

Measuring the Efficiency of English Football Clubs: Empirical Evidence from Professional Football

Michal Tomíček¹

Abstract. The aim of the paper is to propose a method for evaluating the performance of football clubs based on the method of Data Envelopment Analysis. The article also deals with the impact of the reprisals caused by the Covid-19 pandemic, which limited fans' access to the stadiums as well as the football competitions themselves. Professional English football clubs playing in the English Premier League were selected for empirical analysis. The performance evaluation of English football clubs is an important reference for team managers, the management of individual clubs and the selected competition as a whole. One of the two most frequently used DEA models, the CCR model, was used to analyze the relative efficiency of football clubs. The study focuses on clubs operating in the highest English football competition during the seasons 2017/18 to 2020/21.

Keywords: Data Envelopment Analysis, European Football, Efficiency, English Premier League

JEL Classification: C10, L83, C67, C44

AMS Classification: 90B90, 90C90

1 Introduction

The world is slowly returning to normal, and sport is no exception. Especially in Europe, football is at the forefront of popular sports. Fans from Lisbon in the west to Turkey in the east were looking forward to relaunching their national football competitions. The same is true of the so-called cradle of football, England.

Sports organizations try to evaluate its performance: its weaknesses and strengths. Currently, success in the professional football league is linked to successful training and management of the entire team. Efficiency, however, goes beyond this kind of corporate-level view. That is where the board of directors and, crucially, the owner are willing to see the club's performance charts and curves rise. In the business world, comparison with competitors is important to objectively evaluate the direction of the company. One of the most popular methods for this kind of benchmark analysis is Data Envelopment Analysis (DEA).

The presented paper focuses on measuring the performance of clubs in the English Premier League, officially The Football Association Premier League Limited, the top football competition in England. The period under study is the 2017/18 to 2020/21 seasons, these are four consecutive seasons. The comparison covers the seasons before, during and after the Covid-19 restrictions mainly on the number of spectators in the stadiums and the closure of teams to the so-called closed bubbles.

A specific feature of European football competitions is the closed nature of the different competition levels. Football clubs are not free to move between levels, but unlike professional sports competitions in the United States which are usually closed for relegation or promotion, there are clear rules for doing so at the end of the season based on position in the table. Newly formed clubs cannot immediately join the top competitions, they have to go through a long path from the lowest competitions to the professional leagues. Therefore, it is often more profitable for investors to buy an already established club with a history, a fan base and sporting and personnel facilities than to start a club from scratch. At the same time, a single investor may not own more than one football club participating in European cups. This is a safeguard against the oligopolisation of professional European football.

The performance of sports companies has been addressed by a number of authors. Most of them use only sports statistics to assess club efficiency, while the little rest combine sports and economic indicators to achieve more accurate efficiency results.

Palafox-Alcantar and Vargas-Hernández [9] measure the payroll efficiency of 32 National Football League (NFL) teams in the 2014 season using data wrapper analysis. When analyzing American sporting competitions, time

¹ Technical University of Liberec, Faculty of Economics, Studentská 2, Liberec, michal.tomicek@tul.cz.

series can be better used as the set of units of observation is almost constant over time. However, the paper was focused on analyzing the one season only.

In their research, Barros and Douvis [2] estimate productivity changes using DEA analysis applied to a representative sample of football clubs operating in two South European countries: Greece and Portugal. They rank football clubs according to their productivity changes between the 1999/20 and 2002/03 seasons. They conclude that some clubs experienced productivity growth while others experienced productivity decline. In the evaluation of the conclusions of the work, they mainly mention the clubs that participated in all the seasons studied.

Petrovic Djordjevic [10] uses a non-parametric output-oriented DEA model and analyzes the technical efficiency of national football teams in the 2010 World Cup qualification. The DEA model has a two-stage structure. In the first stage inputs are used to produce outputs which then become inputs to the second stage.

Halkos and Tzeremes [5] use the DEA method to compare the actual level of market value of football clubs and their performance. The research shows that the level of market value of football clubs has a negative effect on their performance, more accurately that high value of football clubs does not guarantee higher performance.

Arabzad et al. [1] use the DEA model to identify the best footballers in the English Premier League. A different type of DEA model is used to rank selected players. The proposed approach is investigated in the English Premier League 2010/11 season. The findings show that players Wayne Rooney, Didier Drogba and Carlos Tévez are ranked first, second and third respectively.

