation between these two rankings was 0.853, which is very high. This means that a decision maker can use only one of these two methods and can say that results obtained by it are very representative.

While all other NBA efficiency methods are focused only on player performance on court, the proposed DEA model includes salary which adds a new dimension into the analysis of NBA players. First, the coach can use this method to estimate whether a player should play more or less based on the performance of other, similar players. Second, the executive board is provided a useful insight into a player’s actual performance and decides whether to increase or decrease his annual salary, or even amnesty his contract.

As an improvement of this model, in further work, we plan to analyze all NBA players based on similarity. By similarity we mean players in the same cluster. After that the DEA and the DBA would be conducted on all clusters. With those results team management would gather much more information on players and team performance. Based on that, their decision making process would be much better. Another possible improvement is to perform the DEA and the DBA on team performance, rather than on a player.

REFERENCES


Received: March 2013.
Accepted: June 2013.
Sandro Radovanovic is a master studies student at the University of Belgrade, Faculty of Organizational Sciences. Graduated at the University of Belgrade, Faculty of Organizational Sciences, in 2012, at the Information systems and technologies study program. Areas of his major interests are business intelligence, data mining, machine learning, operational research and computational statistics.

Milan Radojicic is a master studies student at the University of Belgrade, Faculty of Organizational Sciences. Graduated at the Faculty of Organizational Sciences, in 2012, at the Information systems and technologies study program. In 2012 he enrolled master studies at the Faculty of Organizational Sciences at the study program of Operational Research and Computational Statistics, wishing to develop further in that field. Areas of his interests are data analysis, computational and applied statistics, business analytics, operational research, and game theory.

Veljko Jeremic is an Assistant Professor at the Faculty of Organizational Sciences, University of Belgrade. His major interests are applied statistics, computational statistics, business decision making and simulation methods.

Gordana Savic is teaching and research assistant at the Faculty of Organizational Sciences. She took her PhD in May 2012 in the field of Operational Research. She has published more than 50 research and conference papers. The areas of her research interest are mathematical modeling, real-life problem optimization and efficiency evaluation by Data envelopment analysis.