Measuring the efficiency of football players by DEA model

Natalie Pelloneová[1], Michal Tomíček[2]

Abstract. Football is a very popular subdivision of sports not only in our country, but also all around the world. This article expands the ideas from the economic literature on efficiency to develop method for evaluating the performance of football players that take into account many dimensions of football performance. The evaluation of football players has always been an important input for strategic decisions in the football industry. The aim of this paper is to apply an input-oriented data envelopment analysis (DEA) model in order to measure football players' efficiency and to identify the technically efficient players with regard to their position on the field. Efficient players will be further ranked using the Andersen-Petersen super efficiency model. The model is empirically applied to players of the first and second Slovak football leagues in the 2020/21 season. The model proposed in this paper seeks to incorporate more objectivity into decision-making and can thus be an important step in developing a systematic methodology for evaluating football players.

Keywords: football player, DEA, efficiency, super-efficiency, sport

JEL Classification: C10, L83, C67, C44
AMS Classification: 90B90, 90C90

1 Introduction

Economics has long been successfully extending its sphere of influence into diverse areas. Football has not been left out of its interest. Many authors are concerned with football, whether at the level of players and clubs or at the level of national teams and international tournaments. Identifying talented players and then forming a team is a very difficult task for football club managers. Evaluating player performance is currently a key issue in the sports industry. Player performance evaluation itself is quite a challenging problem. All over the world, football fans try to determine the ranking of players based on their subjective views or on the basis of various key parameters. Player performance varies from position to position, but it is also based on competition, time played and team style of play. Recently, there has been an emphasis on evaluating player performance in football using statistical methods, aided by the availability of a wide range of data.

In football, financial aspects, such as the transfer market and club investment, influence player performance. The aim of every club is to achieve the best possible performance. It is therefore important to identify players who meet the requirements, with the best potential for return on investment. In general, the market price of football players is also linked to their performance. However, the game performance of football players is related to specific technical skills in offensive or defensive activities [2]. In this paper, we propose an approach to measure the performance of football players based solely on the use of data envelopment analysis (DEA). This research aims to measure the impact of the market value of players in the 1st and 2nd Slovak football leagues on their efficiency level using data from the 2020/21 season with respect to playing position and technical skills. The CCR-I model was used to analyse the relative efficiency of football players. Using the AP model, a super-efficiency analysis was further performed to differentiate the efficiency scores between the efficient units and to determine the ranking of the players. Determining the ranking of the best player among goalkeepers, defenders, midfielders and forwards is a difficult task. The research will evaluate each group of players separately and different output variables will be selected for each group of players in the CCR-I model. Among goalkeepers, for example, their successful saves will be evaluated, as well as the interventions of defenders, the assists of midfielders and the goals of attackers.

In the past, the DEA approach has been used to measure efficiency in other contexts. In the last 10 to 15 years, DEA has also been recognised as a common methodology for measuring efficiency in sports such as football, handball, volleyball, etc. For example, Cooper, Ruiz and Sirvent [3] apply the DEA method to the evaluation of basketball players in the Spanish Basketball League. Cooper, Ramón, Ruiz and Sirvent [4] further applied DEA to evaluate basketball players using cross-efficiency evaluation. Suk [13] evaluated the relative efficiency of

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

players in a Korean baseball organization using DEA. His research investigates whether a national baseball team composed of the best players can achieve outstanding results in international competitions. Ramón, Ruiz and Sirvent [11] ranked tennis players using the CCR-O model. Santín [12] uses an output-oriented nonincreasing returns to scale super-efficiency data envelopment analysis model in order to measure the performance of Real Madrid football players. The aim of his research is to identify the historically best players of this famous big club. Papahristodoulou [10] measures the efficiency of several scorers for different football teams.

The present research focuses on Slovak football. In global terms, the popularity of football in Slovakia is quite high. Slovakia has a high number of clubs and players per capita and a relatively long history of football. The number of registered football players in 2008 was up to 428 968, which is about 8% of the population of Slovakia. The highest football competition in Slovakia is called Fortuna Liga and 12 teams participate in it. The second highest Slovak football league is called the 2nd Slovak Football League. A total of 16 teams have participated in the Second Slovak Football League since the 2017/18 season. In the 2020/21 season, 375 players started at least one match in the Fortuna Liga. There were 441 players in the 2nd Slovak League. Measuring the efficiency of these players will be the subject of the research below.

