Towards an objective team efficiency rate in basketball

Gilles Celeux, Valérie Robert

Soumission

Journal de la Société Française de Statistique

Towards an objective team efficiency rate in basketball

Titre: Vers une évaluation objective de la performance d'une équipe en basketball

Gilles Celeux[1] and Valérie Robert[2]

Abstract: Taking profit of the numerous statistics on basketball games, we propose a team efficiency rate well related with the ranking of the teams after the regular season. This "objective" efficiency rate is different from the standard efficiency rate used to assess the player performances. The ability of this objective team efficiency rate to recover the season ranking is illustrated for the French PRO A championship and for the NBA championship. Moreover, analyzing the scores we get for the players with this "objective" efficiency rate lead us to propose a specific objective player efficiency rate to better take into account their performances.

Résumé : Tirant profit des nombreuses statistiques issues des matchs de basketball, nous proposons un score d'évaluation de la performance d'une équipe qui retranscrit bien son classement final. Ce score d'évaluation objective est différent des scores classiques d'évaluation des joueurs. Les qualités de notre score sont illustrées pour le championnat français de PRO A et le championnat américain de NBA. De plus, son application pour les joueurs nous a conduit à proposer un score spécifique pour une meilleure prise en compte des performances individuelles.

Keywords: Efficiency rate, Basketball, Regression, Mean Squared Error, Logistic Model, Kendall rank correlation coefficient
Mots-clés : Évaluation, basket-ball, régression, erreur quadratique moyenne, modèle logistique, coefficient de corrélation des rangs de Kendall
AMS 2000 subject classifications: 62P99

1. Introduction

Basketball is a wonderful sport for statistics. After each game, a box score is made available. This box score provides for each player and each team, quantitative information about 15 variables. Besides, a simple Efficiency rate (EFF) is provided to measure the overall performance of the players and the team per game. More generally, other indicators were proposed such as "PER", "Plus-Minus", "Adjusted Plus-Minus", "Wins produced" or "Wins Score" (Martinez and Martinez, 2011 for a review). In Berri (1999, 2008, 2012), Berri et al. (2006) or Berri and Bradbury (2010), such indicators are detailed and their limits are precised (see also Martinez, 2012). Nevertheless, EFF used in many basketball leagues, as the NBA and the French PRO A leagues, remains simple (an additive formula with integer weights), well-established and quite relevant. As a matter of fact, for most games, the winning team has a greater EFF than the losing team. But the aim of the present paper is to go further and to propose an Objective Team EFFiciency rate (OT-EFF)

[1] Inria Saclay.
E-mail: Gilles.Celeux@inria.fr
[2] Université Paris-Sud.
E-mail: Valerie.Robert@math.u-psud.fr

providing the greater possible agreement with the standing of the teams at the end of the regular season. The paper is organized as follows. In the next section the available variables and the EFF are presented and exemplified for the French PRO A seasons. In Section 3, the OT-EFF is introduced and different ways of deriving it are described. The performance of this OT-EFF are analyzed for the French PRO A league and the regular standings of the NBA league. In Section 4, additional experiments are performed. They show that the OT-EFF performs poorly to evaluate the player performances. Thus, a modification of the OT-EFF is proposed to get an objective player efficiency rate which appears to greatly improve the evaluation of the player performances. Notice that alternative efficiency rates have been proposed. They are more sophisticated since they include additional covariables such as opponent's skill set, position of a player or home advantage... (Page et al., 2005, Page et al., 2013, and Fearnhead and Taylor, 2011). But in the present article, we restrict attention to simple efficiency rates based exclusively on the variables used to construct the EFF rate. Finally a discussion section ends this paper.

2. The box score and the efficiency rate

2.1. The box score

After each basketball game, a box score summarizing the performances of the players and the team with 15 variables is made available in sport newspaper and internet sites. Hereunder is an example of such a box score. We chose to show in Table 1 (resp. Table 2) the French (resp Spain) box score of the semi-final France vs. Spain of the 2013 European championship.

Player	FTA	FTM	$2PA$	$2PM$	$3PA$	$3PM$	OR	DR	BS	BA	AST	ST	TO	PF	PFD
Lauvergne	0	0	1	1	0	0	1	1	0	NA	1	0	0	1	0
Batum	0	0	0	0	4	1	0	0	0	NA	2	1	3	3	2
Diot	4	4	2	0	3	2	0	1	0	NA	1	1	0	4	4
Petro	0	0	3	1	0	0	0	2	0	NA	0	1	0	0	0
Kahudi	0	0	0	0	0	0	0	0	0	NA	1	0	0	1	0
Parker	9	8	17	9	2	2	1	5	0	NA	1	2	5	1	11
Pietrus	0	0	3	1	1	1	3	5	1	NA	0	1	0	3	1
De Colo	0	0	1	0	4	1	0	0	0	NA	2	1	2	2	0
Diaw	2	1	8	2	4	1	1	7	0	NA	3	2	3	5	3
Ajinca	2	1	5	1	0	0	2	4	1	NA	0	0	1	3	3
Gelabale	0	0	3	2	3	1	1	4	0	NA	0	0	1	2	0
Team	17	14	43	17	21	9	10	30	2	NA	11	9	15	25	24

TABLE 1. *French box score of the semi-final France vs. Spain of the 2013 European championship.*

The descriptive variables of this box score are the followings:

- x_1 : free throws attempted (FTA), x_2 : free throws made (FTM)
- x_3 : two points attempted ($2PA$), x_4 : two points made ($2PM$)
- x_5 : three points attempted ($3PA$), x_6 : three points made ($3PM$)

Player	FTA	FTM	$2PA$	$2PM$	$3PA$	$3PM$	OR	DR	BS	BA	AST	ST	TO	PF	PFD
Aguilar	0	0	0	0	1	1	1	1	1	NA	1	1	0	1	0
Fernandez	4	3	5	4	6	2	0	3	2	NA	1	1	1	3	7
Rodriguez	1	1	8	2	5	2	0	6	0	NA	9	0	2	3	4
Rey	0	0	2	1	0	0	0	1	0	NA	0	0	0	0	0
Calderon	2	2	3	1	4	0	1	2	0	NA	1	1	1	4	1
Rubio	0	0	4	1	0	0	0	2	0	NA	1	2	4	2	2
Claver	2	1	2	1	0	0	1	3	0	NA	0	0	1	4	2
Emeterio	0	0	2	1	1	1	0	2	0	NA	0	0	0	1	0
Llull	4	3	1	0	2	1	0	0	0	NA	1	0	1	2	2
Gasol	11	9	9	5	1	0	1	8	3	NA	2	0	6	3	7
Mumbru	0	0	0	0	0	0	0	0	0	NA	0	0	0	1	0
Team	24	19	36	16	22	7	7	29	6	NA	16	5	18	24	25

TABLE 2. *Spain box score of the semi-final France vs. Spain of the 2013 European championship.*

- x_7 : offensive rebounds (OR), x_8 : defensive rebounds (DR)
- x_9 : blocks (BS), x_{10} : blocks against (BA)
- x_{11} : assists (AST)
- x_{12} : steals (ST), x_{13} : turnovers (TO)
- x_{14} : personal fouls (PF), x_{15} : personal fouls drawn (PFD)

Remarks:

- Usually, box scores provide the number of minutes a player has played over the 40 (resp. 48) minutes of the game in the French PRO A (resp. NBA). Obviously this information is important. But we omitted it because it enters in none of the efficiency rates considered here.
- On the contrary, the variables PFD and BA are often omitted in the box scores. As a matter of fact, we were unable to find the variable BA for the France vs. Spain game. Actually these two variables do no enter in the formula of the standard efficiency rate. But in our opinion, they are easily gotten and could be relevant to describe the performances of a team and we include them in our study.

