Ranking Players by DEA: An analysis of Czech and Danish football

Michal Tomíček[1], Natalie Pelloneová[1]

[1]*Technical University of Liberec, Faculty of Economics*

Abstract

Team managers and coaches need to choose the best players. The selection relies mainly on the cost and performance of the entire team. It is a common practice that several key players contribute to the overall results of the football team. The quality of players is one of the crucial features determining the failure or success of a sports team. The present article focuses on measuring player efficiency in the Czech and Danish top football competitions during the 2015/16 to 2019/20 seasons. The presented research aims to identify the most technically efficient players, considering their position on the field. The authors used an input-oriented model of data envelopment analysis and subsequently also cluster analysis to determine the best football players. The following article may be of interest to football club managers, football analysts, economists and others interested in the business of football because it combines two methods of measuring the efficiency of football players.

Keywords: *data envelopment analysis, efficiency, performance evaluation, cluster analysis, CCR-I model*

INTRODUCTION

Football is the most popular sport in the Czech Republic and Denmark, although other sports, such as hockey and tennis, are also very popular. The football business has grown significantly over the last few decades, and football clubs have become large companies. To be profitable and thrive in the pitch, they need to improve the efficiency of their business (Pyatunina et al., 2016). The growing worldwide popularity of professional football has made the sport an expanding industry with a substantial financial turnover. The fundamental factors determining the performance of a football team include the selection of players and the team's organisation, taking into account various economic and technical constraints. Player selection has become particularly important as their salaries and transfer costs place increasing restrictions on professional football clubs, which have to select the best possible pool of players within the constraints of their budgets (Nasiri et al., 2018). In addition, the player's market value is an influential economic valuation determining not only the player's price but also various other factors. A player's market value is a score dependent on numerous factors such as the talent, popularity, skill, playing style, efficiency, etc. of a player (Singh & Lamba, 2019). Market value is the economic evaluation of a player that signifies the player's price and performance. In this context, professional football team managers should develop management practices that help clubs evaluate and select the best players to achieve two objectives simultaneously: success on the pitch and commercial performance.

Efficiency and sports are two closely related concepts. A considerable amount of professional literature has been written on sports management. The economic framework of professional sports activities relies on the work of Rottenberg (1956), Jones (1969) and Sloan (1969). Choosing players and the sports team structure represents a fundamental decision of club owners and managers in their efforts to achieve the best possible results. Obviously, the success or failure of any sports team depends primarily on the skills of its players and the organisation of their potential teams. In sports science, data envelopment analysis has been widely used in recent decades to evaluate performance and subsequently support decision-making in various sports. Several of the studies listed below have applied data envelopment analysis (hereinafter DEA) to assess sports teams in different sports disciplines. For example, Chitnis and Vaidya (2014) and Ruiz et al. (2013) deployed the DEA method and evaluated the performance of professional tennis players in terms of the effectiveness of their game. DEA models have also been used to evaluate baseball players, for example, by Cooper et al. (2009), Chen and Johnson (2010). Other authors have applied the DEA method to evaluate golfers (Fried et al., 2004).

In addition, experts have applied parametric and non-parametric analyses to determine the performance of football teams in the last few decades. The efficiency of European professional football clubs has been evaluated by Espitia-Escuer and García-Cebrián (2004), Guzmán and Morrow (2007), Halkos and Tzeremes (2013), Pyatunina et al. (2016), among others. Djordjević et al. (2015) applied the DEA method to rate the performance of the national football team in the qualifications for the 2010 FIFA World Cup. Ribeiro and Lima (2012) employed data envelopment analysis to measure the effectiveness of Portuguese football clubs in the first league from 2002/03 to 2008/09 seasons. Despite the wide popularity of this sport, there are only a limited number of scientific studies available focusing on the statistical, economic dimensions and performance evaluation of individual players. However, only a few studies have concentrated on individual footballers when applying data envelopment analysis. An overview of authors using data envelopment analysis to evaluate player performance is summarised below.

Haas (2003) measured team and player effectiveness in Major League Soccer using data envelopment analysis (DEA). Hirotsu et al. (2012) evaluated the players' performance in Japan's top J-league soccer competition during the 2013 season using DEA. They analysed the players' on-field performance using input-oriented Banker Charnes and Cooper (BCC) and Charnes Cooper and Rhodes (CCR) models of data envelopment analysis. Hirotsu et al. (2018) selected evaluation indicators for individual football players and used data envelopment analysis to identify the player's performance characteristics. The players' on-field performance was then analysed using the DEA super-efficiency model regardless of the players' game position. In their research, Tiedemann et al. (2011) presented a model for evaluating the performance of football players in the field. Based on the data envelopment analysis, they used a non-concave metafrontier approach to estimate players' efficiency scores concerning their playing position. They applied this approach to the data set of German Premier league footballers in the 2002/03 to 2008/09 seasons. Their research included one input variable (playing time) and four output variables (goals, assists, tackle ratio and pass completion ratio) in the DEA model. The research results revealed a positive relationship between the average player efficiency score and the team's placement in the league table at the end of the season. In addition, a metafrontier approach was used to identify the optimal playing position of a football player in the team and, subsequently, quantify the increase in performance when moving to that playing position. Choosing the best players is also very important for team managers and coaches. Player selection is primarily related to the cost and performance of the team. Arabzad et al. (2013) proposed a two-stage approach to selecting and evaluating the best English Premier League football players in 2010/11. An output-oriented BCC model was used to identify the best players and evaluate chosen players. The BCC model identified 29 players as efficient. Efficient players were further sorted using the DEA ranking model. The results show that the best players are Rooney, Drogba and Tevez. Erickson and Callum (2004) used data envelopment analysis to evaluate the National Football League players. Alp (2006) applied the output-oriented CCR model and the Andersen-Petersen super-efficiency model to rank 32 goalkeepers from the 2002 World Cup. The CCR model identified 12 of the 32 goalkeepers as efficient (e.g. Oliver Kahn, Gianluigi Buffon, Iker Casillas and others). Fernández et al. (2020) analysed players' efficiency in the Spanish Soccer League for the 2009/2010 season using a metafrontier version of data envelopment analysis methodology. In their research, they divided the sample of players into three groups according to their game position in the team. They also included specific outputs for each game position to characterise these positions better. In addition, they compared the efficiency of Spanish and foreign players and analysed the correlation between the number of efficient players per team and the points won by each team. Research has found that defenders and midfielders are more efficient than forwards. This is due to the fact that the forwards accumulate the inefficiency of midfielders, who in turn accumulate the inefficiency of the defenders. Research has found that the majority of efficient defenders and midfielders are Spanish. On the other hand, almost 60% of efficient forwards are foreigners. These forwards include, for example Lionel Messi and Cristiano Ronaldo. The correlation coefficient of 0.928 confirmed the existence of a direct relationship between the number of efficient players and the number of points won in the league. Papahristodoulou (2007) evaluated the individual performance of the forwards in the UEFA Champions League. The author warned that only forwards who scored two or more goals were evaluated due to software limitations. The DEA model identified the seven forwards who scored the most points as effective: Kaká, Cristiano Ronaldo, Ronaldo, Miccoli, Mpenza, Fowler and Giggs.

