European Journal of Sport Science
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tejs20

Cricket team selection using data envelopment analysis

Gholam R. Amin [a] & Sujeet kumar Sharma [a]

[a] Department of Operations Management and Business Statistics, College of Commerce and Economics, Sultan Qaboos University, Muscat, Oman

Published online: 27 Jul 2012.

European Journal of Sport Science, 2014
Vol. 14, No. S1, S369–S376, http://dx.doi.org/10.1080/17461391.2012.705333

ORIGINAL ARTICLE

Cricket team selection using data envelopment analysis

GHOLAM R. AMIN & SUJEET KUMAR SHARMA

Department of Operations Management and Business Statistics, College of Commerce and Economics, Sultan Qaboos University, Muscat, Oman

Abstract
This paper suggests a new method for cricket team selection using data envelopment analysis (DEA). We propose a DEA formulation for evaluation of cricket players in different capabilities using multiple outputs. This evaluation determines efficient and inefficient cricket players and ranks them on the basis of DEA scores. The ranking can be used to choose the required number of players for a cricket team in each cricketing capability. A real dataset, Indian Premier League 4 (IPL 2011), cricket players having various capabilities is used to choose the best cricket team. The proposed method has the advantage of considering multiple factors related to the performance of players in multiple capabilities collected from IPL 4 and aggregates their scores using a linear programming DEA model. This DEA Aggregation gives the scores of players objectively instead of using subjective computations. The proposed DEA method can be used to form a national cricket team from several clubs or a team of top cricketers.

Keywords: *Cricket, data envelopment analysis, Indian Premier League, linear programming, aggregation*

Introduction

Studies addressing different research issues related to various dimensions of the cricketing sport can be found in the literature. The studies are limited in numbers and few of them have addressed some issues of cricket using graphical methods (see Barr, Holdsworth, & Kantor, 2008; Bracewell & Ruggiero, 2009; Gerber & Sharp, 2006; Van Staden, 2009). The batting strategies in limited over cricket are discussed in Preston and Thomas (2000). Cricket is a relatively new and promising research area in comparison with other sports such as baseball, soccer, etc. Cricket is an immensely popular sport in the Indian subcontinent (India, Pakistan, Sri Lanka and Bangladesh) and its popularity sometimes affects other sports. Cricket's newest innovation is 'Twenty20', essentially an evening entertainment also known as T20. Based on the popularity of the Twenty20 format, the International Cricket Council (ICC) organised a Twenty20 World Cup in 2007, which was played in South Africa and won by India. Now, India is being considered as the hub of international cricket in the world. To those not familiar with cricket, Preston and Thomas (2000) have introduced the notion of limited over cricket in detail.

Although, there are various similarities between limited over cricket and Twenty20 cricket but some important differences are pointed out in Bhattacharya, Gill, and Swartz (2011). The focus of this study is on the latest format of the cricket, i.e. Twenty20 format. In this format, there are two teams contesting at the same time. There are 11 players in each team. Each player is expected to perform well in his role as a batsman, a bowler and a fielder. While batting or bowing, each team is allowed to bat or bowl for a maximum of 20 overs. No bowler can bowl more than four overs. In general, batting and bowling capability of cricket teams determine the success along with other capabilities including fielding performance such as wicket keeping, captaincy, home ground, etc. In the literature of cricket about team selection, see Sharp, Brettenny, Gonsalves, Lourens, and Stretch (2011) and Lemmer (2011), it is often assumed that selectors choose a team of 11 players. However, in practice, it is common to

Correspondence: G.R. Amin, Department of Operations Management and Business Statistics, College of Commerce and Economics, Sultan Qaboos University, Muscat 123, Oman. E-mail: gamin@squ.edu.om

choose a team of 15 players in order to provide flexibility to the captain and the coach of the cricket team. Selection of a cricket team under various constraints such as number of batsmen, bowlers, all-rounder and a wicket keeper is a complex task. This task can be made easier using mathematical modelling. Ahmed, Jindal, and Deb (2011) proposed a method of cricket team selection from the perspective of multi-objective genetic optimisation whereas Iyer and Sharda (2009) used a neural network approach in cricket team selection. Furthermore, Sharp et al. (2011) and Lemmer (2011) suggested an integer programming model in cricket team selection.

In this study, we suggest a novel application of data envelopment analysis (DEA) for team selections in cricket. A DEA model, which is a linear programming technique, is used for evaluation of cricket players in different abilities. The proposed method finds efficient and inefficient players for each capability and ranks them on the basis of DEA scores. This provides a way to choose the required number of players for a cricket team. A real dataset of cricket players from Indian Premier League 4 (IPL 2011) in different capabilities is used to choose the best cricket team. To the best of our knowledge this paper is the first attempt using DEA method for cricket team selection, however, DEA has been applied to team sports particularly to baseball, see Hadley and Ruggiero (2006), Lewis, Lock, and Sexton, (2009), and Ruggiero (2010), and football, see Hass (2003).