2 Methodology

Data envelopment analysis was used to analyse the relative efficiency of Slovak football players. Data envelopment analysis can be classified as a modern approach to performance evaluation of production units. The first data envelopment analysis models were proposed in the 1970s. The first DEA models were used to evaluate the efficiency of homogeneous production units. The units evaluated can be, for example, institutions, territorial units, enterprises or even athletes as in the case of this paper. Another assumption is that the evaluated units are comparable to each other [5]. Units are evaluated on the basis of inputs used and outputs produced. The advantage of DEA models is the ability to compare multiple inputs and outputs, with no a priori knowledge or weighting of input data required. Due to this assumption, variables with different units of measurement can be included in the model. DEA models are based on the Farrell model, which dealt with the technical efficiency of production units in the 1950s. He was followed up in the 1970s and 1980s by other experts who built the first DEA models. The two basic DEA models are the CCR model by Rhodes, Charnes and Cooper and the BCC model by Banker, Charnes and Cooper [6;7].

The CCR model with constant returns to scale (CRS) is based on the assumption that the same change in inputs causes the same change in outputs. In contrast, the BCC model is considered under the assumption of variable returns to scale (VRS). The models can be further divided into input-oriented and output-oriented models. Using an input-oriented model, it is possible to determine what the quantity of inputs should be to make an inefficient unit efficient. A unit with a technical efficiency coefficient equal to 1 is efficient, a coefficient less than 1 indicates an inefficient unit and determines the degree of input reduction needed to make the unit efficient. Using the output-oriented model, it is possible to determine the quantity of outputs that should be produced to make an inefficient unit efficient. A unit with a technical efficiency coefficient equal to 1 is efficient, a coefficient greater than 1 indicates an inefficient unit and determines the degree of increase in outputs needed to make the unit efficient [7;9].

The CCR-I model is the first model that will be used to analyse the efficiency of Slovak football players. This model assumes constant returns to scale [9] and can be written mathematically by relations (1) and (2). Where λ_j, $j = 1, 2, ..., n$ are weights of all DMUs, s_i^-, $i = 1, 2, ..., m$ and s_k^+, $k = 1, 2, ..., r$ are slack/surplus variables, θ_q is the efficiency score of the DMU_q.

$$Minimize \qquad \theta_q \tag{1}$$

$$S.t. \qquad \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = \theta_q x_{iq}, \qquad i = 1,2,\dots,m,$$

$$\sum_{j=1}^{n} y_{kj}\lambda_j - s_k^+ = y_{kq}, \qquad k = 1,2,\dots,r, \tag{2}$$

$$\lambda_j \geq 0, \quad j = 1,2,\dots,n,$$

Standard DEA models have a large number of applications and modifications. One of the most important extensions of the DEA model is the formulation of super-efficiency models, which are used to determine the ranking of DMUs with a single efficiency score. The best known model is that of Andersen and Petersen (1993). This

model allows an efficient unit to achieve an efficiency greater than 1 for input-oriented models or less than 1 for output-oriented models. The whole concept of super-efficiency is based on extracting a specific efficient production unit from the set under consideration, thus shifting the original efficient frontier. Their main advantage is the possibility to further classify the efficient production units [1]. This is because the arrangement of efficient units is not offered by conventional DEA models. Jablonsky and Dlouhý [9] formulate an input-oriented AP model under the CRS assumption by using relations (3) and (4). If the unit under consideration is identified as efficient, then $\theta_q^{AP} > 1$.

$$Minimize \qquad \theta_q^{AP} \tag{3}$$

$$S.t. \qquad \sum_{j=1}^{n} x_{ij}\lambda_j + s_i^- = \theta_q^{AP} x_{iq}, \quad i = 1, \dots, m,$$

$$\sum_{j=1}^{n} y_{kj}\lambda_j - s_k^+ = y_{kq}, \quad k = 1, \dots, r, \tag{4}$$

$$\lambda_j \geq 0, \quad j = 1, \dots, n, j \neq q,$$

$$\lambda_q = 0.$$

3 Data and variables

Transfermarkt.com and the commercial database InStat were used as data sources. These sources were supplemented with data from Livesport.cz. Only football players who played at least one match in the 2020/21 season were included in the analysis. The database consists of 375 footballers, including 47 forwards, 179 midfielders, 117 defenders and 32 goalkeepers, who played in the Fortuna Liga in the 2020/21 season. Furthermore, the database consists of 441 football players, including 76 forwards, 166 midfielders, 152 defenders and 47 goalkeepers, who played in the 2nd Slovak Football League in the 2020/21 season.