This paper proposes an approach set within the context of football players' evaluation. It focuses on the players in the top Czech and Danish football leagues playing in the 2014/15 to 2019/20 seasons. All available players have been evaluated using an input-oriented DEA model (Banker et al., 1984; Charnes et al., 1978). The first objective of the presented research is to use data envelopment analysis to identify and compare the efficient players of both football competitions under study and, subsequently, to determine whether the most valuable players of the assessed football competitions are efficient. The second objective is to apply cluster analysis to divide the set of players into compact clusters and, subsequently, determine what the individual clusters reveal about the players of the football competitions.

METHODS AND DATA

Data

Players from two top European competitions with potential similarities were selected for analysis and subsequent comparison from the football environment. They are the Czech Fortuna:Liga and the Danish 3F Superliga. These competitions are similar in terms of the same game model and similar positions in the UEFA league coefficients. Considering the data availability in both countries, the player efficiency was analysed for the 2015/16 to 2019/20 seasons. The research period was mainly limited by the data available for the Danish 3F Superliga, as the oldest available data for managed sports came from the 2015/16 season. Data from the 2013/14 season is also available for the Czech top competition; however, they were not included in the research to compare the two competitions.

The research focuses only on players who have played at least one game in a season. Players who did not play in any games can be assigned a market value, but must play in at least one game in the season to be assigned game performance statistics. For players playing for more than one club in the same season, the number of games played in the season for the clubs was the determining factor for club assignment. The player was assigned to the club for which he played more games in the season, with the number of minutes being the tie-breaker. Players on loan were assigned the club they played for in the season.

The research data used have been obtained from the official databases of both sports competitions and completed by private databases of companies from the football environment. The core of the research data came from InStat (2020), which analyses the performance of athletes and sports teams. InStat provides match data and detailed statistics for individual players. The set is supplemented by databases of the Transfermarkt.com server (2021), the Czech Fortuna:Liga (2021), and the Danish 3F Superliga (2021). The Transfermarkt.com database was used mainly to obtain information about the market values of players in each season.

Data Envelopment Analysis

In the presented research, data envelopment analysis was used to evaluate the efficiency of individual players. The data envelopment analysis is a benchmarking tool initially developed by Farrell (1957) and further elaborated by Charnes et al. (1978). DEA is a method based on linear programming that evaluates the efficiency of production units. These units are called decision-making units (DMUs) and can include schools, hospitals, universities, banks, companies or even individuals, such as employees or players (Charnes et al., 1978). In the case of this research, DMUs are football players. The DEA compares inputs and outputs and considers whether the units under study achieved efficient or inefficient results. The best performing player in the set of DMUs is assigned a score of 100 per cent or 1. The remaining DMUs are scored between 0 to 100 per cent or equivalent between 0 and 1 relative to the score of the best performing DMU. The fundamental DEA feature is that a certain weight can be assigned to each input and output, i.e., for each DMU separately. The weights are attributed to maximising the efficiency of each DMU. It is a strong argument if a DMU is inefficient because inefficiency is achieved despite employing weights as favourable as possible to that DMU. The DEA was applied to a set of players from both competitions. For the purpose of the data envelopment analysis and a higher informative value, all players were divided by their game posts into goalkeepers, defenders, midfielders, and forwards.

The first step is to select appropriate variables for the DEA model. The market value of the player was chosen as the input variable. The output variables had to be selected based on three requirements. The first requirement was a minimum correlation within the output variables. The resulting effect could be duplicated with a high correlation of some variables. The second requirement was a high correlation of the output variables with the input variable (i.e., market value). Otherwise, the analysis would lose its informative weight. The third requirement was determining the optimal number of variables concerning the number of evaluated units (i.e., players). For example, Zhu (2014) stated that a large number of inputs and outputs compared to the number of DMUs could negatively affect the discriminatory power of the DEA. Cooper et al. (2011) introduced a rule that the number of DMUs (i.e., players) should be at least double or even triple the number of inputs and outputs. The present research has no problem in fulfilling this condition. In fact, the players' data sets are several times larger than the number of inputs and outputs.

Within the output variables for individual groups of players, factors were selected that correlate to the player's post. For forwards, mainly offensive factors were considered (goals scored, assists, shots on target, etc.). Defenders were assessed based on defensive factors (total number of challenges won, number of ball recoveries, air challenges won, etc.). A balanced mix of defensive and offensive factors was adopted for the midfielder group. Goalkeepers formed a separate group in terms of selecting outcome factors, consisting of general factors (e.g., accurate passes) and typical characteristics such as the number of shots saved, etc.

The output variables characterised by the factors for each player group are summarised in Table 1. Some output factors are common to multiple game posts. These include, for example, the number of matches played during the season, a factor that is common to all game positions. Other factors are specific to a particular group of players.

