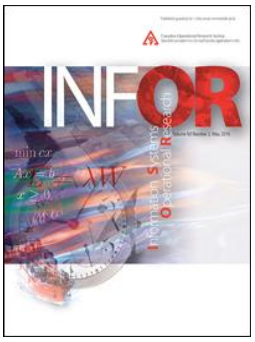

INFOR: Information Systems and Operational Research

ISSN: 0315-5986 (Print) 1916-0615 (Online) Journal homepage: https://www.tandfonline.com/loi/tinf20

Evaluating NBA player performance using bounded integer data envelopment analysis

Ya Chen, Yande Gong & Xiang Li

INFOR: INFORMATION SYSTEMS AND OPERATIONAL RESEARCH, 2017
VOL. 55, NO. 1, 38–51
http://dx.doi.org/10.1080/03155986.2016.1262581

Taylor & Francis
Taylor & Francis Group

Evaluating NBA player performance using bounded integer data envelopment analysis

Ya Chen[a], Yande Gong[b] and Xiang Li[c]

[a]School of Economics, Hefei University of Technology, Hefei, P.R. China; [b]School of Management Science and Engineering, Nanjing Audit University, Nanjing, P.R. China; [c]School of Economics and Trade, Nanjing Audit University, Nanjing, P.R. China

ABSTRACT
Data envelopment analysis (DEA) assumes that data are continuous. However, there are situations where data are integers and bounded. For example, in basketball games, the total number of points that a player has scored is an integer and cannot exceed three times of the number of point field goals that a player has attempted. Without modelling the correct data type, the DEA results can be biased and erroneous. The current paper applies a bounded integer DEA model to evaluating the performance of National Basketball Association (NBA) players when bounded integer data exist. As a result, we correctly capture the data type. The current study also develops a super-efficiency measure under the bounded integer data. The bounded integer data problem is illustrated with data involving a set of NBA shooting guards in the 2013–2014 season.

ARTICLE HISTORY
Received 22 March 2016
Accepted 14 October 2016

KEYWORDS
Data envelopment analysis (DEA); efficiency; performance; bounded; integer; NBA

1. Introduction

Data envelopment analysis (DEA) is a well-known non-parametric approach to measure the relative efficiency of decision-making units (DMUs) that consume multiple inputs and produce multiple outputs. DEA is useful and effective for performance benchmarking and can comprise many performance measures or metrics (Lim & Zhu 2013; Cook et al. 2014). Up to now, it has been successfully applied to many areas such as sports. For example, DEA has been used as a performance evaluation tool for baseball teams, players and team managers (Anderson & Sharp 1997; Sexton & Lewis 2003; Lewis et al. 2007; Volz 2009; Chen & Johnson 2010), soccer or football teams (Barros et al. 2010; González-Gómez & Picazo-Tadeo 2010; Collier et al. 2011), tennis players (Ruiz et al. 2013), formula one constructors teams (Gutiérrez & Lozano 2014), basketball players and teams (Cooper et al. 2009; Cooper, Ramón et al. 2011; Aizemberg et al. 2014; Moreno & Lozano 2014;Yang et al. 2014), etc.

National Basketball Association (NBA) is considered to be the premier men's professional basketball league in the world. It consists of 30 teams (29 in the United States and 1 in Canada). The income of each team depends on its performance in the regular season and playoffs, and its players. There is no doubt that the league operates around

CONTACT Xiang Li ✉ ideallee1976@sina.com

its players. In fact, the performance of players in each team is one of the most important factors for the team's success. In this paper, we examine the performance of NBA players using DEA.

Cooper et al. (2009) evaluate performance of basketball players by selecting non-zero weights and using an output-oriented assurance region model that considers the opinions of experts with regard to the relative value or bounds of the outputs. Cooper, Ramón et al. (2011) extend the multiplier bound approach based on 'model' DMUs for guaranteeing non-zero weights and avoiding large differences in weights in cross-efficiency evaluations. They then apply their proposed approach to the ranking of basketball players. Lee and Worthington (2013) examine the productive efficiency of 62 starting guards by considering undesirable outputs during the 2011/12 NBA season, a period personified by the Linsanity phenomena. Different from Cooper et al. (2009), Cooper, Ramón et al. (2011) and Lee and Worthington (2013), Moreno and Lozano (2014) assess the performance of 30 NBA teams in a two-stage process. They consider the distribution of the budget between first-team players and bench-team players, and compare their slacks-based measure network DEA approach with the black-box conventional DEA. There are also some studies using other multi-attribute decision-making methods such as analytic hierarchy process (AHP) to evaluate sports teams. For example, Sinuany-Stern et al. (2006) predicted the ranking of 11 Israeli basketball teams by using AHP.

Note that all existing studies assume that data are continuous. This is not always true. For example, the total number of points that a player has scored is both an integer and bounded. The value of such a performance measure cannot exceed three times of the number of point field goals that a player has attempted. To deal with such type of non-continuous data, in this paper, we apply a bounded integer DEA model (Chen et al. 2015) for treating bounded integer data, which is developed based on the integer DEA approaches (Lozano & Villa 2006; Kuosmanen & Kazemi Matin 2009; Kazemi Matin & Kuosmanen 2009; Kuosmanen et al. 2016). Different from existing integer or bounded DEA models (Cooper et al. 1999; Cooper, Pastor 2011) and NBA literature above, Chen et al.'s (2015) model considers both bounded and integer values. As a result, we correctly capture the data type.

The rest of the paper is organized as follows. The following section introduces the bounded integer DEA model. An application to the evaluation of NBA players is then provided in Section 3. Section 4 modifies the bounded integer DEA model into a super efficiency DEA model so that the performance of NBA players can be further analysed. The last section concludes.