The only input is the market value of each player (further MV). On the other hand, the research will consider several outputs depending on the playing position of the players. The first common output is the number of minutes played. The importance of a player in a team increases every time he is selected to play. This variable takes into account whether the player played the whole game or came on as a substitute. For the group of goalkeepers, the following output variables were included in the research: close range shots saved, mid range shots saved, long range shots saved, stopped shots, supersaves and accurate passes. For the group of defenders, the following output variables were included in the research: goals, assists, chances created, fouls suffered, successful actions, blocked shots, accurate passes, crosses, defensive challenges won, dribbles successful, ball interceptions, free ball pick ups, ball recoveries. For the group of midfielders, the following output variables were included in the research: goals, assists, chances created, fouls suffered, successful actions, shots on target, accurate passes, crosses, challenges won, dribbles, ball interceptions, free ball pick ups and ball recoveries. For the group of forwards, the following output variables were included in the research: goals, assists, chances created, successful actions, shots on target, accurate passes, crosses and attacking challenges won.

4 Research results

The results of the research are divided into four groups of players according to their position on the field: forwards, midfielders, defenders and goalkeepers. Two variants of the DEA model, the CCR-I model and the AP super-efficiency model, were applied to each group of players. Thanks to the super-efficiency model, the ranking of players can be determined.

The CCR-I model identified a total of 4 forwards as efficient out of a total of 47 forwards in the 1st league (see Table 1). None of the efficient forwards reached the highest market value. These were forwards of average or below average market value. In terms of goals scored, with the exception of Erik Jendrisek, they scored an above average number of goals. Tomas Malec scored the most goals (8 total) of the efficient forwards. The CCR-I model identified a total of 5 forwards out of a total of 76 forwards in the 2nd league as efficient (see Table 1). On average, forwards in the 2nd league needed less minutes played (1658) to be efficient than forwards in the 1st league (1950). On the other hand, efficient forwards in the 2nd league scored more goals on average than efficient forwards in the 1st league. Lukas Gasparovic scored the most goals (11 total) of the efficient forwards. Efficient forwards in the 2nd league also had on average more assists, successful actions, accurate passes and

crosses than forwards in the 1st league. The market value of the efficient forwards of the 2nd league was at a lower financial level.

Name	League	Club	MV (th. EUR)	CCR-I	AP	Ranking
Tomas Malec	1st	Senica	250	1.0000	1.8524	1.
Matej Trusa	1st	Michalovce	200	1.0000	1.4896	2.
Milos Lacny	1st	Sered	150	1.0000	1.4583	3.
Erik Jendrisek	1st	Nitra	200	1.0000	1.1436	4.
Lukas Gasparovic	2nd	Petrzalka	100	1.0000	1.4805	1.
Gabor Toth	2nd	Komarno	50	1.0000	1.4676	2.
Patrik Rumansky	2nd	Poprad	25	1.0000	1.3197	3.
Lutfi Biljali	2nd	Partizan Bardejov	50	1.0000	1.2341	4.
Marek Kuzma	2nd	Dubnica nad Vahom	100	1.0000	1.0995	5.

Table 1 Efficient forwards in the 1st and 2nd Slovak football league

The second group consisted of midfielders. From the perspective of the CCR-I model, a total of 4 midfielders were identified as efficient out of a total of 179 midfielders (see Table 2). None of the efficient midfielders reached the highest market value. These were midfielders of below average market value. In the case of the 2nd league, a set of 166 midfielders was analysed. In terms of the CCR-I model, a total of 8 players were identified as efficient (see Table 2). On average, midfielders in the 2nd league needed smaller values of the outcome variables to be efficient than midfielders in the 1st league. The market value of efficient midfielders in the 2nd league was also in the lower financial level.