Considering multiple factors related to the performance of players simultaneously and comparing the players, with different capabilities, based on these outcomes needs an aggregation method. The contribution of this paper is to propose a DEA model, which can be used to find efficiency of cricket players without using subjective opinion. Also, for improving the performance of inefficient players the proposed DEA method can suggest solution(s). One important issue related to contribution of this paper is that the proposed method is the first attempt of using DEA method to form a cricket team from players of different clubs.

Research motivation

In this section, we illustrate how DEA can evaluate cricket players. To start with, we use a two-dimensional example to analyse the performance of cricket players graphically. We consider a real dataset consisting of 10 batsmen, where each player has two batting parameters, $y_1 =$ Average (AVG) and $y_2 =$ Strike Rate (S/R). The data are taken from 2007 Twenty20 World Cup (see Table 1 in Sharp et al., 2011). The average batting is expressed by R/m where, R indicates the number of runs scored and m denotes the number of times that a batsman was out in a tournament. Also, S/R is defined by R/b where R and b denote the number of runs scored and the number of balls faced by a player, respectively. In the next section, we use the statistics of batsmen with more than two parameters. For illustration, we consider only two parameters related to the performance of batsmen. Nevertheless, the same discussion is true for evaluation of cricket players in other capabilities such as bowling, all-rounders and wicketkeepers. Table I gives the numerical values of Average and SR for 10 batsmen chosen from world cup in 2007.

Figure 1 shows the production possibility set (PPS) corresponding to the players given in Table I.

According to the DEA literature, see Cooper, Seiford, and Tone (2006), the above PPS is constructed using three assumptions. The first DEA axiom concludes that all 10 batsmen, observed players, are in the PPS. The second assumption implies that if a batsman, say Gayle, with average (AVG) of 58.5 and SR of 195, belongs to the PPS then any batsman with average y_1 and SR $y_2$is in the PPS where, $y_1 \le 58.5$ and $y_2 \le 195$. This simply indicates the region which is dominated by statistics of Gayle. For instance, Mubarak and Gibbs are two batsmen whose batting abilities, AVG and SR, are dominated by Gayle. Also, the empirical PPS shown in Figure 1 is made using the third DEA axiom which is convexity. This assumption allows that the line segment connecting any of two players in the PPS should be in the PPS. For example, a convex combination of statistics for two players, say Gayle and Taylor, can be presented as follows:

$$0.5\times\text{Gayle} + 0.5 \times \text{Taylor} = 0.5\times\begin{pmatrix}58.5\\195\end{pmatrix}+0.5\times\begin{pmatrix}107\\127.4\end{pmatrix}=\begin{pmatrix}82.75\\161.2\end{pmatrix}$$

The right hand side of the above equation denotes a batsman, even not observed, who had average of 82.75 and S/R of 161.2. Also, according to the second axiom the region dominated by this batsman

Table I. Ten batsmen with two outputs.

Batsmen	Average (AVG)	Strike rate (S/R)
Chris Gayle	58.5	195
Matthew Hayden	88.3	144.8
Rohit Sharma	88	144.3
Ross Taylor	107	127.4
Justin Kemp	86.5	139.5
Jehan Mubarak	52.5	169.4
Craig McMillan	40.8	181.1
Yuvraj Singh	29.6	194.7
Herschelle Gibbs	55	142.9
Misbah Ul-Haq	54.5	139.7

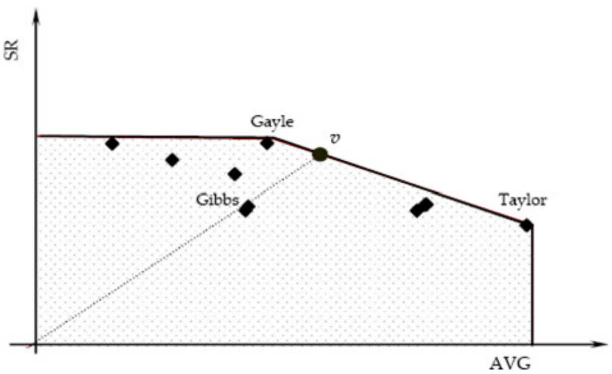

Figure 1. Production possibility set (PPS) for 10 batsmen.

is feasible, that is, any batsman with average of y_1 and S/R of y_2 satisfying the following inequalities:

$$y_1 = \text{Average} \leq 82.75$$
$$y_2 = \text{StrikeRate} \leq 161.2$$

is also in the PPS. In sports, athletes are usually specialised in certain function or task, which might suggest there are economies of specialisation that contradict the convexity assumption, however, cricket experts believe that we should have players in a team who are good in different capabilities. Especially, the convexity assumption is close to the opinion of cricket experts for all-rounders. A good all-rounder is a player who is good in both batting and bowling capabilities.

