Football clubs' salaries and transfer fees efficiency in Portuguese Primeira Liga

Isaac Alberto Monteiro Felizardo [1]

[1]*Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal, isaac.felizardo@tecnico.ulisboa.pt*

Abstract: Association football is the most popular sport in the world and in Portugal. It drives millions of fans every year to the stadiums, to television screens and radios. It is a millionaire industry that draws attention for the gigantic sums of money spent on players' salaries and transfer fees. This dissertation uses Data Envelopment Analysis methodology to study how efficiently these resources were used between the seasons of 2008/2009 and 2020/2021 of the Portuguese Primeira Liga, as well as trying to explain these efficiency estimates, sports performance and stadium attendance. The results show clubs in Primeira Liga operated in a variable returns to scale environment. While all clubs that made it to European competitions were operating in decreasing returns to scale, all the relegated teams were operating under increasing returns to scale. The maximum player salary of a team is very negatively correlated with efficiency, while the minimum player salary is positively correlated with it. The intra-team coefficient of variation of salaries is negatively correlated with efficiency and, for two of the three models, the Cohesive Theory was proven with a high significance level. In order to maximize sports results, teams should spend as much as possible on players' salaries and market values (i.e., transfer fees), but, given a limited budget, teams can maximize their sports performance by being as efficient as possible. What drives stadium attendance seems to be high salaries especially high maximum salary, big estimated market values, big goal differences, wins and inefficiency.

Keywords: Data Envelopment Analysis; Football; Efficiency; Salaries; Tournament Theory; Cohesive Theory; Estimated Market Values.

1. Introduction

Association football, or simply football in most European countries, is the most played sport in the world, exciting millions of fans every year (Kidwell, 2008; *Most Popular Sport by Country*, 2022).

In Portugal, specifically, it is the most widely adored sport (*Most Popular Sport by Country*, 2022) and Portuguese clubs do so well in European competitions that the 1st tier of professional football – Primeira Liga – is ranked 6th in Europe according to UEFA (UEFA, 2022). Obviously, the socio-economic relevance of the sport in the country attracts researchers with several scientific papers being published every year (Carvalho, 2019; P. Mourao & de Cima, 2015; P. R. Mourao, 2016; Ramos et al., 2022; Reilly & Gilbourne, 2008; Ribeiro, Lima, Kraus, & Calabuig, 2021). Despite all this research, no research has been made regarding Portuguese clubs' efficiencies since the 2007/2008 season. We believe this study is of uttermost importance not only to study, with a large dataset spanning 12 seasons, the efficiency of teams, but also what factors can explain it and what relationship it has with factors as important as performance, attendance and salary inequality.

This study aims at finding if there is a link between teams' salary and market value efficiency and their sport performance in Primeira Liga. Efficiency will be measured in relative terms using Data Envelopment Analysis methods over salary data, estimated market value data and the average points won per game during each season. Values of salaries and market values will be adjusted so they are comparable with each other, which allows for a greater comparison and more reliable estimates of our non-parametric production frontier.

Although finding the efficiency of teams' in using their players' salaries and market values is the primary goal of the study, we will also answer what explains efficiency, if there are variable returns to scale, how salary disparities influence team efficiency and performance, how can a team better leverage its resources to achieve maximum performance and efficiency and how is stadium attendance related to team performance and efficiency?

2. Football and the Portuguese Primeira Liga

This chapter gives the reader a full contextualisation of what football is, what type of football this study focuses on, what Primeira Liga is and how relevant is it and the problem characterization.

2.1 Football and Primeira Liga

Football is a group of team sports played with a ball that usually involve contact of the feet with the ball. From all the different football codes, association football became the most famous and even became the most popular sport in the world (Kidwell, 2008; *Most Popular Sport by Country*, 2022). From now on, it will be referred to merely as football for the sake of simplicity. This takes us to the league under study – the Portuguese Primeira Liga. Primeira Liga is the competition and whose teams this work will be studying. It is the highest league in the Portuguese league system and, currently, a total of 18 professional teams competes in it every year (until the 2014/2015 season, only 16 competed).

2.2 Problem Definition

We will be evaluating the sporting success of the team on a season level. Using designed models, we will measure the efficiency of each team in transforming their salaries expenditures and squad market in sports success in the league.

3. Literature Review

This chapter lays the theoretical foundations for the study. In the first section, the history of scientific literature on efficiency, including methodologies, and on football teams' efficiency are shown. After, in the second section, theoretical explanations of possible tools for explaining relationships within our variables and results are presented. In the third and last section, a brief look into the scientific literature of football players' wages, including the consequences of their disparities, and market values is done.

3.1 Efficiency in Football

First and foremost, it would be important to define the term "efficiency". (Farrell, 1957) defined efficiency as "the firm's success to produce the maximum feasible amount of output from a given amount of input or producing a given amount of output using the minimum level of inputs where both the inputs and the outputs are correctly measured".