2.2. *The standard efficiency rate*

The standard efficiency rate (EFF) is obtained by the following formula:

$$\begin{aligned} EFF = \quad & Pts + OR + DR + AST + ST + BS \\ - \quad & ((FGA - FGM) + (FTA - FTM) + TO) \end{aligned} \tag{1}$$

where $FGA = 2PA + 3PA$, $FGM = 2PM + 3PM$, and $Pts = 3 \times 3PM + 2 \times 2PM + FTM$.

This EFF has a lot of qualities. It is simple, well-established and relevant. As a matter of fact, for most games, the winning team has a greater EFF than the losing team. Moreover, it is easy to be interpreted since, roughly speaking, EFF provides values comparable to Pts (the number of points). For instance a player with an $EFF \geq 20$ has played a great game. Thus, this EFF is often added at the end of the box scores of a game and it can be regarded as a reference efficiency rate. Notice that EFF does not make use of the variables PF, PFD and BA.

For instance for the game France vs. Spain the EFFs are given in Table 3.

Player	EFF
Lauvergne	5
Batum	0
Diot	10
Petro	3
Kahudi	1
Parker	27
Pietrus	13
De Colo	0
Diaw	8
Ajinca	4
Gelabale	8
France	79

Player	EFF
Aguilar	8
Fernandez	17
Rodriguez	15
Rey	2
Calderon	2
Rubio	0
Claver	2
Emeterio	6
Llull	3
Gasol	20
Mumbru	0
Spain	77

TABLE 3. *France and Spain EFF scores of the semi-final France vs. Spain of the 2013 European championship.*

A few comments are in order from Table 3. The little difference between the teams EFF (France 79, Spain 77) shows that the game was very tight. Actually France won 75-72 after a prolongation. The best EFF has been obtained by Parker (27) who was actually considered as the most valuable player of this game by the sport journalists. Parker outperformed the Spanish big star, Marc Gasol (20) but this guy has a pretty good EFF too.

3. Towards an objective team efficiency rate

Despite the fact that EFF is a nice score to measure the player performances, it does not allow to retrieve the standing of the team at the end of a regular season. As an example that will be considered further in this paper, we compare the EFF and the ranking at the end of the regular season 2012-2013 of PRO A (see Table 4). Ideally, the EFF as a function of the ranking should be decreasing. It is not the case (see Figure 1). It means that EFF is not closely related to the team ranking.

The aim of this paper is thus to propose an alternative efficiency rate providing the best possible agreement with the team ranking. More precisely, we aim to achieve a linear combination of the 15 descriptive variables x_j, $j = 1,\ldots,15$ allowing to retrieve at best the team ranking y got at the end of the regular season. Using a mean squared error criterion, the problem reduces to find the weights $\widehat{\alpha} = (\widehat{\alpha}_j)_{1\leq j\leq 15}$ such that

$$\widehat{\alpha} \in \underset{\alpha}{\operatorname{argmin}} \left\| y - \sum_{j=1}^{15} \alpha_j x_j \right\|_2^2 . \tag{2}$$

Team	FTA	FTM	$2PA$	$2PM$	$3PA$	$3PM$	OR	DR	BS	BA	AST	ST	TO	PF	PFD
Gravelines 1, 21	618	420	1334	697	536	182	297	759	88	92	437	218	329	522	620
Strasbourg 2, 18	537	423	1153	628	593	206	312	752	88	61	509	197	442	604	578
Villeurbanne 3, 18	548	425	1173	624	536	201	251	737	69	61	457	181	406	527	595
Chalon 4, 18	508	386	1121	579	701	249	329	726	80	61	521	204	436	579	578
Roanne 5, 17	554	409	1100	550	596	221	312	734	80	67	465	174	403	621	592
Le Mans 6, 16	552	421	1211	625	548	177	313	660	71	66	438	204	414	603	599
Dijon 7, 15	457	338	1237	639	491	160	261	611	41	89	410	263	373	669	560
Nanterre 8, 15	481	353	1064	584	736	277	279	599	30	49	428	214	398	597	556
Orleans 9, 15	545	394	1066	573	710	272	254	657	44	84	486	247	426	582	575
Cholet 10, 15	501	358	1102	601	680	236	255	699	97	49	463	217	395	583	555
Le Havre 11, 13	517	364	1209	671	550	199	291	745	81	48	506	205	463	615	574
Paris 12, 13	536	402	1357	689	549	190	307	635	67	75	514	235	341	527	547
Limoges 13, 13	574	415	1326	648	442	149	353	724	73	80	432	194	472	705	609
Nancy 14, 12	558	366	1313	656	585	190	356	702	75	89	450	229	422	544	559
Boulazac 15, 11	520	356	1266	670	487	144	249	742	57	93	371	188	400	581	577
Poitiers 16, 10	574	404	1100	557	601	205	303	729	90	66	371	157	476	555	647

TABLE 4. *Box scores of the 16 PRO A teams ordered according to their ranking for regular season 2012/2013. Each team name is followed by its ranking and its number of wins over the 30 games.*