Mathematically, these three DEA axioms guarantee a batsman with average and SR $(\bar{y}_1, \bar{y}_2)$ is in the PPS if and only if the following constraints are satisfied.

$$\begin{aligned} &58.5\lambda_{\text{Gayle}} + 88.3\lambda_{\text{Hayden}} + \ldots + 55\lambda_{\text{Gibbs}} \\ &\qquad + 54.5\lambda_{\text{Ul-Haq}} \geq \bar{y}_1 \\ &195\lambda_{\text{Gayle}} + 144.8\lambda_{\text{Hayden}} + \ldots + 142.9\lambda_{\text{Gibbs}} \\ &\qquad + 139.7\lambda_{\text{Ul-Haq}} \geq \bar{y}_2 \\ &\lambda_{\text{Gayle}} + \lambda_{\text{Hayden}} + \ldots + \lambda_{\text{Gibbs}} + \lambda_{\text{Ul-Haq}} = 1 \end{aligned} \tag{1}$$

where, $\lambda_j \geq 0$ for all batsmen, $j =$ Gayle, …, Ul-Haq. Now, we give a DEA model for evaluation of 10 batsmen. To see this, assume evaluation of Gibbs. Obviously, this player is not efficient as his batting abilities, average and S/R, is dominated by a convex combination of batting abilities of Gayle and Taylor. Therefore, his performance can be improved if the vector of his batting abilities is projected to the frontier as this is shown in Figure 1 by v. Therefore, the corresponding DEA model for evaluation of Gibbs can be expressed as the following form:

$$\begin{aligned} &\max\ h \\ &s.\ t. \\ &(55, 142.9) \times h \in \text{PPS} \end{aligned} \tag{2}$$

where, $(55,142.9) \times h$ is used to project the vector of batting abilities of Gibbs. We use constraints (1) to simplify model (2) as the following linear programming problem.

$$\begin{aligned} &\max\ h \\ &s.t. \\ &58.5\lambda_{\text{Gayle}} + 88.3\lambda_{\text{Hayden}} + \ldots + 55\lambda_{\text{Gibbs}} \\ &\qquad + 54.5\lambda_{\text{Ul-Haq}} \geq 55\,h \\ &195\lambda_{\text{Gayle}} + 144.8\lambda_{\text{Hayden}} + \ldots + 142.9\lambda_{\text{Gibbs}} \\ &\qquad + 139.7\lambda_{\text{Ul-Haq}} \geq 142.9\,h \\ &\lambda_{\text{Gayle}} + \lambda_{\text{Hayden}} + \ldots + \lambda_{\text{Gibbs}} + \lambda_{\text{Ul-Haq}} = 1 \\ &\lambda_j \geq 0, j = \text{Gayle}, \ldots, \text{Ul} - \text{Haq} \end{aligned} \tag{3}$$

The above linear programming DEA model simply projects the batting abilities of Gibbs to the efficiency frontier (see Figure 1). This means that DEA model (3) is looking for a convex combination of all batsmen to dominate the statistics of Gibbs, if there is any. DEA model (3) is a linear programming problem and can be solved using LP solvers in many popular software packages including spreadsheets, like Excel. An optimal solution of DEA model (3) corresponding to the performance of Gibbs is obtained as follows:

$$\lambda^*_{\text{Gayle}} = 0.778,\ \lambda^*_{\text{Taylor}} = 0.222,\ \lambda^*_j = 0 \forall j,$$
$$j \neq \text{Gayle, Taylor},\ h^* = 1.259512$$

Using the above optimal solution, the batting parameters corresponding to Gibbs will be on the frontier if we multiply his data by $h^\star = 1.259512$. This gives the target for his average and SR as follows:

$$1.259512 \times \begin{pmatrix} 55 \\ 142.9 \end{pmatrix} = \begin{pmatrix} 69.27316 \\ 179.9842648 \end{pmatrix}$$