Table 1. DEA output factors (CCR-I) for individual player posts

Output factors according to the player's game post			
Goalkeepers	Defenders	Midfielders	Forwards
Matches played	Matches played	Matches played	Matches played
Short passes accurate	Goals	Goals	Goals
Medium passes accurate	Assists	Assists	Assists
Long passes accurate	Fouls suffered	Fouls suffered	Fouls suffered
Close range shots saved	Shots on target	Shots on target	Penalties scored
Mid range shots saved	Key passes accurate	Key passes accurate	Key passes accurate
Long range shots saved	Passes accurate	Passes accurate	Chances successful
	Defensive challenges won	Challenges won	Challenges won
	Offensive challenges won	Chances created	Chances created
	Air challenges won	Crosses accurate	Crosses accurate
	Dribbles successful	Dribbles successful	Dribbles successful
	Ball recoveries		Tackles successful
	Ball interceptions		
	Free ball pick ups		

In the next step, the DEA model was applied to the selected variables. The input-oriented CCR-I model was applied to both competitions, considering constant returns to scale. The input-orientation of the model was adopted given that the sports environment belongs among industries where the input variable, i.e., the market value of a player, can be influenced rather than the output variables, representing the sporting factors of individual players (Guzmán-Raja & Guzmán-Raja, 2021). In the basic input-oriented CCR model with constant returns to scale assumption, the objective function (1) is maximised under restrictive conditions (2), see e.g. Jablonský and Dlouhý (2004). The symbol x_j denotes inputs, y_i outputs, u_i is the weight of the output, v_j is the weight of the input and z is the value of the objective function.

$$\mathrm{z} = \sum_{i=1}^{r} u_i y_{iq} \qquad (1)$$

$$\sum_{i=1}^{r} u_i y_{ik} \leq \sum_{j=1}^{m} v_j x_{jk}, \; k = 1, 2, ..., n \qquad (2)$$

$$\sum_{j=1}^{m} v_j x_{jq} = 1$$

$$u_i \geq \varepsilon, i = 1, 2, ... r, \varepsilon \text{ - very small non- Archimedean number } (> 0)$$
$$v_j \geq \varepsilon, \; j = 1, 2, ..., m.$$

A technical efficiency score was determined for each player; the calculations were performed using open software OSDEA-GUI (Open source DEA, 2021). The technical efficiency score obtained using the CCR-I model was termed the overall technical efficiency score.

Cluster analysis

Based on the factors derived from the data envelopment analysis method, a cluster analysis of the players of the Czech Fortuna:Liga and the Danish 3F Superliga was subsequently performed. Cluster analysis is based on the idea of sorting players into clusters so that players belonging to the same group are more similar than players from other groups (Aggarwal & Reddy, 2014). The players were divided into clusters according to the similarity of the values of their individual characteristics. In the process of cluster analysis, the nearest neighbour technique was used (3).

$$D_{NN}(A,B) = min\{d(a;\ b)\}\ (3) \qquad (3)$$

In this case, data were divided into n clusters, one for each observation. Then, the calculation of the minimum distance between all pairs of points that are located in different clusters was done. Two mutually closest clusters were joined together. The process of calculation of the minimum distance between all pairs and joining those mutually closest pairs was repeated until the number of clusters has been reduced to the chosen extent. The optimal number of clusters was chosen by inspecting the dendrogram created using hierarchical clustering. Cluster analysis was performed using Statgraphics Centurion 18 statistical software (Statgraphics Technologies, 2021).

RESEARCH RESULTS

Initially, the CCR-I model was applied to a set of 180 goalkeepers who had played at least one match in one of the evaluated seasons of the Czech Fortuna:Liga and the overall technical efficiency score was determined. The CCR-I model identified three goalkeepers as efficient (see Table 2). All of them had a relatively low market value (input variable). Jan Šeda and Martin Berkovec played 32 games in the 2018/19 season. Jaroslav Drobný played 18 games in the 2019/20 season. All three goalkeepers performed above average in the passing and interventions. However, none of the efficient goalkeepers was the best in any statistics. Ondřej Kolář (6 million euros) and Florin Nita (2.5 million euros) posted the highest market value. However, Ondřej Kolář and Florin Nita achieved a very low overall technical efficiency score (max 0.215 and max 0.092). Thus, according to the CCR-I model, they can be considered inefficient.

Table 2 present the efficient goalkeepers of the Danish 3F Superliga based on the CCR-I model. According to the overall technical efficiency score, Jesper Rask was the most efficient goalkeeper of the 139 goalkeepers who had played at least one game in the top Danish competition. In terms of short and medium passes, this goalkeeper was below average than the goalkeepers of the Danish highest competition but above average in long passes, especially in the statistics of saved shots. The other efficient goalkeeper was Thomas Mikkelsen, who reached the maximum long range shots saved in the 2019/20 season. He also achieved an efficient score with excellent other statistics. Kamil Grabara (5 million euros) and Frederik Ronnow (3 million euros) had the highest market value. However, both achieved a very low overall technical efficiency score and can be classified as inefficient according to the CCR-I model.

Table 2. Efficient goalkeepers of the Czech and Danish leagues according to the CCR-I model

Goalkeeper	League	Club	Season	Market value (ths. €)
Jan Šeda	CZ	FK Mladá Boleslav	2018/19	250
Martin Berkovec	CZ	MFK Karviná	2018/19	250
Jaroslav Drobný	CZ	SK Dynamo České Budějovice	2019/20	100
Jesper Rask	DK	Hobro IK	2015/16	200
Thomas Mikkelsen	DK	Lyngby Boldklub	2019/20	150

Next, the overall technical efficiency score was calculated for a group of 722 defenders who had played at least one match in any of the observed seasons of the Czech Fortuna:Liga. Only one per cent of the defenders (nine defenders in total) of the Czech Fortuna:Liga were considered efficient from the perspective of the CCR-I model. Martin Jiránek was marked as an efficient unit in two seasons (2016/17 and 2017/18). The other players (see Table 3) played a maximum of six matches per season. These were young substitute players who were gaining experience in their first seasons in professional football. Defender Radim Černický is a typical example: He played 130 minutes in four matches for the FC Slovan Liberec in the 2019/20 season and scored two assists and two shots on target.

In the Danish 3F Superliga, the CCR-I model has identified 17 efficient defenders out of a total of 593 defenders (i.e. 2.86 %). A total of 13 of these defenders played more than 15 games in a season and seven of them even more than 30 games. Bjorn Paulsen was the most versatile player in terms of relative value balance compared to other efficient defenders. He also added three goals and two assists to the highest number of won air and offensive challenges. The Danish league's market value of efficient defenders was at a higher financial level than that of the Czech defenders (see Table 3).