Data Envelopment Analysis (DEA) is a methodology used to determine the productive efficiency of a Decision-Making Unit (DMU). By setting the benchmark on the best performance(s) analysed, DEA estimates the best practice efficiency frontier of the units under observation considering multiple inputs and outputs (Farrell, 1957). Also, important to note is that DEA is classified as "nonparametric" because it does not impose assumptions on the error terms, contrasting with "parametric" efficiency estimation approaches that specify how dependent variables are affected by independent variables and how the error term is handled. The obvious advantage of DEA as a non-parametric approach is its "robustness to changes in assumptions about the underlying structure of the error term." (Stolp, 1990)

Many variations of DEA appeared, being the two most popular the one that assumes constant returns to scale (CRS) (Charnes, 1978) and the one that assumes variable returns to scale (VRS) (Banker et al., 1984). These two models diverge only because the latter allows for DMUs that use less inputs to have increasing returns to scale and DMUs that use more inputs to have diminishing returns to scale (Cooper et al., 2007), hence the VRS model efficiencies being greater than or equal than those of CRS.

Both models can be oriented according to inputs or according to outputs.

The inputs are transformed via weights into a single "virtual" input and outputs are transformed into a "virtual" output. Formally, the (Charnes, 1978) CRS model, can be expressed as:

$$\max z_0 = \sum_{r=1}^{m} u_r y_{r0} \tag{1}$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, j = 1,2,\dots,n \tag{2}$$

$$\sum_{i=1}^{m} v_i x_{i0} = 1 \tag{3}$$

$$u_r \geq 0, r = 1,2,\dots,s \tag{4}$$

$$v_i \geq 0, r = 1,2,\dots,m \tag{5}$$

, where u_r and v_i are the weights to be determined by the linear programming problem, the x_{ij} is the amount of the i-th input for the j-th DMU, and the y_{ij} is the amount of the i-th output for the j-th DMU.

The dual of the above linear programming problem can be written as below, in the DEA CRS Input-Oriented Model:

$$\min \Theta \tag{6}$$

$$\sum_j \lambda_j x_{ij} \leq \Theta x_{i0}, i = 1,2,\dots,m; \tag{7}$$

$$\sum_j \lambda_j y_{rj} \geq Y_{r0}, r = 1,2,\dots,s; \tag{8}$$

$$\lambda_j \geq 0, j \in 1,2,\dots,n \tag{9}$$

The essential mathematical difference between the two models is the following constraint present in the VRS model:

$$\sum_j \lambda_j = 1 \tag{10}$$

With this added constraint, the reference set is no longer cone shaped, but rather a convex hull. The result of this addition Is that each DMU will only be compared against a limited number of combinations, resulting in an equal or greater efficiency than in the CRS model. (Luo, 2003).

A survey of DEA applications (Liu et al., 2013) found that, from 1978 to 2010, 63.6% of the papers using DEA embedded empirical data and the remaining were only theoretical. DEA papers regarding the Sport industry accounted for only 0,99% of the sample analysed. The top-five industries under study were banking, health care, agriculture, transportation, and education, but studies went as far as looking at the efficiency of tourism, e-business, real estate and even governments. Football was no exception, and some work has been done on clubs' wage efficiency (Ribeiro & Lima, 2012), total squad market value (Zambom-Ferraresi, Lera-López, et al., 2017), game-related statistics efficiency (Zambom-Ferraresi, García-Cebrián, et al., 2017) and (García-Cebrián et al., 2018), operation-athletic-social efficiency relationship (García-Sánchez, 2007), relationship between value/debt levels and performance (Halkos & Tzeremes, 2013).

3.2 Wages and Market Values

Studies around sports players wages and, especially in this thesis, football players are plentiful. They play such an important role in sports success of teams, that (Szymanski & Kuypers, 1999) found that over a 20-year period and analysing 48 clubs from the English league, 95% of clubs' sports performance variation was explained by players' salaries. (Forrest & Simmons, 2000) found similar relationships for the Italian, English and German first divisions, albeit much weaker – players wages explained between 76% to 22% of the sports performances. Some studies have been performed on the Portuguese Primeira Liga regarding players wages. For example, (Carvalho, 2019) found that teams' expenditures with players wages explained between 48% and 64% of the sports performance differences. (Franck & Nüesch, 2010) Tournament theorists suggest that participants are more motivated if their payments are attributed on a basis of winners and losers than on absolute output measures. (Connelly et al., 2014). One opposing idea – the cohesion theory – argues that organizations would increase their productivity if they reduced their pay gap because more salary equality would increase a sense of unity (Levine, 1991). . Specifically in football, or soccer as the authors call it, a study (Coates et al., 2016) done on MLS teams' salaries tested the two opposing theories. Using the Gini coefficient and the coefficient of variation to measure payroll inequality, the study found "a negative relationship between salary inequality and production" in MLS teams, which implies increasing salary inequality in MLS tends to decrease sporting success. Another study (Franck & Nüesch, 2010), performed with data from the Bundesliga football league, found a U-shaped relationship between intra-team wage gaps and team performance, meaning "team performance is strongest when there is either very high or very low wage inequality" and teams in the middle would perform the worst (Pelloneová & Tomíček, 2022; Pyatunin et al., 2016; Zambom-Ferraresi, Iráizoz, et al., 2018; Zambom-Ferraresi, Lera-López, et al., 2017)(Coates & Parshakov, 2022; He et al., 2015; Herm et al., 2014; Kirschstein et al., 2019; Majewski, 2016; Müller et al., 2017; Romann et al., 2021; Serna Rodríguez et al., 2019; Singh et al., 2019; Velema, 2019)