In this case, the efficiency score of Gibbs can be obtained by

$$\frac{\text{Actual output}}{\text{Target output}} = \frac{\text{Actual average}}{\text{Target average}} = \frac{\text{Actual S/R}}{\text{Target S/R}}$$
$$= \frac{69.27316}{55} = \frac{179.9842648}{142.9} = 0.794$$

where, actual output means the observed batting ability and target output indicates batting ability on the frontier. Therefore, the efficiency score of Gibbs is $\frac{1}{h^*}$, where, $h^\star$ is the optimal value of DEA model (3). Also, positive values of Lambda in the optimal solution denote that the batting abilities of Gibbs are presented in terms of two efficient batsmen, Gayle and Taylor. This is shown in Figure 1 where, the target of batting abilities for Gibbs is given by a convex combination of the batting parameters of Gayle and Taylor. The projection of batting parameters of Gibbs to the frontier is a way to improve the performance of this batsman. Obviously, he will be an efficient batsman if he improves his batting parameters to his target. This means that his average and SR should be increased to $h^\star \times 55 = 1.259512 \times 55 \approx 69.27$ and $h^\star \times 142.9 = 1.259512 \times 142.9 \approx 180$, respectively. If we solve DEA model (3) for the other batsmen we have eight inefficient players and two efficient batsmen, Gayle and Taylor. In the following section, we give a general DEA method for evaluation of cricket players in different abilities.

Evaluation of cricket players using DEA

However, a general DEA model supports multiple inputs and multiple outputs; a wide range of successful DEA applications with multiple outputs and no input or multiple inputs and no output are reported in the literature. These cases are introduced in the areas such as group decision-making, Wang and Chin (2009), preference voting system, Llamazares and Peña (2009), information retrieval and metasearch aggregation, Amin and Emrouznejad (2010) and data mining, Amin, Gattoufi, and Rezaee Seraji (2011). For instance, Amin and Emrouznejad (2010) originated a minimax DEA model with multiple outputs for aggregation of multiple search engines results. In this paper, we introduce a new DEA application for measuring the performance of cricket players in different abilities. This is followed by ranking the players from the most to the least score and selecting a cricket team.

Assume for a given cricket ability, we have n players and for the j^{th} player ($j = 1,\ldots, n$), we consider s outputs $\boldsymbol{y}_j = (y_{1j}, \ldots, y_{sj})$. This is measured in the previously played games by the j^{th} player. A general output-oriented DEA model for evaluation of the k^{th} player ($k = 1, \ldots, n$) can be presented as follows:

$$\begin{aligned} &\max\ h \\ &s.t. \\ &\sum_{j=1}^{n} y_{rj}\lambda_j - y_{rk}h \geq 0 \quad r = 1,\ldots,s \\ &\sum_{j=1}^{n} \lambda_j = 1 \\ &\lambda_j \geq 0 \ \ j = 1,\ldots,n \end{aligned} \tag{4}$$

Notice that model (4) is general form of the output-oriented DEA model illustrated in the previous section. To find out the DEA score of cricket players, we need to solve this model n times, one model for each player. Assume $(\lambda_1^*, \ldots, \lambda_k^*, \ldots, \lambda_n^*, h^*)$ to be an optimal solution of DEA model (4). Clearly, if the optimal value of model (4) is one, i.e. $h^\star = 1$, then the underlying cricket player is efficient. For inefficient player(s) the optimal value of DEA model (4) is strictly greater than one, $h^\star > 1$, which means that the k^{th} player is far from the efficiency frontier. For the later case the performance of the k^{th} player will improve if his outputs are increased to $\hat{\boldsymbol{y}}_k = (\hat{y}_{1k}, \ldots, \hat{y}_{sk}) = \bar{h} \times (y_{1k}, \ldots, y_{sk})$, where $1 < \bar{h} \leq h^*$, where, $\bar{h} = h^*$ indicates a case when the vector of outputs of the k^{th} player is projected to the efficiency frontier. One advantage of using DEA model (4) is related to the interpretation of its dual model. According to duality of linear programming, see Bazarra, Jarvis, and Sherali (2010), DEA model (3) has a dual model. The dual model is a minimisation problem. There is a dual variable corresponding to each constraint of the primal model (3). A pair of primal and dual linear programming can be written as the following models in general form (see Bazarra et al., 2010).

$$\begin{aligned} &\text{Primal model}: \\ &\max \sum_{i=1}^{m} b_i w_i \\ &s.t. \\ &\sum_{i=1}^{m} a_{ij} w_i \leq c_j \qquad j = 1,\ldots,n \\ &w_i \geq 0 \qquad i = 1,\ldots,n \end{aligned}$$

$$\begin{aligned} &\text{Dual model}: \\ &\min \sum_{j=1}^{n} c_j x_j \\ &s.t. \\ &\sum_{j=1}^{n} a_{ij} x_j \geq b_i \qquad i = 1,\ldots,m \\ &x_j \geq 0 \qquad j = 1,\ldots,n \end{aligned}$$

where, $(w_1, \ldots, w_m)$ is vector of decision variables for primal model and $(x_1, \ldots, x_{\text{n}})$ denotes the vector of variables for dual model, where x_j is dual variable corresponding to the j^{th} constraint ($j = 1, \ldots, n$) of primal model. According to the above definition, the dual of DEA model (4) can be written as the following linear programming model.