Table 3. Efficient defenders of the Czech and Danish leagues according to the CCR-I model

Defender	League	Club	Season	Market value (ths. €)
Tomáš Janíček	CZ	FC Fastav Zlín	2015/16	100
Martin Jiránek	CZ	1. FK Příbram	2016/17	50
Martin Jiránek	CZ	FK Dukla Praha	2017/18	50
Martin Nečas	CZ	FC Fastav Zlín	2019/20	100
Matyáš Kazda	CZ	FC Slovan Liberec	2019/20	50
Tomáš Vincour	CZ	1. FC Slovácko	2019/20	50
Lukáš Červ	CZ	SK Slavia Praha	2019/20	75
Sunday Adetunji	CZ	1. FK Příbram	2019/20	50
Radim Černický	CZ	FC Slovan Liberec	2019/20	50
Marc Pedersen	DK	SonderjyskE Fodbold	2015/16	300
Mads Justesen	DK	Hobro IK	2015/16	200
Kevin Mensah	DK	Esbjerg fB	2015/16	250
Bjorn Paulsen	DK	Esbjerg fB	2015/16	400
Thomas Hansen	DK	Hobro IK	2015/16	100
Johnny Thomsen	DK	Randers FC	2016/17	200
Peter Nymann	DK	AC Horsens	2016/17	250
Johan Absalonsen	DK	SonderjyskE Fodbold	2016/17	200
Mads Bech Sorensen	DK	AC Horsens	2016/17	150
Jakob Ahlmann	DK	Aalborg BK	2017/18	400
Kenneth Petersen	DK	Odense BK	2017/18	300
Michael Baidoo	DK	FC Midtjylland	2017/18	150
Melvin Frithzell	DK	FC Helsingor	2017/18	100
Mads Madsen	DK	Silkeborg IF	2017/18	350
Johnny Thomsen	DK	Randers FC	2018/19	200
Markus Halsti	DK	Esbjerg fB	2018/19	200
Peter Nymann	DK	AC Horsens	2018/19	200

The third group of players analysed represented the midfielders from both football competitions. In the case of the Czech Fortuna:Liga, a set of 934 midfielders was assessed. Altogether, seven players were identified as efficient based on the CCR-I model. The analysis of the midfielders of the Czech Fortuna:Liga was influenced by Pavel Zavadil. The experienced midfielder of Zbrojovka Brno or SFC Opava achieved the efficient result of overall technical efficiency three times (in 2016/17, 2018/19, and 2019/20 seasons). Pavel Zavadil managed to transform his market value into selected game statistics efficiently. In the 2019/20 season, besides Pavel Zavadil, three other players were marked as efficient. They comprised young talents of individual teams who had played no more than one half of the season. More information can be found in Table 4. It can be expected that they will appear more frequently on Czech (or world) pitches in the future, as in the case of Pavel Zavadil, one of the group's four 40-year-olds. The highest market value was attributed to Tomáš Souček (12 million euros), who was, however, identified as an inefficient unit according to the CCR-I model.

In the case of the Danish 3F Superliga, a pool of 836 midfielders was analysed. Altogether, six players were marked as efficient based on the CCR-I model, which is very similar to the Czech Fortuna:Liga case. The most valuable efficient midfielders in Denmark in the period under study were Kasper Risgard (33 years old), who had played for Aalborg BK, and Jonas Borring (32 years old), who had joined FC Midtjylland. Both players consistently played around thirty matches in the same season. Adnan Mohammad, the only Pakistani, became efficient in the jersey of the newcomer from Helsingor. In terms of sporting statistics, it was his best season. He participated in 28 games and excelled, especially in the accuracy of his passes and the number of balls recoveries. Lucas Ohlander also played for the same team in the same season but only participated in one match. The overall overview is shown in Table 4. The Danish league's market value of efficient midfielders reached a higher financial level than the Czech midfielders.

Table 4. Efficient midfielders of the Czech and Danish leagues according to the CCR-I model

Midfielder	League	Club	Season	Market value (ths. €)
Pavel Zavadil	CZ	FC Zbrojovka Brno	2016/17	50
Pavel Zavadil	CZ	SFC Opava	2018/19	50
Pavel Zavadil	CZ	SFC Opava	2019/20	25
Paulo Alves Paulinho	CZ	1. FK Příbram	2019/20	75
Petr Janota	CZ	1. FK Příbram	2019/20	25
Vojtěch Patrák	CZ	AC Sparta Praha	2019/20	50
Pavel Osmančík	CZ	Bohemians Praha 1905	2019/20	50
Kasper Risgard	DK	Aalborg BK	2015/16	300
Martin Mikkelsen	DK	Hobro IK	2015/16	150
Jonas Borring	DK	FC Midtjylland	2016/17	300
Adnan Mohammad	DK	FC Helsingor	2017/18	200
Magnus Westergaard	DK	Lyngby Boldklub	2017/18	50
Lucas Ohlander	DK	FC Helsingor	2017/18	50

The largest group of efficient players represented a pool of forwards of both competitions. Altogether, 18 forwards out of 327 forwards of the Czech Fortuna:Liga can be described as efficient based on the CCR-I model. Table 5 below presents the complete list. The highest number (seven in total) of the efficient players played in the last observed season (2019/20). Only three players did not play at least ten matches in their efficient season. The only player who did not score a single goal in the season was the forward Jan Kuchta at Prague Bohemians. However, he dominated in terms of challenges won per match. A pair of foreign players from Zlín, Pedro Martinez and Lamin Jawo, scored the most goals.

The CCR-I model identified a total of 13 forwards as efficient out of 298 forwards participating in the Danish 3F Superliga. Like the forwards in the Czech Fortuna:Liga, the Danish forwards scored a minimum of goals from penalty kicks. On average, the forwards in Denmark had fewer won challenges (190) than their Czech counterparts (220). On the other hand, the Danish efficient forwards scored more goals on average than the Czech players. AC Horsens had the most efficient forwards of the Danish league, all in the 2016/17 season when the club finished tenth in the regular season. The following Table 5 provides a more detailed overview of efficient Danish forwards. The Danish league's market value of efficient forwards reached the same financial level as that of Czech forwards.

