European Journal of Operational Research 255 (2016) 559–569

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Assessing the scoring efficiency of a football match

G. Villa*, S. Lozano

Escuela Superior de Ingenieros, University of Seville, Camino de los Descubrimientos, s/n, 41092 Seville, Spain

ARTICLE INFO

Article history:
Received 20 February 2015
Accepted 12 May 2016
Available online 20 May 2016

Keywords:
OR in sports
Data Envelopment Analysis (DEA)
Football scoring efficiency
Parallel processes
Home/away team

ABSTRACT

Data Envelopment Analysis (DEA) has been widely used for efficiency assessment in many sports and specifically in football. This paper presents a novel approach to measuring the scoring efficiency of football teams in a match. It uses a parallel process network DEA model that takes into account the defensive and attacking variables of the two teams, their respective percentage of possession of the ball and also the team's economic value. The proposed approach provides estimates of how many more goals each team should have scored in each of the matches played in the season. By averaging the scoring efficiency in the different matches played, team scoring efficiencies for the season can be computed. Moreover, the proposed DEA model can be run after each match so that the changes in efficiency can be monitored throughout the season. This can help coaches and managers assess their team's past performance and plan ahead for each match. The proposed approach has been applied to Spanish First Division teams for the season 2013/2014 and a virtual final league table has been computed.

1. Introduction

Association football, also known as soccer in some parts of the world, is one of the most popular team sports, played by circa 300 million people all over the world. It also has the highest global television audience in sport. In football, there is no direct relationship between the number of goals scored and the points earned in a match since, in the end, it is just necessary to score one goal more than the opposing team to win a match (earning 3 points) or to score the same number of goals for a draw (earning 1 point each). However, in most cases, scoring as many goals as possible secures the win, and therefore this is a major objective for each team. Of course, the opposing team will try to neutralize these efforts at the same time that it pursues its own objectives. This competition, this struggle, is the beauty of the game.

An inefficient use of resources on the part of the teams may lead them to score fewer goals than they could have. It may thus be useful to have a way of assessing how many goals each team could have scored had they been efficient in this regard. Note also that, one of the reviewers pointed out, the number of goals scored by the teams is important as in some cases (e.g. Brazilian and Spanish championships) the number of goals scored is used as one of the rules for breaking ties. This paper proposes a method for estimating the maximum number of goals that could have occurred in a match given the resources used by each team. This allows computing the scoring efficiency of both teams (and of the match itself) calculated as a function of the difference between the number of goals scored in the match and the number of goals that could have been scored. This number of goals maximization objective corresponds to a system point of view. Thus, from being a neutral spectator watching the match, the greater the number of goals scored, the more interesting the match normally becomes.

However, the motivation and the usefulness of the proposed approach is not so much related to how interesting the match is or the utility of the match spectators but to provide team coaches and managers with a tool to assess the efficiency of the two teams, given the data of the match. This efficiency assessment, together with other ex post analyses that the coaching personnel may also carry out, can help them to understand the issues and weaknesses (and also the strengths) that each team has shown in the match. Anything that can help correctly diagnose problems in a timely way is helpful in order to try to solve them for subsequent matches. That is why the proposed approach, which can be applied after each match, as a kind of forensic study, can be useful with the added benefit that the proposed approach assesses the efficiency of both teams at the same time and takes into account the competitive, zero-sum nature of the match.

The structure of the paper is the following. In Section 2 a literature review is carried out. In section 3 the network DEA model of scoring efficiency in football matches is presented. Section 4 presents and discusses the results of applying the model to the 38 matches played in the Spanish First Division in the season 2013/14. The last section summarizes and concludes.

* Corresponding author. Tel.: +34 954487207.
E-mail address: gvilla@us.es (G. Villa).

http://dx.doi.org/10.1016/j.ejor.2016.05.024
0377-2217/

2. Literature review

Data Envelopment Analysis (DEA) is a well-known non-parametric technique for evaluating the relative performance of similar Decision Making Units (DMUs). There are many applications of the DEA methodology to sports: Olympic Games (e.g. Lozano, Villa, Guerrero, & Cortés, 2002; Li, Liang, Chen, & Morita, 2008; Soares de Mello, Angulo-Meza, & Branco da Silva, 2009; Wu, Liang, & Chen et al., 2009; Wu, Zhou, & Liang, 2010; Zhang, Li, Meng, & Liu, 2009); basketball (e.g. Cooper, Ruiz, & Sirvent, 2009, Cooper, Ramón, Ruiz, & Sirvent, 2011); baseball (e.g. Sexton & Lewis, 2003; Ruggiero, 2011), etc.

With respect to DEA applications to football, most of them focus on the performance of teams throughout the seasons. Table 1 shows a review of those studies.

Two viewpoints can be distinguished: those papers that determine the efficiency through economic indicators, and those that focus on sports factors. Regarding the first approach, the authors extract information from teams' financial statements to evaluate corporate performance. Staff costs (often distinguishing between the coach's wage and the rest of the salaries), operating costs, stadium facilities expenditures, total season revenues, turnover and tickets sold are some of the variables considered in this approach. Except in Barros and García-del-Barrio (2011), input oriented CCR (Charnes, Cooper, & Rhodes, 1978) and BCC (Banker, Charnes, & Cooper, 1984) models are used to analyze the financial performance of the clubs.