$$\begin{aligned} &\min u_0 \\ &s.t. \\ &-\sum_{r=1}^{s} y_{rj} u_r + u_0 \geq 0 \qquad j = 1,\ldots,n \\ &\sum_{r=1}^{s} y_{rk} u_r = 1 \\ &u_r \geq 0 \qquad r = 1,\ldots,s \end{aligned} \tag{5}$$

where, u_r is the dual variable corresponding to the r^{th} constraint of primal model (4), ($r = 1, \ldots, s$), and u_0 is the dual variable related to the last constraint of model (4). Also, the last constraint of dual model (5) is an equality because the corresponding variable in primal model (4), that is, *h*, is a free variable. According to duality of linear programming, see Bazarra et al. (2010), the dual variables (u_1, …, u_s) can be interpreted as the importance weights of the cricket's parameters. If there is priority between different parameters of cricket abilities then further constraints should be imposed to the dual model (5). For example, in evaluation of batsmen using dual model (5), the number of sixes by a batsmen is not as important as the batting S/R. Therefore, we have to impose a constraint like $u_1 – u_2 \geq \varepsilon$ where, u_1 is the importance weight defined for strike rake and u_2 is the weight assigned to the number of sixes, and ε is a positive number.

Data for IPL 4

We have taken data from IPL session 4 in 2011. In this section, a brief outline of IPL 4 is given, which is followed by a discussion of various player's capabilities and the use of the proposed DEA model to rank the players. IPL4-T20 was concluded in the months of April–May 2011 and was a huge economic and entertainment forum. Spectators on the field and viewers of the television were entertained to the fullest. In this session of IPL, there were a total of 10 teams on the names of famous cities of India such as Chennai Super Kings, Royal Challengers of Bangalore, etc. These teams chose players via an auction where most players are chosen from the list of players playing in India in different formats of the game whereas some players are chosen from different cricket playing countries. In this session, the player's auction for the IPL 4 was held on 8 and 9 January 2011 at the ITC Royal Gardenia, Bangalore, India. The auction comprised 350 players. Each player in the auction pool had a bidding base price under which franchise owners could not bid. Players were allowed to set their base price between \$200,000 and \$400,000. The maximum number of foreign players can be taken into a team was four.

Batting statistics

In this study, we consider five important measures of batting statistics such as highest individual score (HS), average batting performance, strike rate (S/R), numbers of fours (4s) and number of sixes (6s). These parameters have been taken from IPL 4.

These factors are DEA outputs for each batsman and can be defined as follows:

HS: For a batsman, the highest individual score (HS) is the maximum number of runs scored in one match during a tournament.

Average: The average batting performance is expressed by R/*m* where R denotes the number of runs scored and *m* the number of times the batsman was out.

S/R: The batting strike rate can be expressed as R/*b* where R denotes the number of runs scored and *b* denotes the number of balls faced by a player.

4s: The number of 4s hit by a batsman during IPL4 (2011).

6s: The number of 6s hit by a batsman during IPL4 (2011).

We can say that the higher the value for the above five items the better the performance of the batsman. Therefore, these outcomes of batsmen define five outputs for the DEA method. These are all important measures of batsmen in Twenty20 cricket. Sharp et al. (2011) discussed two performance measures named average and strike rate. The purpose of five measures, in this paper, is to consider all possible performance measures simultaneously which is possible through the proposed DEA method. Lemmer (2011) proposed a new approach to standardise average of batsmen in limited over cricket where the probability of a player remaining not out is relatively more in comparison to twenty20 cricket. In our data, we have taken 60 batsmen from IPL 4 and use simple average as one of the outputs in DEA model, however, there is no limitation in replacing this average with any modified version, unlike the model proposed by Lemmer (2011), depending on the format of the cricket. We use data for 60 batsmen from IPL 4 and the data related to other cricket abilities are not shown in this study. Because of space limitation we just show the top 30 batsmen ranked by DEA model in Table II.

Table II shows the DEA scores, obtained by model (4), for the top 30 batsmen. Sehwag, Gayle and Valthaty are the top three batsmen based on DEA scores followed by Michael Hussey, Tirumalsetti Suman and Subramaniam Badrinath. Fans and cricket experts also believe that these were the top batsmen in IPL 4. For inefficient batsmen, the proposed DEA model can be used to improve their performance. For instance, Jean-Paul Duminy has ranked in the 25th place with efficiency score of 0.64554, that is $h^\star = 1.549$. If he plans to improve his efficiency to 0.75, the target for $\bar{h}$ is 1.33333. According to the explanation he will reach to this target if and only if his batting capabilities change as follows:

$$\begin{aligned}\hat{y}_{\text{Diminy}} &= \left(\hat{y}_{1\text{Diminy}}, \ldots, \hat{y}_{5\text{Diminy}}\right)\\ &= \bar{h} \times \left(y_{1\text{Diminy}}, \ldots, y_{5\text{Diminy}}\right)\\ &= (73.33, 30.85, 168.74, 10.67, 8)\end{aligned}$$

Table II. Batting statistics and DEA scores for top 30 batsmen in IPL 4.