2. DEA model with bounded integer data

Assume that there are n DMUs. The inputs are $x_{ij}, i = 1, 2, \ldots, m$ and the outputs are $y_{rj}, r = 1, 2, \ldots, s$, where $j = 1, 2, \ldots, n$ represents DMU_j. The DMUs can represent NBA players, for example. Assume that some inputs are in a form of integer data, and denote this set of inputs by $I_{\text{Int}} \subseteq \{1, 2, \ldots, m\}$. We use subscripts i_{Int} (in $x_{i_{\text{Int}}j}$) to denote the inputs in subsets I_{Int}. Next, we assume that some outputs are in a form of integer data, and denote this set of outputs as $O_{\text{Int}} \subseteq \{1, 2, \ldots, s\}$. Denote the set of bounded outputs as $O_{\text{Bnd}} \subseteq \{1, 2, \ldots, s\}$. We use subscripts r_{Bnd} (in$y_{r_{\text{Bnd}}j}$) and r_{Int} (in $y_{r_{\text{Int}}j}$) to denote the outputs in subsets O_{Bnd} and O_{Int}, respectively. Let $L_{r_{\text{Bnd}}}$ and $U_{r_{\text{Bnd}}}$ represent the lower and upper bounds of the bounded outputs, respectively.

Based upon the work of Chen et al. (2015), we use the following model to treat bounded integer outputs when DMU_o is under evaluation.

$$
\begin{aligned}
\max \quad & \frac{1}{s}\sum_{r=1}^{s}\alpha_r \\
s.t. \quad & \sum_{j=1}^{n}\lambda_j x_{ij} \le \tilde{x}_{io}, \quad i=1,\ldots,m \\
& \tilde{x}_{io} \le x_{io}, \quad i=1,\ldots,m \\
& \tilde{x}_{i_{\text{Int}}\,o} \text{ integer}, \quad i_{\text{Int}} \in I_{\text{Int}} \\
& \sum_{j=1}^{n}\lambda_j y_{rj} \ge \tilde{y}_{ro}, \quad r=1,\ldots,s \\
& \tilde{y}_{ro} \ge \alpha_r y_{ro}, \quad r=1,\ldots,s \\
& L_{r_{\text{Bnd}}} \le \tilde{y}_{r_{\text{Bnd}}o} \le U_{r_{\text{Bnd}}}, \quad r_{\text{Bnd}} \in O_{\text{Bnd}} \\
& \tilde{y}_{r_{\text{Int}}o} \text{ integer}, \quad r_{\text{Int}} \in O_{\text{Int}} \\
& \sum_{j=1}^{n}\lambda_j = 1 \\
& \alpha_r \ge 1, \quad r=1,\ldots,s \\
& \lambda_j \ge 0, \quad j=1,\ldots,n
\end{aligned}
\tag{1}
$$

where λ_j and α_r are intensive and efficiency variables, respectively. $\tilde{x}_{io}$, $\tilde{x}_{i_{\text{Int}}o}$, $\tilde{y}_{ro}$, $\tilde{y}_{r_{\text{Bnd}}o}$ and $\tilde{y}_{r_{\text{Int}}o}$ are unknown variables.

Note that model (1) is a non-radial and output-oriented model. In order to treat outputs differently and fully consider slacks, we choose the non-radial model. According to our DEA application, it is appropriate to develop the output-oriented model. Even though model (1) only model bounded integer outputs, non-integer values can also be bounded as shown in Chen et al. (2015). Note also that model (1) is different from standard integer DEA approaches by including the bounded constraints $L_{r_{\text{Bnd}}} \le \tilde{y}_{r_{\text{Bnd}}o} \le U_{r_{\text{Bnd}}}$, $r_{\text{Bnd}} \in O_{\text{Bnd}}$. Given the inputs/outputs used in this paper, we present model (1) under the assumption of variable returns to scale (VRS). It is the same with Lee and Worthington (2013) or Moreno and Lozano (2014). In fact, it seems unreasonable to expect the scoring performance to double if the total number of minutes is doubled.

3. Performance of NBA Players

This section presents an application of model (1) to evaluating the performance of NBA players when some of the performance metrics are in bounded integers.

3.1. Input and output measures

Consistent with prior literature on basketball performance evaluation and considering the purpose of our analysis, we select the input and output measures as shown in Table 1.

Table 1. Input and output variables for efficiency analysis of NBA players.

	Variable	Meaning	Type
Inputs	MIN	Minutes	Real
	TFGA	Total field goals attempted	Integer
Outputs	TPTS	Total points	Integer, bounded
	FTP	Free throws percentage	Real, percentage, bounded
	REB	Rebounds	Integer
	AST	Assists	Integer
	STL	Steals	Integer
	BLK	Blocks	Integer
	ITOV	Inverse of turnovers	Real
	IFL	Inverse of personal fouls	Real

The definition of each variable is also explained. To be specific, the following inputs and outputs are used for our study:

INPUTS:

- MIN: The total number of minutes a player has played.
- TFGA: The total number of point field goals that a player has attempted.

OUTPUTS:

- TPTS: The total number of points that a player has scored.
- FTP: The percentage of free throws that a player has made. It is defined as $\frac{\mathrm{FT}}{\mathrm{FTA}} \times 100\%$, where FT is the total number of free throws (1 point) that a player has successfully made and FTA is the total number of free throws that a player has taken.
- REB: The total number of rebounds that a player has gained.
- AST: The total number of assists that a player has gained.
- STL: The total number of steals that a player has gained.
- BLK: The total number of blocks that a player has gained.
- ITOV: The inverse of the total number of turnovers that a player has gained.
- IFL: The inverse of the total number of personal fouls that a player has gained.