Name	League	Club	MV (th. EUR)	CCR-I	AP	Ranking
Igor Zofcak	1st	Michalovce	100	1.0000	3.0000	1.
Juraj Piroska	1st	Senica	100	1.0000	1.9041	2.
Alieu Fadera	1st	Pohronie	300	1.0000	1.2314	3.
Dimitris Popovits	1st	Michalovce	200	1.0000	1.0918	4.
Matej Rosenberger	2nd	Slovan Bratislava	25	1.0000	2.1398	1.
Guytho Mijland	2nd	Bardejov	25	1.0000	1.6951	2.
Sebastian Gembicky	2nd	Petrzalka	25	1.0000	1.6250	3.
Dominik Malinak	2nd	Poprad	25	1.0000	1.5860	4.
Vladimir Bajtos	2nd	Trebisov	25	1.0000	1.3390	5.
Filip Szetei	2nd	Komarno	25	1.0000	1.0832	6.
Michal Petras	2nd	Dubnica nad Vahom	50	1.0000	1.0658	7.
Wisdom Uda Kanu	2nd	Trebisov	125	1.0000	1.0500	8.

Table 2 Efficient midfielders in the 1st and 2nd Slovak football league

In the next part of the research the CCR-I and AP model was applied to a set of 117 defenders of the 1st league. Approximately 8% of the defenders were considered efficient in terms of the CCR-I model. These were defenders who played on average around 2000 minutes. In the 2nd league, out of a total of 152 defenders, a total of 13 defenders were marked as efficient. On average, defenders in the 2nd league needed less minutes played to be efficient than defenders in the 1st league. On the other hand, efficient defenders in the 2nd league scored more goals on average. For the other outcome variables, the 2nd league defenders achieved similar or lower values. The market value of efficient defenders in the 2nd league was also in a lower financial level than in the 1st league (see Table 3).

Name	League	Club	MV (th. EUR)	CCR-I	AP	Ranking
Peter Mazan	1st	Pohronie	100	1.0000	2.9980	1.
Lubomir Michalik	1st	Sered	100	1.0000	1.5408	2.
Martin Chren	1st	Zlate Moravce	75	1.0000	1.3853	3.

Tomas Hucko	1st	Sered	100	1.0000	1.3549	4.
Bernard Petrak	1st	Pohronie	150	1.0000	1.3360	5.
C. Blackman	1st	Dunajská Streda	500	1.0000	1.1579	6.
Martin Toth	1st	Zlate Moravce	100	1.0000	1.1044	7.
Jan Maslo	1st	Ruzomberok	200	1.0000	1.0947	8.
J. Mendez	1st	Dunajská Streda	100	1.0000	1.0702	9.
Vladimir Kukol	2nd	Podbrezova	50	1.0000	2.0241	1.
Jakub Balaz	2nd	Slovan Bratislava	25	1.0000	2.0000	2.
Patrik Leitner	2nd	Zilina II	25	1.0000	1.5271	3.
Michal Ranko	2nd	Skalica	175	1.0000	1.4286	4.
Juraj Lacko	2nd	Puchov, Slovan Bratislava	25	1.0000	1.2933	5.
Peter Vojtovic	2nd	Trebisov	25	1.0000	1.2916	6.
Juraj Martincek	2nd	Puchov	75	1.0000	1.1525	7.
Denis Knizka	2nd	Dubnica nad Vahom	50	1.0000	1.1460	8.
Lukas Simko	2nd	Trebisov	50	1.0000	1.1327	9.
Jakub Parkan	2nd	Dubnica nad Vahom	50	1.0000	1.1071	10.
Peter Vosko	2nd	Liptovsky Mikulas	125	1.0000	1.0936	11.
David Kocik	2nd	Poprad	50	1.0000	1.0697	12.
Marek Frimmel	2nd	Banska Bystrica	125	1.0000	1.0587	13.

Table 3 Efficient defenders in the 1st and 2nd Slovak football league

Finally, the CCR-I model and the AP model were applied to the set of 32 goalkeepers of the 1st league. The goalkeepers who emerged from the analysis as efficient are listed in Table 4. These are 5 goalkeepers. It is typical for all three goalkeepers to have above-average values in the variables related to saves. For example, Adrian Chovan was the best goalie in terms of the variables mid-range shots saved and super-saves. The highest market value was attributed to Dominik Greif (1 million euros). However, Dominik Greif has a very low OTE score (0.197) and can be considered inefficient according to the CCR-I model. Table 4 also shows the efficient goalkeepers of League 2. From the set of 47 goalkeepers, the goalkeepers Matej Luksch and Milan Vincler were identified as efficient from the perspective of the CCR-I model. These goalkeepers were above average on all outcome variables. Matej Luksch recorded maximum shots saved, mid and long range shots saved, stopped shots and accurate passes. Thanks to these great stats, he ultimately achieved an efficient score. It was a goalkeeper with a relatively low market value (75 thousand euros). On average, goalkeepers in the 2nd league had to play more minutes to be efficient than goalkeepers in the 1st league. Efficient goalkeepers in the 2nd league also had more saves on average, except for supersaves, and also more accurate passes. The market value of efficient goalkeepers in the 2nd league was also in a lower financial level than in the 1st league.