Player names	Highest individual score	Average	Strike rate (SR)	4s	6s	DEA score
Virender Sehwag	119	38.55	176.67	51	18	1
Chris Gayle	107	99.5	196.06	38	30	1
Paul Valthaty	120	37.6	141.89	46	16	1
Michael Hussey	83	49.12	122.43	48	3	0.98746
Tirumalsetti Suman	36	20	190.48	3	3	0.97153
Subramaniam Badrinath	71	87.25	133.72	33	8	0.87689
Sachin Tendulkar	100	46.38	111.75	42	5	0.87619
Virat Kohli	71	55.29	122.08	40	12	0.87458
Jesse Ryder	60	24.82	154.24	36	10	0.82271
Shaun Marsh	95	36.6	141.31	37	12	0.80684
Rahul Dravid	66	29.9	107.55	39	2	0.76622
Rohit Sharma	87	47.83	130.45	24	12	0.75982
Aiden Blizzard	37	19.33	148.72	12	0	0.75855
Gautam Gambhir	75	38.12	123.48	35	2	0.72828
Brendon McCullum	81	26.82	130.53	31	12	0.70852
Sunny Singh	20	10.75	138.71	6	1	0.70746
Sunny Sohal	62	24.38	138.3	23	6	0.70542
Manoj Tiwary	61	69.25	111.69	19	12	0.69599
Tillakaratne Dilshan	59	28.25	125.56	33	3	0.69156
Dinesh Karthik	69	27.11	130.48	27	7	0.67558
Mahela Jayawardene	76	24.42	115.35	34	5	0.66667
Albie Morkel	30	22.5	130.43	5	5	0.66525
David Warner	77	25.09	116.46	30	7	0.65287
Murali Vijay	74	26	119.08	25	14	0.64683
Jean-Paul Duminy	55	23.14	126.56	8	6	0.64554
Venugopal Rao	60	29.09	119.85	29	12	0.64537
Ajinkya Rahane	52	22.4	125.84	11	2	0.64185
Travis Birt	27	14.75	125.53	7	1	0.64025
Eoin Morgan	66	17.38	124.11	18	3	0.63303
Bharat Chipli	61	27.43	123.87	19	4	0.63179

Thus, Jean-Paul Duminy can improve his performance to 0.75 if his HS increased by 73.33–55 = 18.33, his Average by 7.71, S/R by 42.18, and he has to score 3 and 2 more 4s and 6s, respectively.

Bowling statistics

Sharp et al. (2011) and Lemmer (2011) used three bowling measures such as bowler's economy rate, bowling average and bowling strike rate. In addition to these measures, we have incorporated one more measure, numbers of wickets taken to enhance the flexibility. Bowler's economy rate can be expressed by TR/O where TR is the total number of runs conceded by a bowler and O is the total number of overs bowled by a bowler. Bowling average can be expressed as TR/W where TR is the total runs conceded by a bowler and W is the total number of wickets taken by a bowler. Bowling strike rate can be expressed as TB/W where TB is the total number of balls bowled by a bowler and W is the total number of wickets taken by a bowler. Average, strike rate and economy of a bowler are considered better if these parameters are less so as the DEA output we are using the inverse of these values. Using the proposed DEA model (4) for 60 bowlers in IPL 4 we obtain the highest ranked players shown in Table III.

The top three efficient bowlers are Malinga, Sharma and Marsh followed by bowlers with decreasing efficiency scores. Inclusion of bowlers with high efficiency scores in a team increases chance of success.

All-rounder statistics

An all-rounder in cricket is a player who is good in batting as well as in bowling. In order to choose an all-rounder, the performance measures of batting as well as bowling are considered. Therefore, we use all

Table III. Top 10 bowlers for IPL 4.