Following Lee and Worthington (2013), we use MIN as an input, i.e. the total number of minutes a player has played. But we use an absolute value of this measure rather than an average (i.e. minutes per game) for the purpose of this paper. And we use general statistical data (in one season) in this paper, which is the same with Moreno and Lozano (2014). Different from Moreno and Lozano (2014), we do not discriminate different types of points gained by each player, i.e. 2-point and 3-point. As a result, we think that it would be much easier to consider outputs as just the total number of points scored. We use a measure 'free throws percentage' as output because a player may have no choice whether to take a shot and get the chance only when an opposite player makes a foul on him. Except for ITOV used in existing NBA literature above, we also consider another undesired output, i.e. the number of personal fouls. Considering that the variable may negatively affect a team outcome, we use the inverse of personal fouls (IFL) as the output variable like ITOV. Other four variables (REB, AST, STL and BLK) are regular used in existing literature such as Lee and Worthington (2013) and Moreno and Lozano (2014). For MIN, FTP, ITOV and IFL, the data type is real and its value is continuous. For other measures, the data type is integer and its value is discrete. For TPTS, it is bounded integer because its value is not larger than three times of TFGA, i.e. $0 \leq \text{TFG} \leq \text{TFGA} * 3$. For FTP, it is bounded because its value is not larger than 100%, i.e. $0 \leq \text{FTP} \leq 100\%$.

3.2. Data

All the data are obtained from the websites (http://espn.go.com and www.nba.com) for the 2013–2014 season. We have selected a sample of 40 shooting guards who have played at least 41 games (half a regular season). The players are selected and ranked by their salaries. The data are shown in Table 2.

3.3. DEA analysis

Using the data of 2013–2014 season, we next calculate the results based on the DEA model (1) in Section 2. We also compare the results with the conventional DEA model.

Table 2. Data of 40 NBA players.

		Inputs		Outputs							
DMU	Player	MIN	TFGA	TPTS	FTP	REB	AST	STL	BLK	ITOV	IFL
1	Dwyane Wade	1775	761	839	0.732558	241	252	79	29	0.006211	0.009434
2	Eric Gordon	2057	817	813	0.784753	165	208	74	12	0.007463	0.008772
3	James Harden	2777	1205	1275	0.866165	344	446	115	29	0.003774	0.00565
4	Andre Iguodala	2040	458	502	0.651515	293	263	95	18	0.01	0.009709
5	Tyreke Evans	2028	897	803	0.770968	341	363	84	21	0.005714	0.006452
6	DeMar DeRozan	3017	1407	1272	0.82381	343	313	86	28	0.005682	0.005076
7	Shawn Marion	2409	708	740	0.784615	497	124	90	37	0.010526	0.008
8	Marcus Thornton	1741	639	613	0.803571	198	77	58	11	0.014925	0.009091
9	O.J. Mayo	1346	550	532	0.863636	124	113	28	13	0.010526	0.009009
10	Manu Ginobili	1550	627	678	0.851064	202	293	70	17	0.007194	0.007813
11	Goran Dragic	2668	1093	1226	0.759615	245	447	104	22	0.004695	0.004854
12	Wesley Matthews	2780	1009	1091	0.837209	289	197	76	14	0.009091	0.005682
13	Kyle Korver	2408	609	763	0.925532	282	208	70	24	0.009804	0.006803
14	Evan Turner	2457	1021	918	0.812977	408	262	67	7	0.005291	0.005128
15	Gerald Henderson	2461	930	846	0.760518	310	199	51	32	0.00885	0.006289
16	J.R. Smith	2421	955	981	0.652174	296	219	65	20	0.009259	0.005102
17	Louis Williams	1445	493	473	0.849162	124	210	45	4	0.01087	0.015385
18	Courtney Lee	1973	614	662	0.883929	187	115	65	28	0.013699	0.007874
19	Jamal Crawford	2094	1011	1003	0.86646	158	223	59	12	0.007407	0.008333
20	Corey Brewer	2609	807	835	0.718062	207	135	151	30	0.009524	0.004762
21	Victor Oladipo	2487	936	858	0.779874	329	327	129	37	0.003906	0.004762
22	Bradley Beal	2530	1149	1100	0.78836	273	243	71	18	0.007813	0.006536
23	Jared Dudley	1729	447	473	0.655172	160	104	41	10	0.017857	0.006849
24	Kirk Hinrich	2116	619	571	0.76	192	286	80	26	0.008333	0.004926
25	Dion Waiters	2072	993	952	0.685106	195	209	63	17	0.006494	0.006536
26	Danny Green	1651	505	568	0.793651	229	104	65	61	0.013158	0.009346
27	Gerald Green	2330	1006	1100	0.847826	275	122	70	42	0.006897	0.004545
28	Ray Allen	1936	543	596	0.905172	205	143	54	8	0.011905	0.008696
29	Vince Carter	1973	811	806	0.821429	284	212	61	35	0.009259	0.004785
30	Randy Foye	2485	875	911	0.848649	232	287	67	39	0.006897	0.004878
31	Ben McLemore	2187	679	605	0.804196	235	82	45	18	0.010417	0.004975
32	Kentavious Caldwell-Pope	1583	460	423	0.770492	156	55	75	12	0.035714	0.006897
33	Wayne Ellington	393	126	135	0.909091	43	19	16	2	0.090909	0.030303
34	Austin Rivers	1339	474	420	0.635838	129	160	45	9	0.012987	0.007463
35	Klay Thompson	2868	1259	1341	0.794595	249	181	74	37	0.007407	0.004274
36	C.J. Miles	984	409	439	0.853333	103	52	46	15	0.022727	0.009709
37	Jordan Crawford	1859	788	744	0.863905	183	281	47	5	0.007463	0.008333
38	Jeremy Lamb	1538	609	614	0.796875	189	115	56	26	0.016393	0.006993
39	Iman Shumpert	1962	484	442	0.746479	308	129	92	13	0.012658	0.004785
40	Jodie Meeks	2556	892	988	0.856589	194	138	111	4	0.009009	0.008403

Note that the conventional DEA model is modified as the non-radial and output-oriented version for the sake of comparability. We do not show the model in this paper.