Name	**League**	**Club**	**MV (th. EUR)**	**CCR-I**	**AP**	**Ranking**
Tomas Frystak	1st	Senica	150	1.0000	1.6087	1.
Adrian Chovan	1st	Zlate Moravce	300	1.0000	1.1917	2.
Tomas Jenco	1st	Pohronie	100	1.0000	1.1828	3.
Matej Markovic	1st	Michalovce	150	1.0000	1.1781	4.
Igor Semrinec	1st	Trencin	150	1.0000	1.0093	5.
Matej Luksch	2nd	Liptovsky Mikulas	75	1.0000	1.4427	1.
Milan Vincler	2nd	Trebisov	50	1.0000	1.4403	2.

Table 4 Efficient goalkeepers in the 1st and 2nd Slovak football league

5 Conclusion

The main aim of the paper was to propose an approach for the analysis and evaluation of the best players in two Slovak football competitions. The CCR-I model was used to select the best players. Then the AP super-efficiency model was used to determine the ranking of the best players. The method of data envelopment analy-

sis focused on the analysis of players of two selected football competitions - 1st and 2nd Slovak football leagues. Players were divided according to playing positions on the field and the most appropriate game factors were selected for each group.

The highest number of efficient players in both competitions was found for the defenders. This result is consistent with those published by Tiedemann et al. [14] and Fernández et al. [8]. These authors believe that the largest number of efficient players should be among the defenders. Fernandez et al. [8] believes that the attack actions are initiated in to defense, continue with the midfield and just them in the opposite field. In neither case were the most valuable players of the given positions assigned to the set of efficient players. This fact can be interpreted in such a way that the most expensive players in both competitions should, according to the CCR-I model, achieve higher performance as measured by the aforementioned game statistics. The research also showed that the market value of the efficient players of the 1st league was in a higher financial level than that of the 2nd league players.

Given the current economic and financial situation of football clubs, there is an increased need to know how efficiently a club is using 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 DEA methodology has an advantage to set benchmarks for inefficient players and identifies sources of inefficiency. Along with their general observations and experience, sport managers can take into account the DEA efficiency analysis, when creating teams.

Acknowledgements

The 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] Andersen, P. & Petersen, N. C. (1993). A Procedure for Ranking Efficient Units in Data Envelopment Analysis. *Management Science*, *39*(10), 1261–1264.

[2] Aydemir, A. E., et al. (2021). A Dimension Reduction Approach to Player Rankings in European Football. *IEEE Access*, *9*, 1–1.

[3] 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.

[4] Cooper, W. W., Ramón, N., Ruiz, J. L. & Sirvent, I. (2011). Avoiding large differences in weights in cross-efficiency evaluations: application to the ranking of basketball player. *Journal of centrum cathedra: The Business and Economics Research Journal*, *4*(2), 197–215.

[5] Cooper, W. W. (2011). *Handbook on data envelopment analysis*. New York: Springer.

[6] Cooper, W. W., Seiford, L. M. & Zhu, J. (2004). *Handbook on data envelopment analysis*. Boston: Kluwer Academic.

[7] 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.

[8] 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.

[9] Jablonský, J. & Dlouhý, M. (2015). *Modely hodnocení efektivnosti a alokace zdrojů*. Praha: Professional Publishing.

[10] Papahristodoulou, C. (2007). *The relative efficiency of UEFA Champions League scorers*. Mälardalen University: MPRA paper no. 4943.

[11] Ramón, N., Ruiz, J. L. & Sirvent, I. (2012). Common sets of weights as summaries of DEA profiles of weights: With an application to the ranking of professional tennis players. *Expert Systems with Applications*, *39*(5), 4882–4889.

[12] Santín, D. (2014). Measuring the technical efficiency of football legends: who were Real Madrid's all-time most efficient players? *International transactions in operational research*, *21*(3), 439–452.

[13] Suk, Y. K. (2014). Selecting the Batters of National Baseball Squad using Data Envelopment Analysis. *Journal of the Korea Academia-Industrial cooperation Society*, *15*(1), 165–172.

[14] 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.