Jardin [7] evaluates the efficiency of French football clubs between 2004 and 2007 using DEA. Then the Malmquist indexes are used for the studies of the dynamics of clubs' performance. The first source of inefficiency in the Ligue 1 is linked to size problems and over-investments. The best teams in competition or most profitable clubs are not the most efficient units in the sample. The Kang's study [8] measured the relative efficiency and productivity change of Korean professional sports teams using the DEA model and the Malmquist Index for 2006-2009 with similar results as Jardin's [7].

2 Methods applied

In the main part of the paper the method of DEA, a quantitative non-parametric method, is used. Basic DEA models are divided into input-oriented and output-oriented models. Using input-oriented models, it is possible to estimate the degree of technical efficiency, which determines the reduction of input indicators, so that the unit becomes technically efficient with unchanged output. Efficient units get a score of 1, inefficient ones get a lower score in the interval (0; 1) [3; 4].

The CCR (Charnes, Cooper and Rhodes) model considers constant returns to scale and estimates Overall Technical Efficiency (OTE), which consists of two parts, Pure Technical Efficiency (PTE) and Scale Efficiency (SE). Scale efficiency is then determined by the ratio of OTE and PTE and indicates the extent to which the unit can improve its efficiency by changing its size [3]. Together with the BCC (Banker, Charnes and Cooper) model, the CCR model forms two basic (mathematically simplest) DEA models [6].

The OSDEA-GUI (Open-source DEA Graphic User Interface) software was used for all the calculations related to DEA. The data for each season were entered into the software separately in the form of csv files with inputs and outputs in separate columns. The efficiency limit defines the maximum output combinations that can be selected for a given set of inputs. Assuming a set of n Decision making units ($DMUs$), each DMU_j ($j = 1, \dots, n$) uses m inputs x_{ij} ($i = 1, 2, \dots, m$) to produce s outputs y_{rj} ($r = 1, 2, \dots, s$). Input-oriented models with constant returns to scale can be formulated according to equations (1, 2) to minimize inputs while keeping outputs at their current level [6].

$$\begin{aligned} & \min \theta_q \\ S. t. & \sum_{j=1}^n \lambda_j x_{ij} + s_i^- = \theta_q x_{iq} \quad i = 1, 2, \dots, m, \end{aligned} \quad (1)$$

$$\sum_{j=1}^n \lambda_j y_{kj} + s_k^+ = y_{kq} \quad k = 1, 2, \dots, r \quad (2)$$

$$\lambda_j \geq 0, \quad j = 1, 2, \dots, n. \theta \text{ unrestricted in sign,}$$

where θ_q represents the technical efficiency of entity q and λ_j represents the associated weighting of outputs and inputs of entity j . To find maximal possible inputs excesses and outputs shortfalls while maintaining θ , the slack variables, s_i^- and s_k^+ , reflect these excesses and shortfalls, respectively.

The overall technical efficiency of the DMU is measured in relation to the other units analyzed using the efficiency score. The overall level of technical efficiency (OTE_{CCR-I}) is taking on values in the range (0,1). Technically efficient DMUs achieve efficiency rates of 1, the efficiency rates of inefficient units are less than 1 [3; 6].

The creation of the database is preceded by the collection of data from several sources. In addition to the publicly available information published by the English Football Association itself, the data used in the article also comes from the private database of InStat, a company that analyses sports data. The economic data was obtained from the specialist Transfermarkt server. The complexity of the database used in the paper lies in the combination of these sources and their supplementation with data from the register of companies doing business in the United Kingdom of Great Britain and Northern Ireland. The research process can be divided into the following phases:

1. creation of the list of evaluated companies;
2. collection of sporting and economic metrics:
 - a) sports data from InStat database of football players and clubs;
 - b) financial data from Transfermarkt server;
 - c) attendance data from official English Premier League database and Transfermarkt database.
3. determining the inputs and outputs of the DEA model;
4. determination of technical efficiency values;
5. comparison of the performance of football clubs.

Given the method used and the available data, the input factor is the market value of the team in the season and the average attendance at the team's home stadium. The total market value is only considered for players who have played at least one league game in a given season. The factor of average stadium attendance was used in the article to represent the external football environment. The average attendance factor contains many characteristics that limit its use and power. In future research, the relative attendance to the maximum stadium capacity could be used instead of the average value.