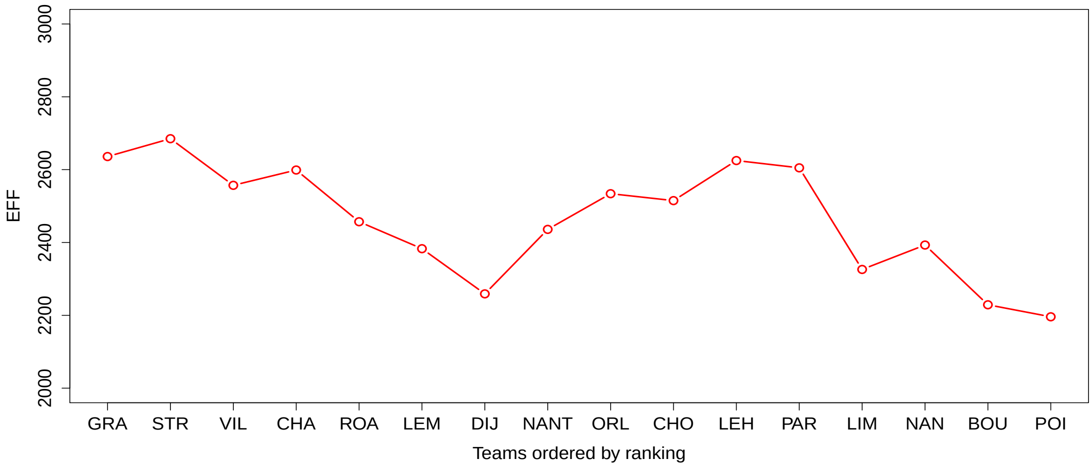

FIGURE 1. *EFF as a function of the teams ranking PRO A 2012/2013.*

Remarks:

- The variable y to be explained may be defined in several ways. It may be the ranking of the teams or the number of wins at the end of the regular season. In any case, it is worthwhile to notice that the data $y_i, i = 1, \ldots, n$, n being the number of teams are not independent. We discuss the choice of y in the next paragraph.
- The intercept in the linear formula is imposed to be zero in order to facilitate the interpretation of the weights α.

- We provide a formula with weights which are increased hundredfold and rounded up or down to the nearest half point to get a more readable formula. But this constraint does not matter when calculating the weights.
- The variable Pts is not included in the study in order to avoid linear combinations between the explanatory variables.

Choosing the response variable y. Choosing y to be equal to the team ranking produces no ties and can be thought of as natural. But this choice can produce unstable results. As it is apparent from Table 4, many teams have the same number of wins and can be considered as tied. For instance, Dijon ranked 7 and Cholet ranked 10 have the same number of wins (15). It does not seem reasonable to differentiate too much these two teams in our study. Moreover, if we choose the team ranking for y, the interpretation of the model coefficients should be counterintuitive if the constant is set to 0. Thus we choose the response variable y to be equal to the number of wins. If there are not many ties, this choice would make little difference with the rankings. But if there are many ties, as it often happens, this choice will produce more stable and reliable results.

When considering the number of wins as the response variable, an alternative logistic model can be considered for choosing the weights α. This logistic model is

$$\log\left[\frac{p}{1-p}\right] = \sum_{j=1}^{15} \alpha_j x_j, \tag{3}$$

where p denotes the win probability of a team.

3.1. Numerical experiments with PRO A data set

We used the models (2) and (3) on the data provided by the LNB [1]. Nine seasons were available from 2004 to 2013. For three seasons, the number of teams was 18 and the number of games for a team was 34. For six seasons, the number of teams was 16 and the number of games for a team was 30. Thus, we get 150 observations to explain the number of wins with a linear equation.

In Table 5, the weights provided by model (2), called $\alpha_{OT\text{-}EFF}$, are compared with the weights of EFF.

	FTA	FTM	$2PA$	$2PM$	$3PA$	$3PM$	OR	DR	BS	BA	AST	ST	TO	PF	PFD
α_{EFF}	-1	2	-1	3	-1	4	1	1	1	0	1	1	-1	0	0
$\alpha_{OT\text{-}EFF}$	-3	2.5	-4	4.5	-4.5	6.5	4	4.5	1.5	-0.5	0.5	6	-6	-2	3.5

TABLE 5. *Comparing the weights of EFF and $OT\text{-}EFF$ for the PRO A league.*

Then, the ability of both criteria to recover the ranking induced by the response variable y is illustrated for the nine seasons 2005-2013 in Figure 2.

The orderings induced by EFF and by $OT\text{-}EFF$ are compared with the Kendall rank correlation coefficient (Kendall (1938)) in Figure 3. The Kendall coefficient is preferred to the Spearman rank correlation coefficient, because it is supposed to be more robust towards ties. That being said,

[1] (www.lnb.fr)

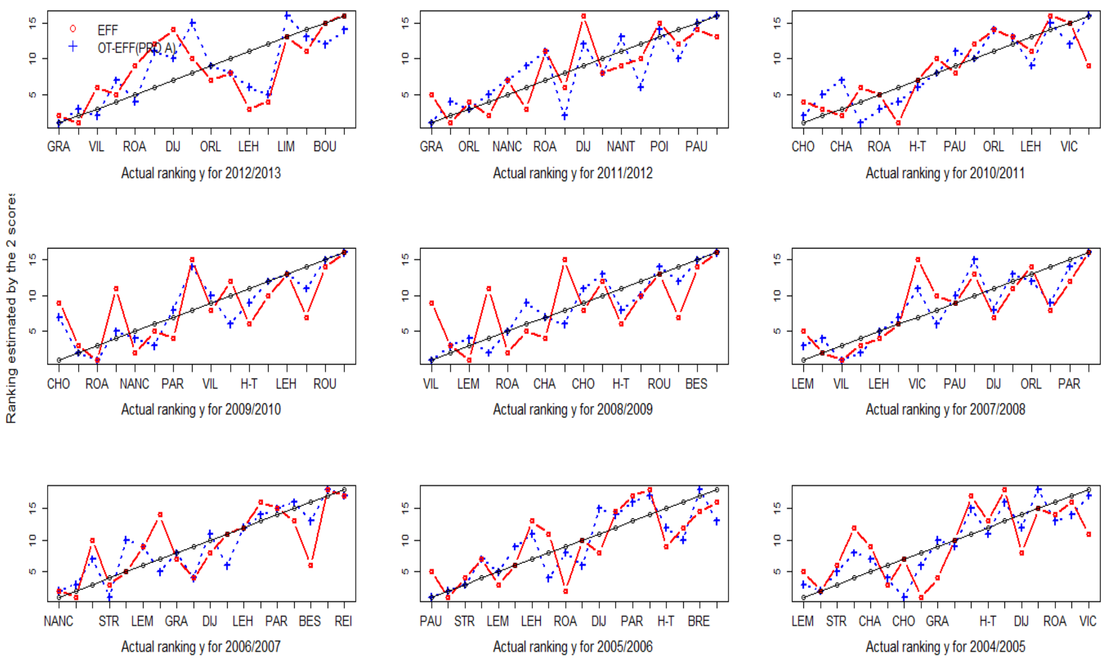

FIGURE 2. *Comparing the ability of EFF (in red) and OT-EFF (in blue) to recover y for the nine PRO A seasons.*

the Spearman rank correlation coefficient provides similar results to those reported here.