The results based on the conventional DEA model are reported in Table 3. It shows that 19 out of 40 DMUs are inefficient. The 19 inefficient players are Eric Gordon, Marcus Thornton, O.J. Mayo, Wesley Matthews, Evan Turner, Gerald Henderson, J.R. Smith, Louis Williams, Courtney Lee, Bradley Beal, Jared Dudley, Dion Waiters, Ray Allen, Randy Foye, Ben McLemore, Kentavious Caldwell-Pope, Austin Rivers, Jordan Crawford and Jeremy Lamb. It is interesting that DeMar DeRozan is efficient even though he has two largest inputs (MIN and TFGA). It is also reasonable because he has a very high points per game, i.e. 22.7 in this season. Moreover, Jordan Crawford is the most inefficient player among the 40 players with an efficiency score 0.483. Note that a smaller efficiency score indicates a worse performance and a score of 1 indicates best practice or efficient.

The performance ratings ($\alpha_r, r = 1, \ldots, 8$) and the DMUs' projections ($\tilde{y}_r, r = 1, \ldots, 8$) are shown in 2–9 and 11–18 columns of Table 3, respectively, which is corresponding to the outputs in Table 2. The performance ratings indicate inefficiencies of the outputs and the projections provide an optimal path for performance improvement for inefficient DMUs. For instance, Eric Gordon should increase FTP, REB, AST, STL, BLK, ITOV and IFL by 12.6%, 34.6%, 31.2%, 1.2%, 50.5%, 423.8% and 78.3%, respectively. And the corresponding target quantities are 9.9%, 57.016, 64.953, 0.879, 6.058, 0.032 and 0.007. Eric Gordon should have an increase in ITOV (423.8%), which means that he loses so many balls to the defence that should have been avoided. And he does not need to improve the TPTS. This shows that the player has a good ability to score.

However, as shown in Table 1, one input (TFGA) and five outputs (TPTS, REB, AST, STL and BLK) are integers. Some results in Table 3 for all inefficient DMUs are not correct with respect to their data type. For example, the projections of REB, AST, STL and BLK for Eric Gordon are 222.016, 272.953, 74.879 and 18.058, respectively. All of them should be integers.

We now turn to our model (1). Table 4 shows the results. The same 19 inefficient DMUs are obtained as compared with those in Table 4. Model (1) successfully constrains some of the data to be in integer, namely, the projections of TPTS, REB, AST, STL and BLK are all in integer and correctly calculated now.

The projections shown in Table 4 also provide a detailed path for inefficient DMUs to be efficient. For example, for Marcus Thornton, he should increase TPTS, FTP, AST, STL, BLK, ITOV and IFL by 10.9%, 9%, 176.6%, 12.1%, 90.9%, 191.6% and 87.2%, respectively, which corresponds to the targets 67, 7.2%, 136, 7, 10, 0.029 and 0.008. For Jeremy Lamb, he should increase all the output variables except BLK.

It should be noted that Jeremy Lamb's efficiency score is higher than Eric Gordon's, indicating that Jeremy Lamb has a better overall performance than Eric Gordon. It seems incredible at first glance because Eric Gordon is a fat-salary player but Jeremy Lamb is just a bench-team player. But it is true when we carefully analysis the results. All input measures of Eric Gordon are higher than those of Jeremy Lamb. Moreover, Jeremy Lamb has a better performance than Eric Gordon on FTP (0. 796875/0. 784753), REB (189/165), BLK (26/12) and ITOV (0.016393/0. 007463). In other words, Jeremy Lamb has better ability to grab rebounds, block shots and less total number of turnovers. Thus, all the measures should be considered to fully evaluate a player's performance.

Table 3. Results based on conventional DEA model.