Table 5. Efficient forwards of the Czech and Danish leagues according to the CCR-I model

Forward	League	Club	Season	Market value (ths. €)
Lukáš Magera	CZ	FK Mladá Boleslav	2015/16	350
Roman Bednář	CZ	1. FK Příbram	2015/16	300
Jakub Mareš	CZ	FK Dukla Praha	2015/16	400
Michael Rabušic	CZ	FC Vysočina Jihlava	2016/17	200
Milan Baroš	CZ	FC Slovan Liberec	2016/17	300
Pavel Černý	CZ	FC Hradec Králové	2016/17	175
Jean-David Beauguel	CZ	FK Dukla Praha	2016/17	200
Jan Kuchta	CZ	Bohemians Praha 1905	2016/17	150
Eric Ramirez	CZ	MFK Karviná	2017/18	150
Miroslav Slepička	CZ	1. FK Příbram	2018/19	75
David Puškáč	CZ	Bohemians Praha 1905	2018/19	150
Jiří Kladrubský	CZ	SK Dynamo České Budějovice	2019/20	200
Antonín Fantiš	CZ	FC Fastav Zlín	2019/20	300
Martin Bukata	CZ	MFK Karviná	2019/20	250
Tomáš Pilík	CZ	FK Jablonec	2019/20	200
Pedro Martinez	CZ	FC Fastav Zlín	2019/20	250
Lamin Jawo	CZ	FC Fastav Zlín	2019/20	250
Michal Papadopulos	CZ	MFK Karviná	2019/20	100
Mads Agesen	DK	Randers FC	2015/16	200
Jeppe Kjaer	DK	Lyngby Boldklub	2016/17	250
Andre Bjerregaard	DK	AC Horsens	2016/17	400
Lasse Kryger	DK	AC Horsens	2016/17	250
Kim Aabech	DK	AC Horsens	2016/17	250
Mohammed Fellah	DK	FC Nordsjaelland	2016/17	200
Mikkel Vendelbo	DK	Silkeborg IF	2016/17	150
Mustapha Bundu	DK	Aarhus GF	2016/17	300
Quincy Antipas	DK	Hobro IK	2017/18	200
Mikael Antonsson	DK	FC Copenhagen	2017/18	250
Morten Hegaard	DK	FC Helsingor	2017/18	100
Julian Kristoffersen	DK	Hobro IK	2018/19	200
Adam Jakobsen	DK	Vejle Boldklub	2018/19	100

Consecutively, the data were subjected to cluster analysis. First, a cluster analysis of goalkeepers in both evaluated football competitions was performed. As a result, the goalkeepers were divided into three clusters. Figure 1 below presents the observed factors for the goalkeepers and the course of the average values of each set.

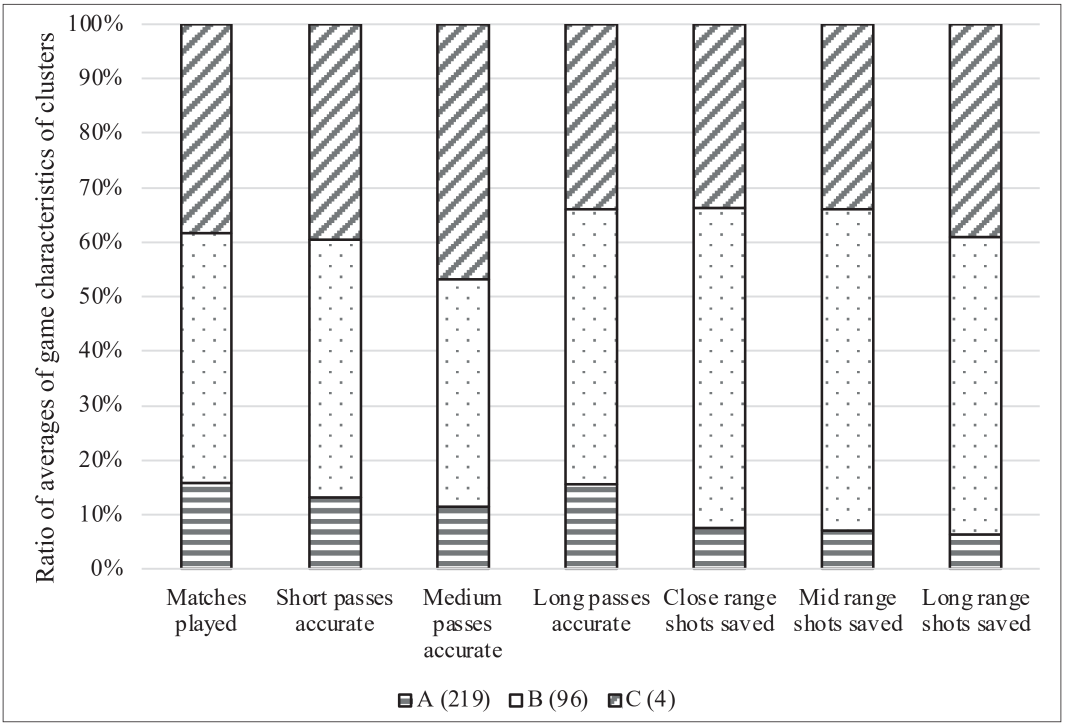

Figure 1. Cluster analysis of goalkeepers in both evaluated football competitions

Figure 1 displays data on the average sports performance of individual groups of goalkeepers. The number of goalkeepers in each group is indicated in brackets next to the group description in the chart legend. The goalkeepers of Cluster A achieved an average market value of 350,000 euros. We have observed that this most numerous group of goalkeepers is, on average, the least sport performing group compared to the other two. The second-largest group of goalkeepers is represented in Cluster B. On average, they reached a market value of 630 thousand euros. This cluster achieved a lower average weight than the most valuable group of players (Cluster C) only in the statistics of accurate medium passes. Goalkeepers Ondřej Kolář, Florin Nita and Kamil Grabara averaged a market value exceeding 4.5 million euros. In all statistics, they reached similar values as the Cluster B goalkeepers. However, based on their high market value, they should dominate the other goalkeepers of Clusters A and B in all statistics.

The same type of analysis was performed for on-field players. The selected characteristics are indicated on the horizontal axis and their frequency on the vertical. In the legend of Figure 2, the number of players in each cluster is stated in brackets. Only the characteristics of the number of games in a season are common to the goalkeeper pool. Clusters A to F are ordered by the average market value of the players. Players in the most numerous Cluster A reached an average market value of 480 thousand euros. In comparison, the players of the least numerous and simultaneously, the most valuable Cluster F reached approximately 8 million euros. Players in Clusters B to E had a market value ranging from 950 thousand to 6 million euros.