Some comments are in order:

- EFF is much simpler than OT-EFF. But, EFF does not take all the variables into account. Moreover, it is unchanging.
- On the contrary, OT-EFF may change over the years and the leagues and it is depending on all the available variables. But it is complex and difficult to read.
- OT-EFF is doing the job for which it has been conceived and is in greater agreement with the response variable y than EFF.
- The interpretation of the OT-EFF weights is of great interest:

  — The turnovers have definitively an important negative impact on the performances of the teams while the steals have a quite positive effect.
  — The assists seem to have no effect on the team performances.
  — The offensive and defensive rebounds on the contrary have an important positive impact, and defensive rebounds have a slightly better positive impact than offensive rebounds, which could be thought of as counter-intuitive.
  — Missed shots have a quite negative impact greater than the impact of successful shots. This is an other big difference with EFF. We comment further this important point in Section 4.

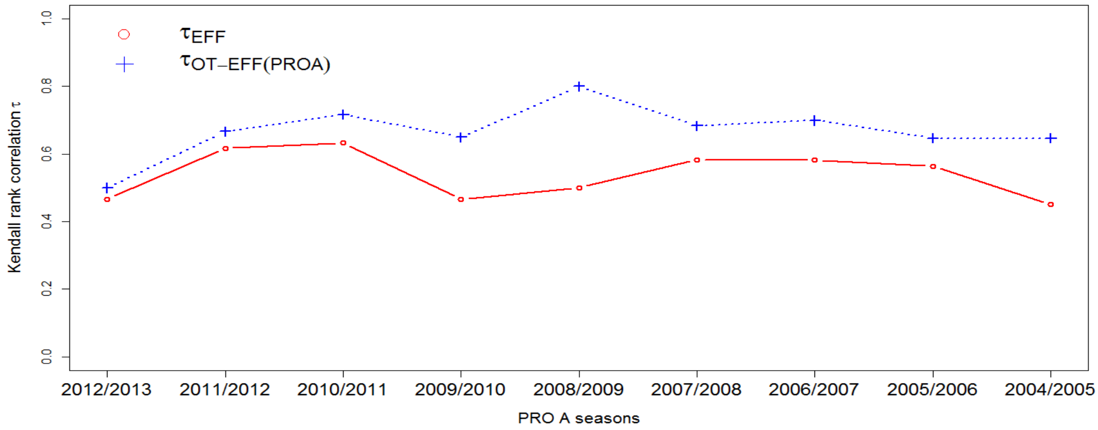

FIGURE 3. *Comparing the Kendall rank correlation coefficients of y with EFF (in red) and y and $OT\text{-}EFF$ (in blue) for the nine PRO A seasons.*

```
Coefficients:
          Estimate Std. Error z value Pr(>|z|)
FTA     -0.0040018  0.0016509  -2.424 0.015349 *
FTM      0.0039827  0.0017647   2.257 0.024020 *
2PA     -0.0062802  0.0007390  -8.498  < 2e-16 ***
2PM      0.0061860  0.0013148   4.705 2.54e-06 ***
3PA     -0.0059823  0.0009091  -6.580 4.70e-11 ***
3PM      0.0081563  0.0023300   3.501 0.000464 ***
OR       0.0060970  0.0009953   6.126 9.02e-10 ***
DR       0.0059368  0.0007842   7.571 3.71e-14 ***
BS       0.0026445  0.0015109   1.750 0.080063 .
BA      -0.0017067  0.0026461  -0.645 0.518947
AST      0.0002767  0.0008100   0.342 0.732648
ST       0.0075942  0.0010676   7.113 1.14e-12 ***
TO      -0.0083443  0.0009264  -9.007  < 2e-16 ***
PF      -0.0006405  0.0006651  -0.963 0.335577
PFD      0.0035990  0.0015565   2.312 0.020761 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

FIGURE 4. *Logistic regression on PRO A data set with the glm function of R.*

Finally, we also experimented the logistic model for the same data using the `glm` function in `R`. The results are summarized in Figure 4 and it is easy to see that the logistic model does not bring new element and the two models highlight the same variables. Therefore, we do not consider the logistic model in the sequel.

3.2. Numerical experiments with NBA data set

We compared on data provided by the NBA [2]. Seven seasons were available from 2006 to 2013. (We discard the season 2011-2012 which has been reduced drastically because of the players' strike.) For each season, the number of teams was 30 and the number of games for a team was 82. Thus, we get 210 observations to explain the number of wins with the $OT\text{-}EFF$ criterion. Table 6 and Figure 5 are analogous to Table 5 and Figure 2.

	FTA	FTM	2PA	2PM	3PA	3PM	OR	DR	BS	BA	AST	ST	TO	PF	PFD
α_{EFF}	-1	2	-1	3	-1	4	1	1	1	0	1	1	-1	0	0
$\alpha_{OT\text{-}EFF(PROA)}$	-3	2.5	-4	4.5	-4.5	6.5	4	4.5	1.5	-0.5	0.5	6	-6	-2	3.5
$\alpha_{OT\text{-}EFF(NBA)}$	-3.5	2	-5	4.5	-6	9.5	6	6	1	-1	1	6	-6.5	0	4

TABLE 6. *Comparing the weights of EFF, $OT\text{-}EFF$ gotten from PRO A league and $OT\text{-}EFF$ gotten from NBA league.*

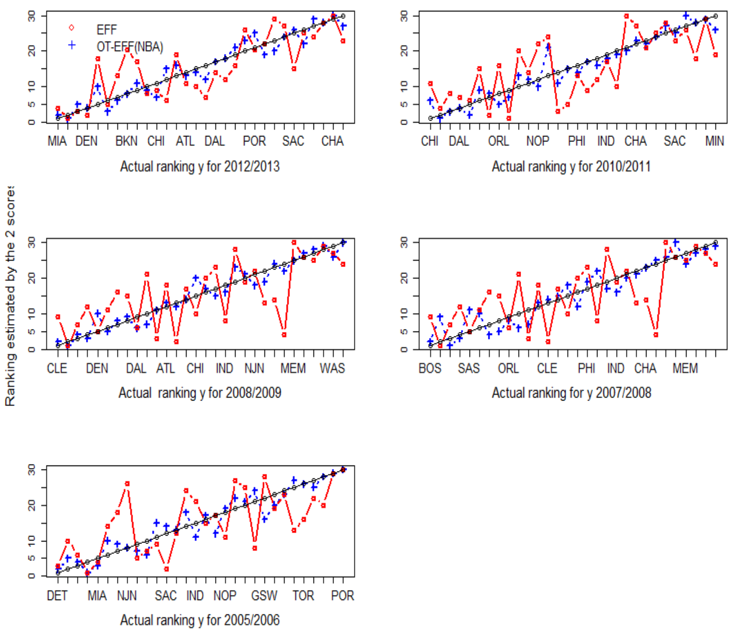

FIGURE 5. *Comparing the ability of EFF (in red) and $OT\text{-}EFF$ (in blue) to recover y for the seven NBA seasons.*

Some comments are in order:

- There is little difference between the $OT\text{-}EFF$ derived from the PRO A and the NBA. Maybe the influence upon three-points field goals is greater for NBA. The fact that the two $OT\text{-}EFF$s are analogous is an information by its own. It means that the way of playing games is not really different for the two leagues. It is one of the interest of $OT\text{-}EFF$ to

[2] www.nba.com

detect possible changes in the way of playing basketball. For instance, it could be of interest to compare $OT\text{-}EFF$s for different periods (for example the fifties and in this day and age).