Rank	Player names	Efficiency
1	Lasith Malinga	1
2	Rahul Sharma	1
3	Mitchell Marsh	1
4	Daniel Vettori	0.939
5	Ali Murtaza	0.896
6	Munaf Patel	0.885
7	Iqbal Abdulla	0.877
8	Doug Bollinger	0.856
9	Chris Gayle	0.856
10	Harbhajan Singh	0.854

the above mentioned five batting and four bowling measures for 10 available all-rounders in IPL 4. Using DEA model (4) with nine outputs, we find four efficient all-rounders, Yuvraj Singh, Johan Botha, James Hopes and Yusuf Pathan. To rank these all-rounders, we can use dual model (5) by imposing non-zero weights in the model, that is, $u_r \geq \varepsilon$, $r = 1, \ldots 9$, where ε is a non-Archimedean value. This is defined to consider the importance of all nine parameters in the evaluation of all-rounders. In fact, if we ignore to impose these additional constraints then some parameters may have zero weight in the optimal solution of dual model (5). This means that some parameters with zero weights will not affect the performance of all-rounders. To find a suitable non-Archimedean value we can solve the following DEA model:

$$
\begin{aligned}
&\max \varepsilon \\
&s.t. \\
&-\sum_{r=1}^{9} y_{rj} u_r + u_0 \geq 0 \quad j = 1, \ldots, 10 \\
&\sum_{r=1}^{9} y_{rk} u_r = 1 \\
&u_r - \varepsilon \geq 0 \quad r = 1, \ldots, 9
\end{aligned}
$$

Assume ε_k^* be the optimal value of the above model, a suitable value for Epsilon in the dual DEA model can be $\varepsilon = \min\{\varepsilon_k^* : k = 1, \ldots, 9\}$. It is not difficult to show that for this Epsilon the corresponding dual model (4) for all 10 all-rounders will be feasible. For 10 all-rounders, this value is $\varepsilon = 0.003533$. Now, we impose constraints $u_r \geq 0.003533$ (for all $r = 1 \ldots, 9$) and re-rank all-rounders. The best two all-rounders identified by dual DEA model (5) are Yuvraj Singh and Yusuf Pathan, respectively. This ranking is very close to the opinion of cricket fans and experts in India.

Wicketkeeper statistics

A wicketkeeper should be a good batsman, so all performance measures of batsman are included. Using DEA model (4) for 10 wicketkeepers who played in IPL 4 the top two wicketkeepers were MS Dhoni and AB de Villiers, respectively.

It is evident that the best wicketkeeper in the whole IPL 4 was MS Dhoni. Therefore, it will be a big asset to have such wicketkeeper with the highest performance.

Cricket team selection

One important issue in the cricket team selection is the way of aggregation of multiple performance measures related to players' available statistics. This paper deals with the benefit of introducing a linear programming based DEA model for the aggregation. This avoids the inclusion of subjective computation for finding players' scores. In previous sections, we obtained top best cricket players who have the highest performance in IPL 4. According to the results, finding a cricket team is equivalent to solving the following integer programming model:

$$
\begin{aligned}
&\max \sum_{i=1}^{4} \sum_{j=1}^{n_i} \theta_j^i x_{ij} \\
&s.t. \\
&\sum_{j=1}^{n_i} x_{ij} = k_i \qquad i = 1, 2, 3, 4 \\
&x_{ij} \in \{0, 1\}
\end{aligned}
$$

Where, θ_j^i is the efficiency score of the j^{th} player in the i^{th} group, $i = 1$for batting, $i = 2$ for bowling, $i = 3$ for all-rounders, and $i = 4$ for wicketkeepers and

$$
x_{ij} = \begin{cases} 1 \text{ if the } j\text{th player is selected in the } i\text{th dicipline} \\ 0 \qquad\qquad \text{otherwise} \end{cases}
$$

In fact, the optimal solution of the above integer programming problem is at hand without solving the model. We can pick six batsmen ($k_1 = 6$), five bowlers ($k_2 = 5$), two all-rounders and two wicket-keepers ($k_3 = k_4 = 2$). This team of 15 players can be considered as chosen by selectors. A team of 11 players is to be chosen by the captain and the coach depending on the weather and playground conditions among others. Therefore, the best cricket team identified by DEA method for IPL 4 is given in Table IV.

One important issue related to the proposed DEA method in this paper is that the objective is not to compare the efficiency of a cricket team in terms of the efficiencies of individual players. In fact, it is

Table IV. Cricket team chosen for IPL 4.

Capabilities	Players in the team
Batsmen	Virender Sehwag
	Chris Gayle
	Paul Valthaty
	Michael Hussey
	Tirumalsetti Suman
	Subramaniam Badrinath
Bowlers	Lasith Malinga
	Rahul Sharma
	Mitchell Marsh
	Daniel Vettori
	Ali Murtaza
All-rounders	Yuvraj Singh
	Yusuf Pathan
Wicketkeepers	MS Dhoni
	AB de Villiers

incorrect to assume efficiency of a team is the same as the sum of efficiencies of individual players (see Kuosmanen et al., 2010). In fact, this paper is the first attempt using DEA to have a cricket team from players who played in a tournament, IPL 4, for different clubs. This can be used for choosing a national team or a team of cricket stars from players of different clubs. Further research needs to compare and evaluate the performance of a cricket team based on the performance of individual players using DEA.