DMU	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	Efficiency[a]	$\tilde{y}_1$ [b]	$\tilde{y}_2$	$\tilde{y}_3$	$\tilde{y}_4$	$\tilde{y}_5$	$\tilde{y}_6$	$\tilde{y}_7$	$\tilde{y}_8$
1	1	1	1	1	1	1	1	1	1	839	0.733	241	252	79	29	0.006	0.009
2	1	1.126	1.346	1.312	1.012	1.505	5.238	1.783	1.790	813	0.884	222.016	272.953	74.879	18.058	0.039	0.016
3	1	1	1	1	1	1	1	1	1	1275	0.866	344	446	115	29	0.004	0.006
4	1	1	1	1	1	1	1	1	1	502	0.652	293	263	95	18	0.010	0.010
5	1	1	1	1	1	1	1	1	1	803	0.771	341	363	84	21	0.006	0.006
6	1	1	1	1	1	1	1	1	1	1272	0.824	343	313	86	28	0.006	0.005
7	1	1	1	1	1	1	1	1	1	740	0.785	497	124	90	37	0.011	0.008
8	1.112	1.087	1	2.758	1.124	2.016	2.847	1.844	1.724	681.828	0.874	198	212.386	65.176	22.173	0.042	0.017
9	1.089	1.030	1.313	1.650	1.997	1	5.329	2.268	1.960	579.453	0.890	162.870	186.460	55.927	13	0.056	0.020
10	1	1	1	1	1	1	1	1	1	678	0.851	202	293	70	17	0.007	0.008
11	1	1	1	1	1	1	1	1	1	1226	0.760	245	447	104	22	0.005	0.005
12	1	1.027	1	1.906	1.288	1.745	1.543	1.465	1.372	1091	0.860	289	375.506	97.882	24.426	0.014	0.008
13	1	1	1	1	1	1	1	1	1	763	0.926	282	208	70	24	0.010	0.007
14	1.020	1.012	1	1	1.445	4.503	2.077	1.545	1.700	936.281	0.823	408	262	96.824	31.522	0.011	0.008
15	1	1.107	1	1.228	1.659	1	2.379	1.720	1.387	846	0.842	310	244.291	84.598	32	0.021	0.011
16	1	1.329	1	1.478	1.410	1.206	2.463	2.172	1.507	981	0.867	296	323.635	91.629	24.118	0.023	0.011
17	1.123	1	1.390	1	1.307	3.421	3.252	1	1.687	531.042	0.849	172.369	210	58.823	13.686	0.035	0.015
18	1.063	1	1.265	1.725	1.038	1	1.852	1.492	1.304	703.928	0.884	236.604	198.352	67.458	28	0.025	0.012
19	1	1	1	1	1	1	1	1	1	1003	0.866	158	223	59	12	0.007	0.008
20	1	1	1	1	1	1	1	1	1	835	0.718	207	135	151	30	0.010	0.005
21	1	1	1	1	1	1	1	1	1	858	0.780	329	327	129	37	0.004	0.005
22	1	1.107	1.091	1.566	1.406	1.381	2.195	1.443	1.399	1100	0.873	297.794	380.452	99.803	24.855	0.017	0.009
23	1.047	1.269	1.153	1	1.333	4.206	1.932	2.192	1.766	495.216	0.831	184.467	104	54.654	42.059	0.034	0.015
24	1	1	1	1	1	1	1	1	1	571	0.76	192	286	80	26	0.008	0.005
25	1	1.245	1.057	1	1.060	1.319	4.556	1.880	1.640	952	0.853	206.089	209	66.782	22.427	0.030	0.012
26	1	1	1	1	1	1	1	1	1	568	0.794	229	104	65	61	0.013	0.009
27	1	1	1	1	1	1	1	1	1	1100	0.848	275	122	70	42	0.007	0.005
28	1.044	1	1	1.268	1.082	2.084	3.362	1.805	1.581	622.158	0.905	205	181.334	58.407	16.671	0.040	0.016
29	1	1	1	1	1	1	1	1	1	806	0.821	284	212	61	35	0.009	0.005
30	1.029	1	1.244	1	1.338	1	1.448	1.580	1.205	937.461	0.849	288.655	287	89.638	39	0.010	0.008
31	1.187	1.087	1	2.703	1.542	1.027	4.025	3.317	1.986	717.982	0.874	235	221.628	69.404	18.482	0.042	0.017
32	1.108	1.002	1.384	3.507	1	1.512	1	2.244	1.595	468.495	0.772	215.865	192.881	75	18.144	0.036	0.015
33	1	1	1	1	1	1	1	1	1	135	0.909	43	19	16	2	0.091	0.030
34	1.198	1.397	1.121	1	1.097	1.215	4.715	2.944	1.836	503.147	0.888	144.668	160	49.345	10.932	0.061	0.022
35	1	1	1	1	1	1	1	1	1	1341	0.795	249	181	74	37	0.007	0.004
36	1	1	1	1	1	1	1	1	1	439	0.853	103	52	46	15	0.023	0.010
37	1.122	1.022	1.244	1	1.633	3.713	5.017	1.821	2.071	834.429	0.883	227.704	281	76.749	18.567	0.037	0.015
38	1.044	1.086	1.028	1.648	1.121	1	2.515	2.360	1.475	640.972	0.866	194.263	189.490	62.749	26	0.041	0.017
39	1	1	1	1	1	1	1	1	1	442	0.746	308	129	92	13	0.013	0.005
40	1	1	1	1	1	1	1	1	1	988	0.857	194	138	111	4	0.009	0.008

[a]The scores are shown as reciprocals to the aggregated values.
[b]($\sim$) indicates DEA projections.

Table 4. Results based on model (1).

DMU	α_1	α_2	α_3	α_4	α_5	α_6	α_7	α_8	Efficiency[a]	$\tilde{y}_1$[b]	$\tilde{y}_2$	$\tilde{y}_3$	$\tilde{y}_4$	$\tilde{y}_5$	$\tilde{y}_6$	$\tilde{y}_7$	$\tilde{y}_8$
1	1	1	1	1	1	1	1	1	1	839	0.733	241	252	79	29	0.006	0.009
2	1	1.126	1.345	1.313	1.014	1.5	5.228	1.781	1.788	813	0.883	222	273	75	18	0.039	0.016
3	1	1	1	1	1	1	1	1	1	1275	0.866	344	446	115	29	0.004	0.006
4	1	1	1	1	1	1	1	1	1	502	0.652	293	263	95	18	0.01	0.010
5	1	1	1	1	1	1	1	1	1	803	0.771	341	363	84	21	0.006	0.006
6	1	1	1	1	1	1	1	1	1	1272	0.824	343	313	86	28	0.006	0.005
7	1	1	1	1	1	1	1	1	1	740	0.785	497	124	90	37	0.011	0.008
8	1.109	1.090	1	2.766	1.121	1.909	2.916	1.872	1.723	680	0.876	198	213	65	21	0.044	0.017
9	1.096	1.032	1.306	1.646	1.964	1	5.349	2.278	1.959	583	0.891	162	186	55	13	0.056	0.021
10	1	1	1	1	1	1	1	1	1	678	0.851	202	293	70	17	0.007	0.008
11	1	1	1	1	1	1	1	1	1	1226	0.760	245	447	104	22	0.005	0.005
12	1	1.024	1	1.904	1.289	1.786	1.498	1.447	1.368	1091	0.857	289	375	98	25	0.014	0.008
13	1	1	1	1	1	1	1	1	1	763	0.926	282	208	70	24	0.010	0.007
14	1.017	1.010	1	1	1.448	4.571	1.970	1.515	1.691	934	0.821	408	262	97	32	0.010	0.008
15	1	1.107	1	1.231	1.667	1	2.367	1.713	1.386	846	0.842	310	245	85	32	0.021	0.011
16	1	1.329	1	1.475	1.4	1.2	2.463	2.172	1.505	981	0.867	296	323	91	24	0.023	0.011
17	1.123	1	1.371	1	1.289	3.25	3.364	1.019	1.677	531	0.849	170	210	58	13	0.037	0.016
18	1.062	1	1.262	1.722	1.031	1	1.855	1.493	1.303	703	0.884	236	198	67	28	0.025	0.012
19	1	1	1	1	1	1	1	1	1	1003	0.866	158	223	59	12	0.007	0.008
20	1	1	1	1	1	1	1	1	1	835	0.718	207	135	151	30	0.010	0.005
21	1	1	1	1	1	1	1	1	1	858	0.780	329	327	129	37	0.004	0.005
22	1	1.106	1.084	1.547	1.394	1.389	2.206	1.442	1.396	1100	0.872	296	376	99	25	0.017	0.009
23	1.042	1.261	1.175	1	1.366	4.3	1.850	2.134	1.766	493	0.826	188	104	56	43	0.033	0.015
24	1	1	1	1	1	1	1	1	1	571	0.76	192	286	80	26	0.008	0.005
25	1	1.246	1.072	1	1.063	1.353	4.507	1.862	1.638	952	0.854	209	209	67	23	0.029	0.012
26	1	1	1	1	1	1	1	1	1	568	0.794	229	104	65	61	0.013	0.009
27	1	1	1	1	1	1	1	1	1	1100	0.848	275	122	70	42	0.007	0.005
28	1.045	1	1	1.266	1.074	2.125	3.322	1.789	1.578	623	0.905	205	181	58	17	0.040	0.016
29	1	1	1	1	1	1	1	1	1	806	0.821	284	212	61	35	0.009	0.005
30	1.029	1	1.241	1	1.328	1	1.448	1.580	1.203	937	0.849	288	287	89	39	0.010	0.008
31	1.183	1.087	1	2.695	1.533	1	4.031	3.321	1.981	716	0.874	235	221	69	18	0.042	0.017
32	1.106	1.002	1.385	3.509	1	1.5	1.003	2.248	1.594	468	0.772	216	193	75	18	0.036	0.016
33	1	1	1	1	1	1	1	1	1	135	0.909	43	19	16	2	0.091	0.030
34	1.198	1.396	1.124	1	1.089	1.222	4.706	2.940	1.834	503	0.888	145	160	49	11	0.061	0.022
35	1	1	1	1	1	1	1	1	1	1341	0.795	249	181	74	37	0.007	0.004
36	1	1	1	1	1	1	1	1	1	439	0.853	103	52	46	15	0.023	0.010
37	1.120	1.019	1.251	1	1.638	3.8	4.906	1.795	2.066	833	0.881	229	281	77	19	0.037	0.015
38	1.039	1.085	1.026	1.643	1.125	1	2.515	2.358	1.474	638	0.865	194	189	63	26	0.041	0.016
39	1	1	1	1	1	1	1	1	1	442	0.746	308	129	92	13	0.013	0.005
40	1	1	1	1	1	1	1	1	1	988	0.857	194	138	111	4	0.009	0.008