- Figure 5 shows clearly that the adequacy between y and $OT\text{-}EFF$ is better for the NBA league. There are two reasons for that: (i) The number of observations is greater (210 vs. 150 for the PRO A), (ii) and, above all, the number of games per season is higher (82 vs. 30 for the PRO A) and thus $y =$ "number of wins" is a more precise and sensitive variable.

3.3. *Assessing the performances of* $OT\text{-}EFF$

The $OT\text{-}EFF$s have been computed using the whole data set. Thus, when computing the Kendall rank correlation coefficients of y with $OT\text{-}EFF$, each season is used twice. Each season enters in the computing of $OT\text{-}EFF$ and assessing its ability to achieve a high Kendall rank correlation coefficients with y. Consequently the Kendall rank correlation coefficients shown in Figure 3 could be too optimistic. To get a fair assessing of the $OT\text{-}EFF$ performances, we use a cross-validation procedure (see Hastie et al., 2009, chapter 7 for instance), namely *a leave one season out* procedure: the $OT\text{-}EFF$ is first calculated by discarding the season s, then the Kendall rank correlation coefficient of y with this $OT\text{-}EFF$ is computed on the data from season s. Acting in such a way, we almost eliminate any optimistic bias.

This *leave one season out* procedure has been used for the nine PRO A seasons and the seven NBA seasons analyzed in this article. In this occasion, we add some alternative efficiency rates in this comparison:

- The Euro League efficiency rate, so called *PIR* (Player Index Rating), created in 1991 by the Spanish ACB league which differs from EFF by taking the fouls and the blocks against into account. It gives a 1-weight to the drawn fouls and a (-1)-weight to the committed fouls and the blocks against. For more coherence, this score will be denoted here by $EL\text{-}EFF$.
- The Hollinger efficiency rate, so called Game Score (Hollinger, 2002) but denoted here by $HO\text{-}EFF$. Its formula is the following (Page et al., 2013):

$$\begin{aligned} HO\text{-}EFF = & \quad Pts + 0.4FGM - 0.7FGA - 0.4FTM + 0.7OR + 0.3DR \\ + & \quad ST + 0.7AST + 0.7BS - 0.4PF - TO. \end{aligned}$$

  Remark that this score assumes the weight of free throws attempted to be zero, which could be surprising. Moreover, it tends to overrate poor shooters (Berri, 2012). As a result, Hollinger proposed a more elaborate score called *PER* (Player Efficiency Rate) often used in NBA which is a per-minute and pace-adjusted efficiency rate that we do not study here (Hollinger, 2005, see also Page et al., 2013). Indeed, we aim at competing with EFF which does not consider such covariables as the number of minutes played, for example.
- The Win Score proposed by Berri in 2007 and modified in 2011 including the fact that defensive rebounds count 0.5 times offensives rebounds (this score is available on this web

page [3]). It is denoted here by $WIN\text{-}EFF$ and its expression is as follows:

$$\begin{aligned} WIN\text{-}EFF = \quad & Pts + ST + OR + 0.5DR + 0.5AST + 0.5BS - TO \\ - \quad & FGA - 0.5FTA - 0.5PF. \end{aligned}$$

An important characteristic of this score, which does not consider the variable PFD, is to be almost as simple as EFF as it appears from Table 7.

	FTA	FTM	$2PA$	$2PM$	$3PA$	$3PM$	OR	DR	BS	BA	AST	ST	TO	PF	PFD
$\alpha_{EL\text{-}EFF}$	-1	2	-1	3	-1	4	1	1	1	-1	1	1	-1	-1	1
$\alpha_{HO\text{-}EFF}$	0	0.6	-0.7	2.4	-0.7	3.4	0.7	0.3	0.7	0	0.7	1	-1	-0.4	0
$\alpha_{WIN\text{-}EFF}$	-0.5	1	-1	2	-1	3	1	0.5	0.5	0	0.5	1	-1	-0.5	0

TABLE 7. *Comparing the weights of $EL\text{-}EFF$, $HO\text{-}EFF$ and $WIN\text{-}EFF$.*

- Anticipating a discussion introduced in Section 4, we add a constrained $OT\text{-}EFF$, named $COT\text{-}EFF$, where the absolute values of the weights for FTA and FTM are assumed to be equal, as the absolute values of the weights for $2PA$ and $2PM$ and the variables $3PA$ and $3PM$ are replaced by $3PM = (3PA - \frac{3}{2}3PM)$, a variable which measures the impact of the number of missed three point field goals.

The Kendall rank correlation coefficients of y with all these efficiency rate criteria has been computed for the PRO A (Figure 6) and the NBA (Figure 7) data sets by using *the leave one season out* procedure.

Some comments are in order:

- There is little difference between the standard efficiency rates (EFF, $EL\text{-}EFF$, $HO\text{-}EFF$, and $WIN\text{-}EFF$). Maybe $WIN\text{-}EFF$ appears to be slightly better. But all of them produce smaller Kendall rank correlation coefficients than $OT\text{-}EFF$ in most seasons.