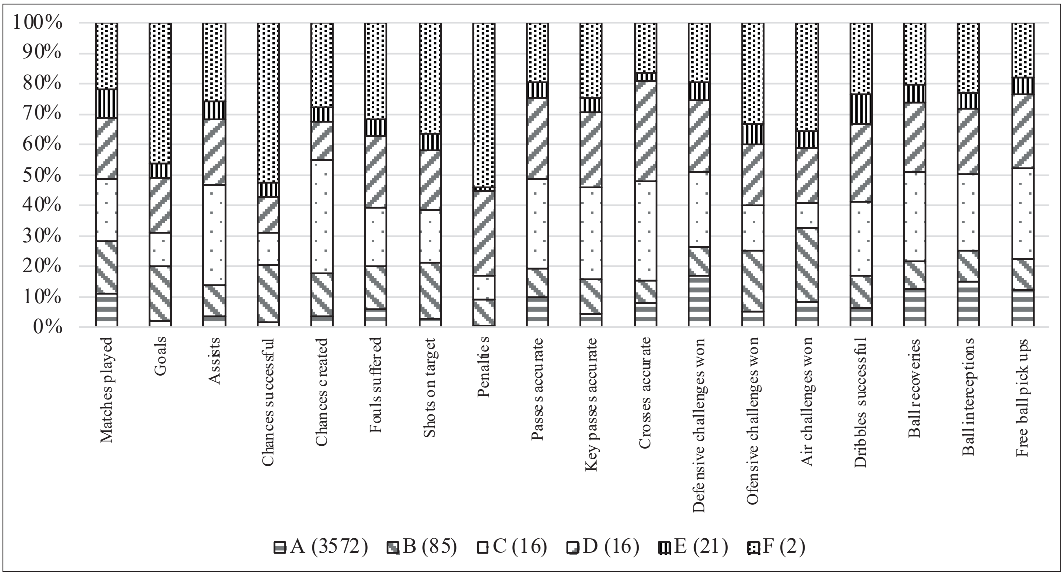

Figure 2. Cluster analysis of on-field players in both evaluated football competitions

The data shows that Cluster A is dominated primarily by defensive players (relatively high average number of defensive challenges won, balls recoveries or opponents' broken passes). They are the players with the lowest average market value among the clusters. More than 95% of all players are incorporated in Cluster A. The other five clusters contain particular groups of players. Cluster B contains offensive players who have, on average, a high number of matches played during the season, an increased number of created chances, shots on target or won offensive challenges. Compared to Clusters C to F, these players are relatively "cheap" in terms of market value, as indicated by the Transfermarkt.com server. Cluster C mainly comprise creative midfielders with a high average value of accurate passes and crosses or created chances. At the same time, the players of Cluster C have achieved a relatively high average value of successful dribbles or free ball pick ups. Players in Cluster D have reached relatively constant average values of the given characteristics compared to other clusters. This Cluster comprises an equal representation of defensive and offensive players. Clusters E and F account for the most valuable players in the group on average. Cluster F contains the most valuable players who have scored the highest number of goals, shots on target, converted penalty kicks, won offensive and air challenges, and successful chances; Cluster E includes players with high market value but low game performance. In most cases, the players in Cluster E do not reach the average values of Cluster A (players with the lowest average market value).

DISCUSSION

The presented research aimed at identifying and subsequent comparing the efficient players of the two competitions under study using data envelopment analysis; by doing so, it was possible to determine whether the most valuable players are also the most efficient. Efficient players were identified by applying the CCR-I model to both football competitions. The number of efficient players at each game position was similar in both leagues, except for one. In the case of the Danish 3F Superliga defenders, the CCR-I model, despite the smaller number of evaluated players, marked almost twice the number of efficient players compared to the Czech Fortuna:Liga. The highest number of efficient players in both football competitions was among the forwards (31 in total). This result is in line with the widely held view that forwards are the most efficient players. This result contrasts with the result published by Tiedemann et al. (2011) and Fernández et al. (2020). These authors believe that the most efficient players should be among the defenders. Fernández et al. (2020) believes that the attack actions are initiated in to defense, continue with the midfield and just them in the opposite field. Therefore, the inefficiencies of each positions are accumulated towards the more advanced positions. Hence, the inefficiency of the defenders is moved to midfielders, which along with own inefficiency is moved to forwards. The present research also supported some correlation with this opinion. In the case of defenders, 26 players were identified as efficient by applying the CCR-I model. It represents the second-highest number of efficient units.

The most valuable players in each position were not assigned to the group of efficient players in either the Czech Fortuna:Liga or the Danish 3F Superliga, as confirmed by Alp (2006). Therefore, this can imply that the most expensive players in both competitions should have better individual game stats. The CCR-I model provides, among other things, model values of the output characteristics inefficient players should achieve in order to become efficient. Research demonstrated that the Danish league's market value of efficient defenders and midfielders was at a higher financial level than in the case of Czech defenders and midfielders. In the case of the market value of goalkeepers and forwards, there are no significant differences between the Danish and Czech leagues.

The second objective was to apply the cluster analysis to divide the player pool into compact clusters and, consecutively, determine the characteristics of each cluster. This approach can be used as a basis for further research on individual players' game positions and performance, as confirmed by studies by Mulitz (2015) and Martineau (2022). In contrast to this research, Mulitz (2015) uses cluster analysis to evaluate defenders entering the National Football League draft. The clusters are created based on the data collected at the training camp. The cluster analysis results in three defender clusters that divide defenders into offensive defenders, defensive defenders and central defenders. Martineau (2022) applied clustering to football player data in 2016 and identified nine clusters, e.g. defensive technical players, natural goalscorers, players proficient in the aerial game but less skilled during other game phases, players characterized by an ability to accelerate and to keep the ball, versatile midfield players, physical defenders etc. This research also arrived at a similar breakdown by player skill.