[a]The scores reported are reciprocals to the optimal values to model (1).
[b]($\sim$) indicates DEA projection.

Note also that the results based on the regular integer DEA model (non-radial and output-oriented version), which are not shown in this paper, are almost the same with those based on model (1). It implies that both models can be used to calculate the efficiency of DMUs in this example. However, as shown above, model (1) should be used if bounded integer data exist.

4. Performance ranking

Section 3 analyses the performance of NBA players based on model (1). It shows that our results provide very useful and valuable information for differentiating inefficient and efficient players, and setting out a way for performance improvement. However, most players are efficient as shown in Table 4. But it is important to provide a full ranking for the players. The decision-makers can use the ranking to evaluate the players. In order to discriminate the efficient DMUs, a super-efficiency DEA model is developed by Andersen and Petersen (1993) in which a DMU under evaluation is excluded from the reference set. In this section, we further examine the super-efficiency scores and provide a full ranking of the players based on the super-efficiency DEA model.

Before calculating the results, we first extend model (1) to its super-efficiency version.

$$
\begin{aligned}
\max\ & \frac{1}{s}\sum_{r=1}^{s}\alpha_r \\
s.t.\ & \sum_{\substack{j=1 \\ j\neq k}}^{n}\lambda_j x_{ij} \le \tilde{x}_{io},\quad i=1,\ldots,m \\
& \tilde{x}_{io} \le x_{io},\quad i=1,\ldots,m \\
& \tilde{x}_{i_{\text{Int}}\,o}\ \text{integer},\quad i_{\text{Int}} \in I_{Int} \\
& \sum_{\substack{j=1 \\ j\neq k}}^{n}\lambda_j y_{rj} \ge \tilde{y}_{ro},\quad r=1,\ldots,s \\
& \tilde{y}_{ro} \ge \alpha_r y_{ro},\quad r=1,\ldots,s \\
& L_{r_{\text{Bnd}}} \le \tilde{y}_{r_{\text{Bnd}}o} \le U_{r_{\text{Bnd}}},\quad r_{\text{Bnd}} \in O_{\text{Bnd}} \\
& \tilde{y}_{r_{\text{Int}}o}\ \text{integer},\quad r_{\text{Int}} \in O_{\text{Int}} \\
& \sum_{\substack{j=1 \\ j\neq k}}^{n}\lambda_j = 1 \\
& \alpha_r \ge 1,\quad r=1,\ldots,s \\
& \lambda_j \ge 0,\quad j=1,\ldots,n
\end{aligned}
\tag{2}
$$

Model (2) may be infeasible as it is a VRS super-efficiency DEA model. Fortunately, many studies have examined and solved the super-efficiency infeasibility (Seiford & Zhu 1999; Chen 2005; Cook et al. 2009; Chen & Liang 2011; Lee et al. 2011). In this paper, we use the method in Chen and Liang (2011) to calculate efficiency if model (2) is infeasible.

Table 5. Super-efficiency and full ranking of the players.