- The performances of $OT\text{-}EFF$ are satisfactory. It means that the optimistic bias is not very high. No surprisingly, the results are dramatically better and more stable for the NBA data set which has more observations and a more precise response variable y.
- Finally, there is no sensitive difference between $OT\text{-}EFF$ and $COT\text{-}EFF$.

4. From team efficiency rate to player efficiency rate

The $OT\text{-}EFF$ criterion has been conceived to propose an efficiency rate the most related possible to the intrinsic value of a team and to give a sensible measure of its ability to win games. Now, it could be interesting to consider $OT\text{-}EFF$ as a criterion to assess the efficiency rate of the players. We performed some numerical experiments not reported here that show that $OT\text{-}EFF$ is not a good rate to measure the individual performances of players. We illustrate its disappointing behavior on the game France vs. Spain of EURO 2013 presented in Section 2. Analyzing its weaknesses, we then propose a simple adaptation of $OT\text{-}EFF$ to get a reasonable efficiency rate for players.

[3] http://wagesofwins.com/2011/12/11/wins-produced-comes-back-better-and-stronger

FIGURE 6. *The* leave one season out *Kendall rank correlation coefficients of different efficiency rates for the PRO A data set: EFF is in red, EL-EFF in pink, HO-EFF in orange, WIN-EFF in purple, OT-EFF in blue, and COT-EFF in sky blue.*

FIGURE 7. *The* leave one season out *Kendall rank correlation coefficients of different efficiency rates for the NBA data set: EFF is in red, EL-EFF in pink, HO-EFF in orange, WIN-EFF in purple, OT-EFF in blue, and COT-EFF in sky blue.*

The EFF and OT-EFF values for the players of the game France vs. Spain are given in Table 8.

The differences between the EFF and OT-EFF scores are important. First, OT-EFF indicates

Player	EFF	OT-EFF
Lauvergne	5	11
Batum	0	-15
Diot	10	9.5
Petro	3	7
Kahudi	1	2
Parker	27	17.5
Pietrus	13	34
De Colo	0	-18
Diaw	8	3
Ajinca	4	7.5
Gelabale	8	2
France	79	60.5

Player	EFF	OT-EFF
Aguilar	8	20
Fernandez	17	19
Rodriguez	15	26
Rey	2	1
Calderon	2	-14.5
Rubio	0	-8.5
Claver	2	3.5
Emeterio	6	5.5
Llull	3	-10
Gasol	20	8
Mumbru	0	-2
Spain	77	48

TABLE 8. *France and Spain EFF and OT-EFF scores of the semi-final France vs. Spain of the 2013 European championship.*

a greater difference between the two teams than EFF. It could make sense but OT-EFF scores of some players are amazing. Parker score (17.5) is outperformed by Pietrus' score (34). It is an interest of OT-EFF to underline the efficiency of Pietrus in this game. But, it is clear that OT-EFF reduces dramatically the efficiency of Parker and this is typically due to the important number of the turnovers of this player. Nevertheless, all the specialists agree to consider that Parker has played a great game and was the most influential player. There is something wrong here. Roughly speaking, OT-EFF favors big men and penalizes small guards.

Thus, OT-EFF is not a reliable player efficiency rate. In order to improve it, we analyze the weights involved in OT-EFF and deduce from this analysis a way to adapt it to the players.

Looking at OT-EFF values gotten from PRO A and NBA, it appears that the weights for FTA and FTM are almost the same, as the weights for $2PA$ and $2PM$, and that the weight of $3PM$ is approximately the weight of $3PA$ multiplied by $3/2$. Notice that assuming equal weights for FTA and FTM is a crude approximation in the OT-$EFF(NBA)$. But, this approximation is natural and will help to get a relevant simplification of OT-EFF:

$$\begin{aligned} OT\text{-}EFF \approx \quad & \alpha_1'(FTA - FTM) + \alpha_3'(2PA - 2PM) + \alpha_5'\left(3PA - \frac{3}{2}3PM\right) \\ & + \alpha_7 DR + \alpha_8 OR + \alpha_9 BS + \alpha_{10} BA + \alpha_{11} AST + \alpha_{12} ST \\ & + \alpha_{13} TO + \alpha_{14} PF + \alpha_{15} PFD. \end{aligned}$$

This approximation of OT-EFF, denoted COT-EFF (Constrained OT-EFF) is depending on 12 variables, namely, LFM, $2PMi$, $3PMi$, DR, OR, BS, BA, AST, ST, TO, PF and PFD where $LFMi = FTA - FTM$ is the number of missed free throws, $2PMi = 2PA - 2PM$ is the number of missed two point field goals and $3PMi = (3PA - \frac{3}{2}3PM)$ measures the impact of the number of missed three point field goals. To get the efficiency rate COT-EFF, we estimate the weights of model (2) from these 12 variables aggregated at the team level instead of the 15 variables considered when computing OT-EFF.