CONCLUSION

The main objective of the research was to outline an integrated approach to analysing and evaluating the best players in two selected European football competitions over five consecutive seasons. Although player selection is an important decision-making problem, researchers have paid little attention to this area. The main reason for conducting this research was to expand the research literature by furthering the approach for player evaluation. A two-phase method was proposed in the paper. First, the CCR-I model was used to determine the best players. The DEA methodology has the advantage of establishing benchmarks for inefficient DMUs and identifying sources of inefficiency. Cluster analysis was then used to evaluate the players, which allowed to divide the players into compact clusters. Subsequently, cluster analysis helped to determine what the clusters revealed about the players.

The methods described above focused on analysing players from two selected football competitions – the Czech Fortuna:Liga and the Danish 3F Superliga. The players of both leagues under study were arranged according to their game positions, and the most relevant game factors were determined for the individual groups. Forwards had the highest number of efficient players in both competitions. The most valuable players at the given positions were not included in the efficient players' pool. It can be interpreted that the most expensive players in both competitions should, according to the input-oriented CCR model, achieve higher performance as measured by the mentioned game statistics.

The obtained data were subjected to cluster analysis in the second part of the research. After cluster analysis, the goalkeepers' pool was divided into three clusters, with Cluster A containing more than two-thirds of all players at this position. The average market value of the goalkeepers in the first cluster was the lowest compared to the other two. Cluster B included relatively more valuable goalkeepers than Cluster A and, at the same time, the most efficient in terms of given game statistics. The last cluster, C, represented the goalkeepers with the highest market value who did not achieve the corresponding values of the game statistics. The cluster analysis was also applied to the in-field players in the second step. Cluster A contained more than 95% of players, primarily defensive players (defenders and defensive midfielders); the players in the group can be characterised as cheap and reasonably efficient. Cluster B included combat offensive players with a relatively high average number of shots per goal per season. Cluster C involved players with high average values of accurate passes and crosses. Cluster D players achieved consistent and relatively high average values for all given characteristics. In this perspective, they constitute versatile players. Cluster E players can be characterized as "overpriced". Their average market value is high, not matching game statistics. On the other hand, players from Cluster F attained even higher market values than those from Cluster E; however, their average values of individual statistics were relatively high, and they achieved the best results in some of them (e.g., number of goals scored, successful chances, successful penalty kicks, etc.).

Given football clubs' current economic and financial situation, it is increasingly important to know how efficiently a club uses its resources. Efficiency analysis is used to calculate the performance scores of the players and also to determine the lack of aspects and the amount of lack of the inefficient players. The proposed approach is appropriate when there is a large amount of data with different criteria and alternatives. The inputs and outputs can be changed according to the research needs, allowing managers or coaches to find out additional player characteristics. The DEA methodology has the advantage of establishing benchmarks for inefficient DMUs and identifying sources of inefficiency. The methodology and empirical approach adopted and drafted in this paper can illustrate the potential of individual player efficiency assessment as a valuable tool to support decision-making in team sports management. The results obtained from the model show that scientific decision support can bring positive returns to club owners, managers and the whole team and facilitate its success. Future studies may consider multi-period models (such as WDEA or Malmquist index) or applying the proposed approach to player selection in other team sports.

References

3F Superliga. (2021). *Superliga.dk: Performance Center*. Retrieved 15 March 2021. Available from https://www.superliga.dk/performance-center-superliga-2020-2021

Aggarwal, C. C., & Reddy, C. K. (2014). *Data Clustering: Algorithms and Applications*. Boca Raton: CRC Press.

Akhanli, S. E. (2019). *Distance construction and clustering of football player performance data*. (Doctoral thesis). University College London. Retrieved from https://discovery.ucl.ac.uk/id/eprint/10065964/

Alp, I. (2006). Performance evaluation of goalkeepers of the world cup. G.U. *Journal of Science*, *19*(2), 119–125. https://dergipark.org.tr/tr/download/article-file/83094

Arabzad, S. M., Mazaher G., & Shahin, A. (2013). Ranking players by DEA the case of English Premier League. *International Journal of Industrial and Systems Engineering*, *15*(4), 443–461. https://doi.org/10.1504/IJISE.2013.057479

Banker R. D., Charnes, A., & Cooper, W. W. (1984). Some Models for Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, *30*(9), 1078–1092. https://doi.org/10.1287/mnsc.30.9.1078

Charnes A., Cooper, W. W., & Rhodes, E. L. (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operational Research*, *2*(6), 429–444. https://doi.org/10.1016/0377-2217(78)90138-8

Chen, W. C., & Johnson, A. L. (2010). The dynamics of performance space of major league baseball pitchers 1871–2006. *Annals of Operations Research*, *181*(1), 287–302. https://doi.org/10.1007/s10479-010-0743-9

Chitnis, A., & Vaidya, O. (2014). Performance assessment of tennis players: Application of DEA. *Procedia – Social and Behavioral Sciences*, *133*(2014), 74–83. https://doi.org/10.1016/j.sbspro.2014.04.171

Cooper, W. W., Ruiz, J. L., & Sirvent, I. (2009). Selecting non-zero weights to evaluate effectiveness of basketball players with DEA. *European Journal of Operational Research*, *195*(2), 563–574. https://doi.org/10.1016/j.ejor.2008.02.012

Cooper, W. W., Lawrence, M. S., & Zhu, J. (2011). *Handbook on Data Envelopment Analysis*. Berlin: Springer Science and Business Media.

Djordjević, D. P., Vujošević, M., & Martić, M. (2015). Measuring efficiency of football teams by multi-stage DEA model. *Tehnički vjesnik*, *22*(3), 763–770. https://doi.org/10.17559/TV-20140306134047

Erickson, H. D., & Callum, R. (2004). Who's the Best? Data Envelopment Analysis and Ranking Players in the National Football League. In *Economics, Management, and Optimization in Sports*, Butenko et al.(Ed.), 15–30. New York: Springer-Verlag. https://doi.org/10.1007/978-3-540-24734-0_2

Espitia-Escue, M., & García-Cebrián, L. I. (2004). Masuring the Efficiency of Spanish First-Division Soccer Teams. *Journal of Sports Economics*, *5*(4), 329–346. https://doi.org/10.1177/1527002503258047

Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society*, *120*(3), 253–290. https://doi.org/10.2307/2343100