DMU	Super-efficiency[a]	Ranking on super-efficiency	PIE	Ranking on PIE
1	1.021	11	14.40%	2
2	0.559	36	8.90%	27
3	1.093	3	15.60%	1
4	1.130	2	9.90%	9
5	1.036	8	12.90%	5
6	1.004	20	12.80%	6
7	1.056	5	9.30%	19
8	0.580	34	7.40%	32
9	0.510	38	7.40%	32
10	1.044	6	13.20%	4
11	1.015	14	13.40%	3
12	0.731	24	9.40%	15
13	1.031	10	9.40%	15
14	0.591	33	9.30%	19
15	0.722	25	9.20%	21
16	0.664	28	9.10%	23
17	0.596	32	9.60%	13
18	0.767	23	9.40%	15
19	1.005	19	11.10%	7
20	1.041	7	6.60%	36
21	1.034	9	9.70%	11
22	0.716	26	9.50%	14
23	0.566	35	5.30%	38
24	1.006	18	7.60%	31
25	0.611	31	9.10%	23
26	1.091	4	9.10%	23
27	1.014	15	9.70%	11
28	0.634	29	9.10%	23
29	1.000	21	9.80%	10
30	0.831	22	8.30%	30
31	0.505	39	4.20%	40
32	0.627	30	5.00%	39
33	1.229	1	6.80%	34
34	0.545	37	6.70%	35
35	1.019	12	8.40%	29
36	1.016	13	9.20%	21
37	0.484	40	10.40%	8
38	0.678	27	9.40%	15
39	1.013	16	6.10%	37
40	1.009	17	8.70%	28

[a]The scores reported are reciprocals to the optimal values to model (2).

Table 5 reports the results when model (2) is applied. The second column of Table 5 shows the super-efficiency scores from model (2). The scores reported in Table 5 are reciprocals to the optimal values to model (2). As a result, a larger super-efficiency score indicates a better performance.

The third column of Table 5 shows the full ranking for the players based on super-efficiency scores in Column 2. As shown in the NBA official website (www.nba.com), player impact estimate (PIE) measures a player's overall statistical contribution against the total statistics in games they play in and yields results which are comparable to other advanced statistics (e.g. the Player Efficiency Rating (PER)) using a simple formula. It is a major improvement to the stat 'efficiency' (EFF) rating. The last two columns thus show the PIE and the corresponding ranking ordered by the PIE.

In Table 5, we find that the super-efficiency scores for most efficient players are close to each other with the exception of two players, i.e. Andre Iguodala and Wayne Ellington. These two players have a clearly higher efficiency scores than the others. Andre Iguodala, who is an All-Star player and wins NBA Finals MVP in 2015, is the second efficient player with a super-efficiency of 1.130. Wayne Ellington is the most efficient player with a super-efficiency of 1.229. It is unexpected and surprising. But we can find the answer if we have a look at Wayne Ellington's data of inputs and outputs in Table 2. He has the minimum for MIN and TFGA and the maximum for ITOV and IFL even though Wayne Ellington has the minimum for REB, AST, STL and BLK, which helps Wayne Ellington dominating others successfully.

The ranking based on our proposed super-efficiency model totally differs from that based on the PIE. The Spearman correlation is 0.369 under 5% level of significance (two-tailed). The players in the top positions of the ranking for the super-efficiency approach are even in the bottom positions of the ranking for the PIE approach. It is not surprising because these two rankings are based on different methods. In fact, our proposed approach is based on DEA efficiency analysis. Namely, both inputs and outputs are considered to assess a player's performance while the PIE approach only considers output metrics. Moreover, the measurement for the input and/or output variables is different for the super-efficiency approach and the PIE approach. Of course, the approach itself for calculation is also different. Anyway, we think that both approaches investigate the efficiency of the players from a particular perspective and can complement each other. We believe that both approaches will provide useful information and supports for decision-makers and readers.

5. Conclusions

The current paper applies a bounded integer model to capturing bounded integer data in performance evaluation of NBA players. In the existing basketball literature on performance evaluation, data type of bounded integer is just neglected from analysis. However, it is important to include this factor because the results can be totally different if the bounded integer data are considered. In this paper, we successfully solve this problem.

We apply the bounded integer model to a set of NBA players and compare it with the conventional non-radial and integer DEA models. We further obtain the super-efficiency scores for the players. The full ranking ordered by the super-efficiency scores is totally different from that ordered by the PIE. But both rankings provide useful information for decision support.

Note that the efficiency scores of the conventional DEA approach are very similar to those of the bounded integer DEA approach used in this paper. This may be a special case and the results may be different for different examples/applications (Chen et al. 2015). Nevertheless, it is interesting and deserves for further study on the potential reasons.

We should note that it is difficult and even impossible for a player to increase an output or decrease an input tremendously. So, it may be inappropriate for an inefficient player such as Kentavious Caldwell-Pope to reach a 350.9% AST even though his AST is small. Similarly, it is hard for a player to increase ITOV or decrease the total number of turnovers multiply. It implies that a player should improve his performance by combining with other performance metrics. For example, Kentavious Caldwell-Pope can increase his FTP and TPTS more in total, and further decrease the total number of turnovers to dominate others. Moreover, we can also add an appropriate upper bound for FTP, TPTS and ITOV that a player can achieve when evaluating a player's performance. Hence, future research can include reasonable assurance region or preference information from experts into the performance evaluation for players.

In this paper, the improvement for the outputs does not consider the difference of players' abilities. Each player has an advantage and/or disadvantage in the output measures. The improvement for certain measures may be easy for some players but difficult for other players. For a more realistic and effective consideration, the performance improvement should consider the players' characteristics. However, it is beyond the scope of the current paper to develop a DEA model for this consideration.

Moreover, our application only considers shooting guards in this paper. In fact, the players have been classified into the five groups according to their position: point guard, shooting guard, small forward, power forward and centre. Thus, if we want to examine the performance of all players in different positions, one could run separate DEA analyses on each of the groups or develop new models for non-homogenous DMUs when data are discrete and bounded. Note that Cook et al. (2013) have developed a DEA model for non-homogenous DMUs. For further research, this model can be used or new DEA models can be proposed to investigate the performance of all players with discrete and bounded data.