On the output side of the model, there are more factors: goals scored, challenges won, supersaves and accurate key passes. Outputs of the model represent the game characteristics for the main football roles on the pitch, i.e. goalkeepers, defenders, midfielders and forwards. Challenges won consist of the sum of defensive and offensive challenges and thus represent a universal physical component of the factors used. The output factors include the four different components of a football match, namely the offensive and defensive parts, the goalkeeper factor and the creativity factor. The chosen model therefore examines how efficiently clubs can convert the market value of their players and fan support into positive game characteristics in the form of goals scored, challenges won, supersaves and accurate key passes.

3 Research results

The results of the paper are divided into four parts according to the season studied (2017/18, 2018/19, 2019/20 and 2020/21). This is followed by a summary of the results in the seasons studied and an evaluation of the clubs' efficiency in the seasons. The first table (Table 1) shows position in the table (Pos.) and the value of OTE_{CCR-I} of the clubs that participated in a given seasons. The sports successful clubs, which were ranked in the table up to 8th place, did not reach the efficient OTE_{CCR-I} value even once in the examined seasons.

DMU	2017/18		2018/19		2019/20		2020/21	
	Pos.	OTE_{CCR-I}	Pos.	OTE_{CCR-I}	Pos.	OTE_{CCR-I}	Pos.	OTE_{CCR-I}
MCI	1	0.4526	1	0.5503	2	0.6073	1	0.4113
MAN	2	0.2641	6	0.3177	3	0.4081	2	0.4602
TOT	3	0.2709	4	0.4172	6	0.3701	7	0.5232

LIV	4	0.3655	2	0.5292	1	0.5286	3	0.5521
CHE	5	0.3484	3	0.5348	4	0.6012	4	0.4163
ARS	6	0.3392	5	0.4503	8	0.4147	8	0.4201
BUR	7	0.8226	15	0.9717	10	1.0000	17	1.0000
EVE	8	0.4023	8	0.5142	12	0.4755	10	0.5917
LEI	9	0.5594	9	0.7073	5	0.8343	5	0.5878
NEW	10	0.6325	13	0.8222	13	0.6866	12	0.6627
CRY	11	0.6622	12	0.8020	14	0.8490	14	0.8781
BOU	12	1.0000	14	1.0000	18	1.0000		
WHU	13	0.6942	10	0.4778	16	0.6003	6	0.8190
WAT	14	0.7263	11	1.0000	19	1.0000		
BRI	15	0.9191	17	0.8877	15	0.8067	16	0.7480
HUD	16	1.0000	20	1.0000				
SOU	17	0.4970	16	0.7441	11	0.9585	15	1.0000
SWA	18	0.9109						
STO	19	0.8112						
WBA	20	0.9203					19	1.0000
WOL			7	0.8221	7	0.6988	13	0.9827
CAR			18	1.0000				
FUL			19	0.9294			18	1.0000
SHU					9	1.0000	20	0.9967
AVL					17	0.6571	11	1.0000
NOR					20	1.0000		
LEE							9	1.0000

Table 1 DEA efficiency of Premier League clubs in the seasons 2017/18 to 2020/21

Since the efficient clubs are always ranked 9th to 20th in the table, the result suggests that the team market value of the clubs in the top eight positions in the table is so much higher than the rest of the clubs cannot achieve the efficient OTE_{CCR-I} value. In the seasons under review, the ratio of the average market values of the Premier League clubs ranked in the Top 8 to those ranked 9th to 20th ranged from 2.6 to 3.6 (see Table 2).

Market Value Average (Th. €)	2017/18	2018/19	2019/20	2020/21
Top 8 Clubs	704 550	784 313	681 731	688 588
9 th – 20 th Clubs	199 916	235 071	219 556	258 750
Ratio	3.5242	3.3365	3.1050	2.6612

Table 2 Market value for clubs according to their position in the table

Of the 17 efficient clubs (see Table 1), there have been a total of six occasions when clubs rated as efficient at the end of the season have been relegated from the top flight of English football. These clubs had the lowest budgets and the market value of their team was also the lowest compared to other Premier League clubs. In each season, there has always been one club newly promoted to the top competition that has managed to finish in the top ten and in two cases even achieve an efficient value of OTE_{CCR-I} . In the 2018/19 season it was Wolverhampton Wanderers F.C. in 7th place (0.8221), in the 2019/20 season it was Sheffield United F.C. in 9th place (1.0000) and in the 2020/21 season it was Leeds United F.C. in 9th place (1.0000).