Recall that COT-EFF appears in the Figures 6 and 7 and that these figures show that COT-EFF

and $OT\text{-}EFF$ behave the same.

Notice that the weights α'_1, α'_3, and α'_5 are associated to missed shot and are negative. Looking at $OT\text{-}EFF$ from this view, it could be thought that the players paid an important price when they miss a shot. Thus they have interest avoiding shooting to increase their $OT\text{-}EFF$! Obviously, it does not make sense. It leads us to propose the following modified version of $OT\text{-}EFF$ for each player j. *This modified efficiency rate is based on $COT\text{-}EFF$ and is favoring players whose successful shot percentage is greater than the successful shot percentage of their team.* For a player j, this efficiency rate is as follows:

$$
\begin{aligned}
COT\text{-}EFF(j) = \quad & \alpha_1^{'}(FTA(j) - FTM(j)) + \alpha_3^{'}(2PA(j) - 2PM(j)) \\
& + \alpha_5^{'}\left(3PA(j) - \frac{3}{2}3PM(j)\right) + \alpha_7 DR(j) + \alpha_8 OR(j) \\
& + \alpha_9 BS(j) + \alpha_{10} BA(j) + \alpha_{11} AST(j) + \alpha_{12} ST(j) \\
& + \alpha_{13} TO(j) + \alpha_{14} PF(j) + \alpha_{15} PFD(j) \\
& - \alpha_1^{'}\left(FTM(j) - FTA(j)\frac{FTM}{FTA}\right) \\
& - \alpha_3^{'}\left(2PM(j) - 2PA(j)\frac{2PM}{2PA}\right) \\
& - \alpha_5^{'}\left(3PM(j) - 3PA(j)\frac{3PM}{3PA}\right). \qquad (4)
\end{aligned}
$$

Notice that, as it is desirable, we have $\sum_j COT\text{-}EFF(j) = COT\text{-}EFF$. In Table 9, the EFF, $OT\text{-}EFF$ and $COT\text{-}EFF$ for the players of the game France vs. Spain are given.

Player	EFF	OT-EFF	COT-EFF
Lauvergne	5	11	15
Batum	0	-15	-14
Diot	10	9.5	17
Petro	3	7	6
Kahudi	1	2	4
Parker	27	17.5	29
Pietrus	13	34	35
De Colo	0	-18	-19
Diaw	8	3	-1
Ajinca	4	7.5	3
Gelabale	8	2	4
France	79	60.5	79

Player	EFF	OT-EFF	COT-EFF
Aguilar	8	20	22
Fernandez	17	19	27
Rodriguez	15	26	33
Rey	2	1	1
Calderon	2	-14.5	-13
Rubio	0	-8.5	-10
Claver	2	3.5	5
Emeterio	6	5.5	7
Llull	3	-10	-9
Gasol	20	8	14
Mumbru	0	-2	-1
Spain	77	48	76

TABLE 9. *France and Spain EFF, $OT\text{-}EFF$ and $COT\text{-}EFF$ scores of the semi-final France vs. Spain of the 2013 European championship.*

The differences between $OT\text{-}EFF$ and $COT\text{-}EFF$ are important for the players, but not for the teams. In particular, the great performance of Parker is more fairly acknowledged with $COT\text{-}EFF$. But, the best score remains the Pietrus' score. On the Spanish side, it appears that the performance of Rodriguez is highlighted by $COT\text{-}EFF$, while this score confirms that the performance of

Gasol was below his usual standards. Moreover, it is worthwhile to notice that for the teams, contrary to $OT\text{-}EFF$, $COT\text{-}EFF \approx EFF$.

5. Discussion

We have proposed an efficiency rate $OT\text{-}EFF$ different of the standard efficiency rate as EFF. $OT\text{-}EFF$ is aiming to explain at best the final ranking of the teams in a championship. This team oriented efficiency rate allow to highlight the important negative impact of turnovers and more surprisingly, the negligible impact of assists on the team performances. In summary, $OT\text{-}EFF$ appears to answer the purpose for which it has been conceived quite well. For instance, using the $OT\text{-}EFF$ criterion computed for PRO A, we got for the season 2014 a Kendall τ of 0.75 with the team ranking, while the Kendall τ is 0.52 with EFF and 0.47 with $EL\text{-}EFF$ and $HO\text{-}EFF$, and 0.53 with $WIN\text{-}EFF$. For the 2014 NBA season, the results are analogous: the Kendall τ is 0.79 with $OT\text{-}EFF$, 0.58 with EFF, 0.60 with $EL\text{-}EFF$, 0.53 with $HO\text{-}EFF$ and 0.61 with $WIN\text{-}EFF$. The estimated ranking for PRO A and NBA seasons are displayed in Figures 8 and 9.

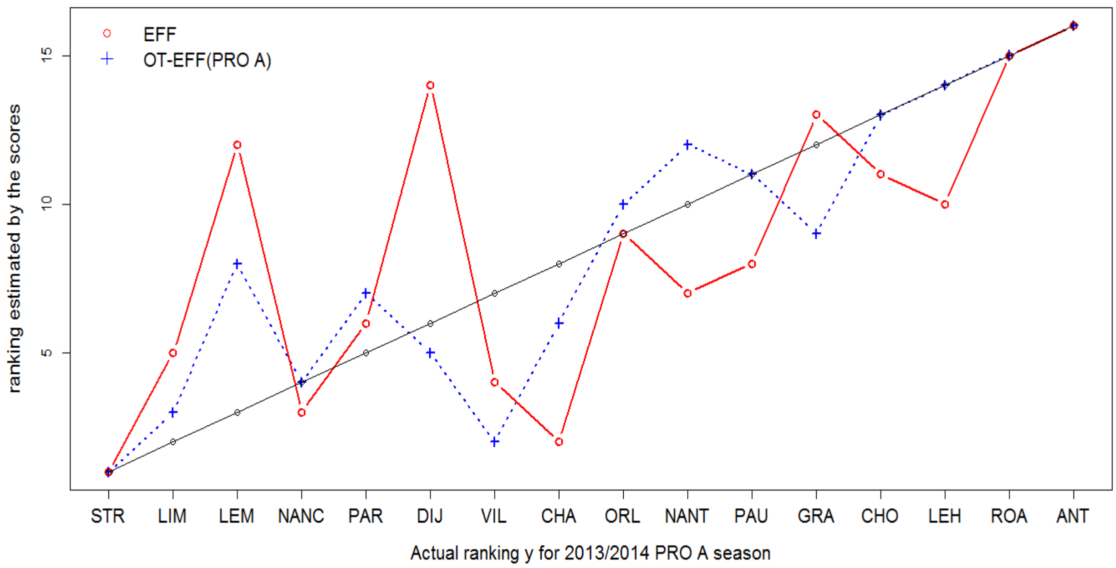

FIGURE 8. *The estimated ranking for the PRO A data set of 2013/2014: EFF is in red, $OT\text{-}EFF$ in blue.*

Furthermore, it is interesting to represent as supplementary points the scores discussed in this article on the first plane of the PCA of the PRO A teams described by the 15 variables in 2014. The graph of teams in Figure 10 shows that the two first axes are not closely related to the ranking. Indeed, neither y nor the scores are well represented on this plane (see the correlation circle on Figure 11). Actually, PCA is not useful to explain y.