Since Espitia-Escuer and García-Cebrián (2004) used DEA methodology to assess the sporting performance of Spanish football teams by considering attacking moves throughout the length of the season, many other studies followed this approach. Although the majority of the studies focus on the performance of the teams when playing in the national First Division League, some papers have analyzed the efficiency of teams when playing in other competitions, such as the Champions League or Spanish King's Cup.

With the aim of incorporating the details involved in the matches played into the performance assessment as accurately as possible, most of the articles make a distinction between attacking and defensive variables, although some authors do not include the latter because the effect of the opposing teams is homogeneous for the entire sample when the data refer to a whole season (Espitia-Escuer & García-Cebrián, 2004, 2006, 2008). When dealing with these two types of variable, authors consider two separate models, in order to obtain the defensive and the attacking efficiencies of the teams. The attacking DEA model variables are the attack skills of the team (inputs), and the goals and points obtained in a season (outputs). On the other hand, the defensive model inputs can be either the defensive abilities of the team whose performance is measured (e.g. García-Sánchez, 2007) or the inverse, i.e. the attacking abilities of the opposing team (e.g. Boscá, Liern, Martínez, & Sala, 2009; Sala, Liern, Martínez, & Boscá et al., 2009) while the defensive model output is the inverse of the number of conceded goals.

The most common results found are the technical and scale efficiency of the teams in a season, and the productivity change between seasons computed using the Malmquist index. However, other conclusions can be extracted from solving these DEA. Thus, for example, the potential final league table is calculated by Espitia-Escuer and García-Cebrián (2006), by calculating how many points the teams would have achieved had they been efficient. Also, González-Gómez and Picazo-Tadeo (2010) assess the degree of satisfaction of the supporters of Spanish football teams, evaluating the gap between the performances obtained in League, King's Cup, and European competitions and their potential. Picazo-Tadeo and González-Gómez (2010) focus their study on the influence that the extra matches played in other competitions have on the performance of the Spanish league's football teams. Finally, Boscá et al. (2009) distinguish two types of efficiency scores, depending on whether the team plays at home or away.

Also, some studies apply DEA to other issues related to football, that are different from teams' performance. Thus, for example, Alp (2006) applies a CCR-O model to evaluate the performance of 2002 FIFA World Cup goalkeepers while Tiedemann, Francksen, and Latacz-Lohmann (2011) analyze the performance of a sample of German football players from season 2002/03 to 2008/09 by considering their playing positions using a metafrontier approach based on a BCC-O model.

From a methodological point of view, this paper uses network DEA, which assumes that the system consists not just of a single process but of distinct processes which are usually interrelated. Thus, it is common in network DEA to identify intermediate products produced by some processes and consumed by others. Network DEA models provide a more fine-grained level of analysis and has more data requirements. The results are also more detailed and more useful, since they can provide information not only about the performance of the whole system but also of each of its subsystems. Thus, for example, contrary to what occurs in conventional DEA, in network DEA it is often the case that no DMU is relative efficient. The reason for this apparent anomaly is that in network DEA for a DMU to be efficient all its processes must be efficient, something which, in general, is not easy to achieve. That is because the operation of each process is benchmarked against the operation of the corresponding process of all other DMUs. In other words, each process has its own production possibility set (i.e. its own technology).

The literature on network DEA has expanded greatly in recent years (e.g. Kao, 2009; Tone & Tsutsui 2009; Cook, Zhu, Bi, & Yang, 2010; Fukuyama & Mirdehghan, 2012; Chen, Cook, Kao, & Zhu, 2013; Sahoo, Zhu, Tone, & Klemen, 2014; Kao & Hwang, 2014; Kao 2014a, 2014b). The number of applications has also grown significantly and spans many sectors, such as banking (e.g. Lozano, 2015), environmental performance (e.g. Chen, Zhu, Yu, & Noori, 2012), transportation (e.g. Lozano & Gutiérrez, 2014), supply chains (e.g. Chen & Yan, 2011), etc. There are also a number of network DEA approaches to sports. Thus, for example, Lewis, Lock, and Sexton (2009) study the Major League Baseball considering that each team consists of three processes: offense (that consumes offensive measures and produces runs gained), defense (that consumes defensive measures to prevent runs surrendered) and integration process (that consumes the intermediate products runs gained and surrendered and produces the final output, wins). Yang, Lin, and Chen (2014) propose a two-stage approach for measuring the efficiency of NBA teams using total players' salaries as input and games won and gate receipts as outputs, and an aggregated players' performance index as the intermediate product. Moreno and Lozano (2014) propose a more complex five-process approach for modeling basketball first team and bench team, defensive/attacking and win generation performance. Two stages (namely athlete preparation and athlete competition) are also considered by Li, Lei, Dai, and Liang (2015) to assess the performance of participating nations at the Olympic Games. Lei, Li, Xie, and Liang (2015) present a network DEA approach, also with two processes (Summer Olympics and Winter Olympics) but in a parallel-process configuration and with shared inputs.

There are not, however, many network DEA approaches to football. The existing approach that most closely resembles network DEA is that of García-Sánchez (2007), which considers three stages. In the first one, two DEA models are proposed to evaluate the defensive and attacking efficiencies. These efficiencies are used as inputs in a third DEA model (named the second stage), to evaluate the operating effectiveness by considering the position in the final league table of the team as output. Finally, a fourth DEA model