Conclusion remarks

This paper suggested a novel application of DEA for selection of a cricket team. We proposed a multiple outputs DEA formulation to assess cricket players in different capabilities. A real dataset consisting of cricket players in Indian Premier League for 2011, known as IPL 4, is used to find the best cricket team. The proposed method has the benefit of using an optimisation DEA model for aggregation of multiple performance measures. This avoids the inclusion of subjective computation in the cricket team selection. This paper is the first attempt to form a cricket team from players who played in a tournament for different clubs. Further research needs to compare and evaluate the performance of a cricket team from performance of individual team's players.

Acknowledgements

The authors thank two anonymous reviewers and the editor of EJSS for their constructive suggestions that improved the current paper substantially. The authors also appreciate Shubhranshu R. Singh at HASS, University of California, Berkeley, for his useful comments on the revised version of this paper.

References

Ahmed, F., Jindal, A., & Deb, K. (2011). Cricket team selection using evolutionary multi-objective optimization. *Evolutionary, and Memetic Computing, Lecture Notes in Computer Science, 7077*, 71–78.

Amin, G. R., & Emrouznejad, E. (2010). Finding relevance search engines results: A minimax linear programming approach. *Journal of the Operational Research Society, 61*, 1144–1150.

Amin, G. R., Gattoufi, S., & Rezaee Seraji, E. (2011). A maximum discrimination DEA method for ranking association rules in data mining. *International Journal of Computer Mathematics, 88*, 2233–2245.

Barr, G. D. I., Holdsworth, G. C., & Kantor, B. S. (2008). Evaluating performances at the 2007 cricket world cup. *South African Statistical Journal, 42*, 125–142.

Bazarra, M., Jarvis, J. J., & Sherali, H. D. (2010). *Linear programming and network flows* (4th ed.). Hoboken, NJ: John Wiley & Sons.

Bhattacharya, R., Gill, P. S., & Swartz, T. B. (2011). Duckworth-Lewis and twenty20 cricket. *Journal of the Operational Research Society, 62*, 1951–1957.

Bracewell, P. J., & Ruggiero, K. (2009). A parametric control chart for monitoring individual batting performances in cricket. *Journal of Quantitative Analysis in Sports, 5*, 1–19.

Cooper, W. W., Seiford, L. M., & Tone, K. (2006). *Introduction to data envelopment analysis and its uses with DEA-solver software and references*. New York, NY: Springer.

Gerber, H., & Sharp, G. D. (2006). Selecting a limited overs cricket squad using an integer programming model. *South African Journal for Research in Sport, Physical Education and Recreation, 28*, 81–90.

Hadley, L., & Ruggiero, J. (2006). Final-offer arbitration in major league baseball: A nonparametric analysis. *Annals of Operations Research, 145*, 201–209.

Hass, D. J. (2003). Productive efficiency of English football teams – A data envelopment analysis approach. *Managerial and Decision Economics, 24*(5), 403–410.

Iyer, S. R., & Sharda, R. (2009). Prediction of athletes performance using neural networks: An application in cricket team selection. *Expert Systems with Applications, 36*, 5510–5522.

Kuosmanen, T., Kortelainen, M., Sipiläinen, T., & Cherchye, L. (2010). Firm and industry level profit efficiency analysis using absolute and uniform shadow prices. *European Journal of Operational Research, 202*, 584–594.

Lemmer, H. H. (2011). Team selection after a short cricket series. *European Journal of Sport Science*. doi:10.1080/17461391.2011.587895.

Lewis, H. F., Lock, K. A., & Sexton, T. R. (2009). Organizational capability, efficiency, and effectiveness in major league baseball: 1901–2002. *European Journal of Operational Research, 197*(2), 731–740.

Llamazares, B., & Peña, T. (2009). Preference aggregation and DEA: An analysis of the methods proposed to discriminate efficient candidates. *European Journal of Operational Research, 197*, 714–721.

Preston, I., & Thomas, J. (2000). Batting strategy in limited overs cricket. *Journal of the Royal Statistical Society, Series D (The Statistician), 49*, 95–106.

Ruggiero, J. (2010). *Frontiers in major league baseball: Nonparametric analysis of performance using data envelopment analysis* (1st ed). New York, NY: Springer.

Sharp, G. D., Brettenny, W. J., Gonsalves, J. W., Lourens, M., & Stretch, R. A. (2011). Integer optimization for the selection of a Twenty20 cricket team. *Journal of the Operational Research Society, 62*(9), 1688–1694.

Van Staden, P. J. (2009). Comparison of cricketers' bowling and batting performances using graphical displays. *Current Science, 96*, 764–766.

Wang, Y. M., & Chin, K. S. (2009). A new data envelopment analysis method for priority determination and group decision making in the analytic hierarchy process. *European Journal of Operational Research, 195*, 239–250.