Moreover, we have proposed an adaptation of $OT\text{-}EFF$, the so-called $COT\text{-}EFF$ criterion, in order to improve its ability to give a relevant estimation of the players impact. This new criterion rewards or penalizes the player dexterity with respect to the mean dexterity of its team. The differences between $OT\text{-}EFF$ and $COT\text{-}EFF$ could be quite important for some players and these differences are always relevant. These results are encouraging to show the usefulness of

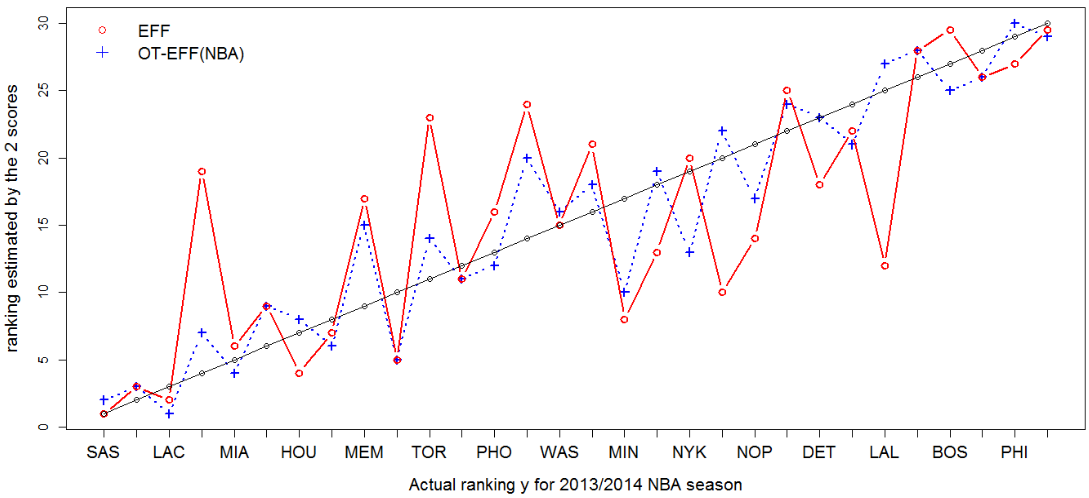

FIGURE 9. *The estimated ranking for the NBA data set of 2013/2014: EFF is in red, OT-EFF in blue.*

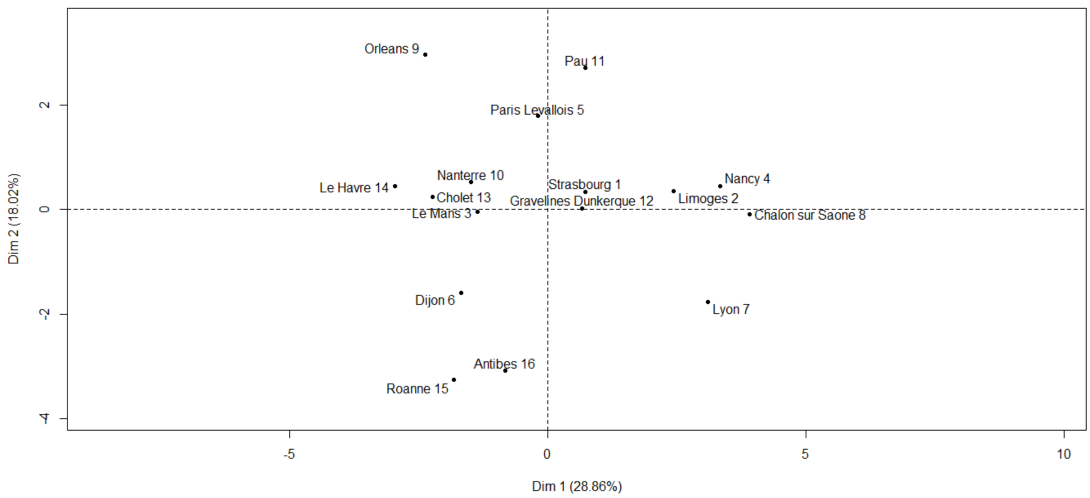

FIGURE 10. *Principal Component Analysis (PCA) of the 2014 PRO A season (graph of individuals).*

the score COT-EFF in assessing the player performances. Further investigations are needed. We test COT-EFF and OT-EFF for the 20 best PRO A players for the 2013 season. Results are not reported here but we see that OT-EFF gave amazing poor scores for brilliant small guards and forwards (see Diabate, King or Diot). This anomaly was corrected by COT-EFF (see Table 10). Besides, the COT-EFF score gives a more realistic ranking to the first class forwards Schlib and Greer which were somewhat underrated by the OT-EFF score.

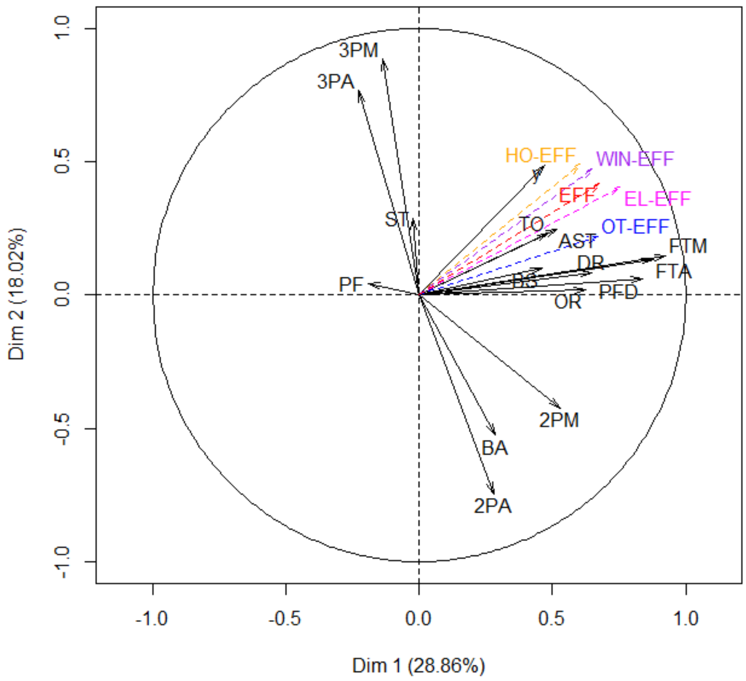

FIGURE 11. *Principal Component Analysis (PCA) of the 2014 PRO A season (graph of variables).*

Player	Position	OT-EFF ranking	COT-EFF ranking
May	C	5	8
Greer	F	6	1
Schilb	F	11	5
Monroe	C	3	3
King	F	20	12
Ajinca	C	10	16
Brockman	C	1	2
Diabate	G	18	9
Nivins	C	7	13
McKenzie	C	9	11
Greene	C	12	10
Williams	C	8	15
Sommerville	C	15	17
Green	C	14	14
Dobbins	G	2	4
Williams	C	17	18
Diot	G	13	6
Buycks	G	19	20
Jackson	G	16	19
Collins	C	4	7

TABLE 10. *Comparison between OT-EFF and COT-EFF rankings for 20 best players of the 2013 PRO A season where G, F, C denote respectively guard, forward and center.*

In conclusion, we do not claim that $OT\text{-}EFF$ should replace EFF or the simple and impressive Win score of Berri. These criteria are definitively reference criteria to measure the efficiency rate of a player. They are simple, easy to interpret and in most general cases relevant. Maybe, it may be suggested to include with a positive weight of one the variable *BA* (block against) in the formula of EFF. But, we think that $OT\text{-}EFF$ and $COT\text{-}EFF$ can bring some additional interesting information. Moreover, $OT\text{-}EFF$ can be thought of as useful to analyze the change in the style of games over the years inside a league or the differences in the style of games between leagues.

References

Berri, D. (2012). David Berri educates us on John Hollinger. (URL: http://www.3sob.com/december-2012) david-berri-educates-us-on-john-hollinger/5515/.

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. *The business of sport*, 3:1–40.

Berri, D. J. and Bradbury, J. C. (2010). Working in the land of the metricians. *Journal of Sports Economics*, 11(1):29–47.

Berri, D. J., Schmidt, M. B., and Brook, S. L. (2006). *The wages of wins: Taking measure of the many myths in modern sport*. Stanford University Press.

Fearnhead, P. and Taylor, B. M. (2011). On estimating the ability of nba players. *Journal of Quantitative Analysis in Sports*, 7(3):11.

Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The elements of statistical learning*, volume 2. Springer.

Hollinger, J. (2002). *Pro Basketball Prospectus.* Brassey's Incorporated.

Hollinger, J. (2005). *Pro Basketball Forecast.* Potomac Books.

Kendall, M. G. (1938). A new measure of rank correlation. *Biometrika*, 30(1–2):81–93.

Martinez, J. A. (2012). Factors determining production (fdp) in basketball. *Economics and Business Letters*, 1(1).

Martinez, J. A. and Martinez, L. (2011). A stakeholder assessment of basketball player evaluation metrics. *Journal of Human Sport & Exercise*, 6(1).

Page, G. L., Barney, B. J., and McGuire, A. T. (2013). Effect of position, usage rate, and per game minutes played on nba player production curves. *Journal of Quantitative Analysis in Sports*, 9(4):337–345.

Page, G. L., Fellingham, G. W., and Reese, C. S. (2005). *Using box-scores to determine a position's contribution to winning basketball games*. PhD thesis, Brigham Young University. Department of Statistics.