Acknowledgments

This research was supported by National Natural Science Foundation of China under Grants (Nos. 71271196, 71471053, 71601064), the Foundation for Innovative Research Groups of the National Natural Science Foundation of China (No. 71121061), the Fund for International Cooperation and Exchange of the National Natural Science Foundation of China (No. 71110107024), and the Foundation of Humanities and Social Sciences of the Ministry of Education (No. 13YJC790081). This paper was finished when Ya Chen was visiting Worcester Polytechnic Institute with financial support from the China Scholarship Council.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

National Natural Science Foundation of China [grant number 71271196], [grant number 71471053]; Foundation for Innovative Research Groups of the National Natural Science Foundation of China [grant number 71121061]; International Cooperation and Exchange of the National Natural Science Foundation of China [grant number 71110107024]; Foundation of Humanities and Social Sciences of the Ministry of Education [grant number 13YJC790081]; China Scholarship Council.

References

Aizemberg L, Roboredo MC, Ramos TG, de Mello JCCS, Meza LA, Alves AM. 2014. Measuring the NBA teams' cross-efficiency by DEA game. Am J Oper Res. 4:101–112.

Andersen P, Petersen NC. 1993. A procedure for ranking efficient units in data envelopment analysis. Manag Sci. 39:1261–1264.

Anderson TR, Sharp GP. 1997. A new measure of football batters using DEA. Ann Oper Res. 73:141–155.
Barros CP, Assaf A, Sá-Earp F. 2010. Brazilian football league technical efficiency: a Simar and Wilson approach. J Sports Econ. 11:641–651.
Chen WC, Johnson AL. 2010. The dynamics of performance space of major league baseball pitchers 1871-2006", Ann Oper Res. 181:287–302.
Chen Y. 2005. Measuring super-efficiency in DEA in the presence of infeasibility. Eur J Oper Res. 161:545–551.
Chen Y, Cook WD, Du J, Hu HH, Zhu J. 2015. Bounded and discrete data and Likert scales in data envelopment analysis: application to regional energy efficiency in China. Ann Oper Res. doi: 10.1007/s10479-015-1827-3.
Chen Y, Liang L. 2011. Super-efficiency DEA in the presence of infeasibility: one model approach. Eur J Oper Res. 213:359–360.
Collier T, Johnson AL, Ruggiero J. 2011. Measuring technical efficiency in sports. J Sports Econ. 12:579–598.
Cook WD, Harrison J, Imanirad R, Rouse P, Zhu J. 2013. Data envelopment analysis with nonhomogeneous DMUs. Oper Res. 61:666–676.
Cook WD, Liang L, Zha Y, Zhu J. 2009. A modified super-efficiency DEA model for infeasibility. J Oper Res Soc. 69:276–281.
Cook WD, Tone K, Zhu J. 2014. Data envelopment analysis: prior to choosing a model. Omega. 44:1–4.
Cooper WW, Park KS, Pastor JT. 1999. RAM: a range adjusted measure of inefficiency for use with additive models, and relations to other models and measures in DEA. J Prod Anal. 11:5–42.
Cooper WW, Pastor JT, Borras F, Aparicio J, Pastor D. 2011. BAM: a bounded adjusted measure of efficiency for use with bounded additive models. J Prod Anal. 35:85–94.
Cooper WW, Ramón N, Ruiz JL, Sirvent I. 2011. Avoiding large differences in weights in cross-efficiency evaluations: application to the ranking of basketball players. J Centrum Cathedra. 4:197–215.
Cooper WW, Ruiz JL, Sirvent I. 2009. Selecting non-zero weights to evaluate effectiveness of basketball players with DEA. Eur J Oper Res. 195:563–574.
González-Gómez F, Picazo-Tadeo AJ. 2010. Can we be satisfied with our football team? Evidence from Spanish professional football. J Sports Econ. 11:418–442.
Gutiérrez E, Lozano S. 2014. A DEA approach to performance-based budgeting of formula one constructors. J Sports Econ. 15:180–200.
Kazemi Matin R, Kuosmanen T. 2009. Theory of integer-valued data envelopment analysis under alternative returns to scale axioms. Omega. 37:988–955.
Kuosmanen T, Kazemi Matin R. 2009. Theory of integer-valued data envelopment analysis. Eur J Oper Res. 192:658–667.
Kuosmanen T, Keshvaril A, Kazemi Matin R. 2016. Discrete and integer valued inputs and outputs in data envelopment analysis. In: J Zhu, editor. Data envelopment analysis: a handbook of models and method. New York, NY: Springer. 67–103.
Lee HS, Chu CW, Zhu J. 2011. Super-efficiency DEA in the presence of infeasibility. Eur J Oper Res. 212:141–147.
Lee BL, Worthington AC. 2013. A note on the 'Linsanity' of measuring the relative efficiency of National Basketball Association guards. Appl Econ. 45:4193–4202.
Lewis HF, Sexton TR, Lock KA. 2007. Player salaries, organizational efficiency, and competitiveness in major league baseball. J Sports Econ. 8:266–294.
Lim S, Zhu J. 2013. Incorporating performance measures with target levels in data envelopment analysis. Eur J Oper Res. 230:634–642.
Lozano S, Villa G. 2006. Data envelopment analysis of integer-valued inputs and outputs. Comput Oper Res. 33:3004–3014.
Moreno P, Lozano S. 2014. A network DEA assessment of team efficiency in the NBA. Ann Oper Res. 214:99–124.
Ruiz JL, Pastor D, Pastor JT. 2013. Assessing professional tennis players using data envelopment analysis (DEA). J Sports Econ. 14:276–302.

Seiford LM, Zhu J. 1999. Infeasibility of super-efficiency data envelopment analysis models. INFOR. 37:174–187.
Sexton TR, Lewis HF. 2003. Two-stage DEA: an application to major league baseball. J Prod Anal. 19:227–249.
Sinuany-Stern Z, Israeli Y, Bar-Eli M. 2006. Application of the analytic hierarchy process for the evaluation of basketball teams. Int J Sport Manag Marketing. 3:193–207.
Volz B. 2009. Minority status and managerial survival in major league baseball. J Sports Econ. 10:522–542.
Yang CH, Lin HY, Chen CP. 2014. Measuring the efficiency of NBA teams: additive efficiency decomposition in two-stage DEA. Ann Oper Res. 217:565–589.