Fernández, R. C., Núñez, T. G., & Sala-Garrido, R. (2020). Analysis of the Efficiency of Spanish Soccer League Players (2009/10) *Using the Metafrontier Approach. Studies of Applied Economics*, *30*(2), 565. https://doi.org/10.25115/eea.v30i2.3541

Fortuna:Liga. (2021). *FortunaLiga.cz: Statistiky*. Retrieved 10 March 2021. Available from https://www.fortunaliga.cz/statistiky?unit=1&parameter=1

Fried, H. O., Lambrinos, J., & Tyner, J. (2004). Evaluating the performance of professional golfers on the PGA, LPGA and SPGA tours. *European Journal of Operational Research*, *154*(2), 548–561. https://doi.org/10.1016/S0377-2217(03)00188-7

Guzmán-Raja, I., & Guzmán-Raja, M. (2021). Measuring the Efficiency of Football Clubs Using Data En-velopment Analysis: Empirical Evidence From Spanish Professional Football. *SAGE Open*, *11*(1), 1–13. https://doi.org/10.1177/2158244021989257

Guzmán, I., & Morrow, S. (2007). Measuring efficiency and productivity in professional football teams: Evidence from the English Premier League. *Central European Journal of Operations Research*, *15*(4), 309–328. https://doi.org/10.1007/s10100-007-0034-y

Haas, D. J. (2003). Technical efficiency in the major league soccer. *Journal of Sport Economics*, *4*(3), 203–215. http://dx.doi.org/10.1177/1527002503252144

Halkos, G., & Tzeremes, N. (2013). A two-stage double bootstrap DEA: the case of the top 25 European football clubs' efficiency levels. *Managerial and Decision Economics*, *34*(2), 108–115. https://doi.org/10.1002/mde.2597

Hirotsu, N., Iguchi, Y., Yoshimura, M. (2018). Development of Simple Indicators to Evaluate Performance of Soccer Player by Use of Data Envelopment Analysis. *Juntendo Sports and Health Science Research*, *9*(2), 53–63.

Hirotsu, N., Osawa, K., Aoba, Y., & Yoshimura, M. (2012). A DEA Approach to Evaluating Characteristics of J-League Players in terms of Time played and Player Similarity. *Football Science*, *13*, 9–25. https://www.shobix.co.jp/jssf/tempfiles/journal/2016/110.pdf

InStat. (2020). *Objective performance rating for players & teams*. Retrieved 23 March 2021. Available from https://instatsport.com/football

Jablonský, J., & Dlouhý, M. (2004). *Modely hodnocení efektivnosti produkčních jednotek*. Praha: Professional Publishing.

Jones, J. C. H. (1969). The Economics of the National Hockey League. *Canadian Journal of Economics*, *2*(1), 1–20. https://doi.org/10.2307/133568

Martineau, D. (2022). *Ai applied to football: Uncover the most promising talent*. Sia Partners – Global Management Firm. Retrieved 5 June 2022. Available from https://www.sia-partners.com/en/news-and-publications/from-our-experts/ai-applied-football-uncover-most-promising-talent

Mulitz, M. I. (2015). *Rigorous Cluster Analyses For Prospective Player Evaluation In The National Football League*. (Thesis). Brown University. Retrieved from https://sites.wustl.edu/francescodiplinio/mentoring-outreach/

Nasiri, M. M., Ranjbar, M., Tavana, M., Santos Arteaga, F. J., & Yazdanparast, R. (2018). A novel hybrid method for selecting soccer players during the transfer season. *Expert Systems*, *36*(1), 1–19. https://doi.org/10.1111/exsy.12342

Open source DEA. (2021). *Introducing OSDEA GUI*. Retrieved 27 January 2021. Available from https://opensourcedea.org/osdea-gui/

Papahristodoulou, C. (2007). *The relative efficiency of UEFA Champions League scorers*. MPRA Paper Number 4943. School of Business, Mälardalen Univerity.

Pyatunin, A. V., Vishnyakova, A. B., Sherstneva, N. L., Mironova, S. P., Dneprov, S. A., & Grabozdin, Y. P. (2016). The economic efficiency of European football clubs – Data Envelopment Analysis (DEA) approach. *International Journal of Environmental and Science Education*, *11*(15), 7515–7534. http://www.ijese.net/makale/1045.html

Ribeiro, A. S., & Lima, F. (2012). Portuguese football league efficiency and players' wages. *Applied Economics Letters*, *19*(6), 599–602. https://doi.org/10.1080/13504851.2011.591719

Rottenberg, S. (1956). The baseball player's labor-market. *Journal of Political Economy*, *64*(3), 242–258. http://dx.doi.org/10.1086/257790

Ruiz, J. L., Pastor, D., & Pastor, J. T. (2013). Assessing Professional Tennis Players Using Data Envelopment Analysis (DEA). *Journal of Sports Economics*, *14*(3) 276–302. https://doi.org/10.1177/1527002511421952.

Sloane, P. J. (1969). The labour market in professional football. *British Journal of Industrial Relations*, *7*(2), 181–199. https://doi.org/10.1111/j.1467-8543.1969.tb00560.x

Singh, P., & Lamba, P. S. (2019). Influence of crowdsourcing, popularity and previous year statistics in market value estimation of football players. *Journal of Discrete Mathematical Sciences and Cryptography*, *22*(2), 113–126. https://doi.org/10.1080/09720529.2019.1576333

Statgraphics Technologies. (2021). *Download statgraphics centurion 18*. Retrieved 27 January 2021. Available from https://www.statgraphics.com/download18

Tiedemann, T., Francksen, T., & Latacz-Lohmann, U. (2011) Assessing the performance of German Bundesliga football players: a non-parametric metafrontier approach. *Central European Journal of Operations Research*, *19*(4), 571–587. https://doi.org/10.1007/s10100-010-0146-7

Transfermarkt. (2021). *Transfermarkt.com: Intern*. Retrieved 25 February 2021. Available from https://www.transfermarkt.com/intern/stellenangebote

Zhu, J. (2014). *Quantitative Models for Performance Evaluation and Benchmarking: Data Envelopment Analysis with Spreadsheets*. Cham: Springer.

Contact Information

Natalie Pelloneová, email: natalie.pelloneova@tul.cz, tel.: +420 775 976 164
Michal Tomíček, email: michal.tomicek@tul.cz, tel.: +420 485 352 242