On the other hand, four clubs failed to keep up (sportingly and financially) with the Premier League clubs and immediately after promotion they left the top competition and were relegated back to the EFL Championship (officially English Football League Championship). In doing so, all these clubs achieved an efficient value of OTE_{CCR-I} .

The four most successful clubs in terms of table position in the seasons under review were Manchester City, Liverpool, Manchester United and Chelsea. Of these most successful English clubs in recent times and in European

competitions, only Manchester City achieved an average value of OTE_{CCR-1} of more than 0.5. 14 clubs participated in all four seasons. Of these, the highest average OTE values were achieved by Burnley (0.9486), Brighton (0.8404), Southampton (0.7999) and Crystal Palace (0.7978).

If we look at the season before (2018/19) the covid pandemic and the season that was affected by the covid pandemic (2019/20), then 10 of the 17 clubs that participated in both of these seasons saw an improvement in their OTE_{CCR-1} values compared to the 2018/19 season. This is due to the combined effects of a reduction in average attendance at large English club's stadiums, a cooling of the market for players and therefore their market value, while maintaining a decent sporting performance of the clubs.

4 Conclusion

This paper measured the efficiency of professional football clubs playing in the English Premier League. For this purpose, a time span of four seasons from 2017/18 to 2020/21 was chosen. The main aim of the paper was to propose an approach to the analysis and evaluation of the best football clubs of the English Premier League. The efficiency of football clubs was measured using a non-parametric DEA method. Total market value of the team and average stadium attendance were chosen as inputs of the model. Average stadium attendance was added to standard models used by several other authors. Output was measured by the total number of goals, challenges won, supersaves and accurate key passes. This particular specification proved to be appropriate for this application (adequate number of efficient DMUs), but it can also be used to analyze the efficiency of other team sports.

Given the current economic and financial situation, there is an increasing need to determine how efficiently a club is using its resources and assets – its players. Based on the seasons analyzed, several conclusions emerge from the research: It turns out that the gap between the top clubs and the rest of the league in terms of the input factor of the total market value of a Premier League club's squad combined with average attendances is too great for the wealthy section of clubs who can afford the best and therefore most valuable football players in the world, and cannot outweigh the better sporting performance of these clubs compared to average or below average English clubs. This leads to the fact that the sport's best Premier League clubs were not rated as efficient in any of the seasons monitored by the chosen DEA model.

As other authors, for example Jardin [7], mentioned in their works, the best clubs from the sport point of view are not the most efficient by the DEA model results.

Acknowledgements

Creation of this paper was supported by the Technical University of Liberec grant no. 21456 "Factors influencing the economic performance of European sports enterprises" (SGS-2022-1017).

References

- [1] Arabzad, S. M., Ghorbani, M., & Shahin, A. (2013). *Ranking players by DEA the case of English Premier League*. International Journal of Industrial and Systems Engineering, 15(4), 443.
- [2] Barros, C. P., & Douvis, J. (2009). *Comparative analysis of football efficiency among two small European countries: Portugal and Greece*. International Journal of Sport Management and Marketing, 6(2), 183.
- [3] Cooper, W. W. (2011). *Handbook on data envelopment analysis*. New York: Springer.
- [4] Cooper, W. W., Seiford, L. M. & Tone, K. (2007). *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software*. New York: Springer.
- [5] 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.
- [6] Jablonský, J. & Dlouhý, M. (2015). *Modely hodnocení efektivnosti a alokace zdrojů*. Praha: Professional Publishing.
- [7] Jardin, M. (2009). *Efficiency of French football clubs and its dynamics*. Munich: Personal RePEc.
- [8] Kang, H. J. (2010). *Productivity Change and Relative Efficiency of Professional Sport Teams*. The Journal of the Korea Contents Association. Vol. 10, Issue 10, 456–463.
- [9] Palafox-Alcantar, P. G. & Vargas-Hernández, J. G. (2015). *Measuring the efficiency of the 32 franchises in the NFL during the 2014 season*. Journal of Sports Economics & Management, 5(1), 37–53.
- [10] Petrović Djordjević, D. (2015). *Measuring efficiency of football teams by multi-stage DEA model*. Tehnicki vjesnik-Technical Gazette, 22(3), 763–770.