A bombastic part of the football world is the amount of money football clubs spend on transfer fees every year to sign players from other clubs. The goal of this global and formal transfer market is to "organize the acquisition and exchange of players" and to "facilitate the movement of players between clubs in their search for better opportunities". These transactions are usually supervised, and, in many cases, restrictive controls apply (Carmichael & Thomas, 1993). Transfer fee is defined as "financial compensation agreed to be paid between clubs in the course of a player transfer" (FIFA TMS, 2020). One alternative to the usually undisclosed transfer fee is to use a proxy for a probable transfer fee – the market value of the player. The market value is the estimate of "the transfer fee if the players were transferred during the present season" (He et al., 2015). The most popular source of player market values is an online platform called Transfermarkt.com, which leverages the wisdom of the crowds to make predictions regarding the players' market values. Data from *Transfermarkt.com* have been used both in predictive models of transfer fees/market values (Coates & Parshakov, 2022; He et al., 2015; Herm et al., 2014; Kirschstein et al., 2019; Majewski, 2016; Müller et al., 2017; Romann et al., 2021; Serna Rodríguez et al., 2019; Singh et al., 2019; Velema, 2019).

4. Methodology

The Methodology is written in a simple, clear and comprehensive manner so that any reader, being a scholar or a football decision-maker, can both reproduce the study in a different context and understand the limitations of the results and conclusions to be drawn. The rationale behind each of the methodological choices is also given.

To measure the efficiency of the Portuguese football teams over the period, we chose to use DEA – DEA was chosen to follow-up the previous study done on the Portuguese Primeira Liga (Ribeiro & Lima, 2012). Both a DEA-CRS and DEA-VRS were used to enrich the study, spark discussion of results and draw conclusions. Because football clubs have direct control over their expenditures, but much less control over the results that are sometimes affected by external factors and even by chance, we chose to use the input-oriented version of the DEA models, thereby increasing the value of the conclusions of this study for football decision-makers.

As explained in the literature review chapter, salaries have an indisputable reputation to predict sports results in football, so it was chosen to include them in our study. Not having access to trustworthy sources of transfer fees of clubs during the season transfer windows, we decided to include estimated market values of squads. However, it is important to note that the models developed using estimated market values do not estimate the efficiency of teams' investments on transfer fees, but rather a proxy of the efficiency of teams regarding using their assets in terms of players' contracts. Output(s) were trickier to decide upon because starting on the 2014/2015 season, the Portuguese Primeira Liga structure changed, and it started having 18 teams competing instead of 16. That made using directly available information such as points obtained, goals difference or final ranking impossible because we couldn't compare it timewise. It was chosen to create a variable - points per game (PPG) – that would show proportionally the sports success of teams in each season. This was performed by dividing the points obtained in the league by the number of matches played.

Wages aggregated on a team level by season come from a matched employer-employee dataset – the Quadros de Pessoal (QP). This is a mandatory administrative survey collected by the Portuguese government which has been used extensively in labour research. One of the limitations of the data collected from QP is that October is the reference month for collected data. This means that, on an individual level, data could be biased by not knowing if the player was paid a performance prize during or after the season. It may even cause some players to be untracked because of players on loan (do not belong to the team's staff), players hired in January or simply because of missing records (Ribeiro, Lima, Kraus, Calabuig, et al., 2021).

The estimated aggregate market value of each team's squad comes from the website *Transfermarkt.com*, that, as previously discussed, leverages the 'Wisdom of the Crowds' to assess the market value of football players. The main limitations of data provided by *Transfermarkt.com* are the loss of accuracy due to some members having less experience/knowledge, the lack or rarity of estimates for less well-known players and teams and the time it takes between updates to a player's value (Behravan et al., 2020).

Sports metrics for each team by season were retrieved from *fbref.com*, a free online platform whose data is powered by Data Sports Group and StatsBomb, with the purpose of becoming "the trusted source of information and tools that inspire and empower […] users to enjoy, understand, and share the sports they love." (Sports-Reference.com, 2022).

We wanted the dataset to be as big as possible, as advised, but first we needed to worry with data homogeneity – for data to be comparable in time, we only used salaries adjusted with a wage deflator. The same deflator was used to adjust player market values to 2020 levels, originating the new variable "TotalMV_real". The same deflator was used due to lack of consensus in the literature for an inflation rate for market values and since market values are used as a proxy for wages (Poli et al., 2021), this approximation was chosen. It is however seen as another limitation of the study.

The models built to help us achieve our study's objectives are described below. The Decision-making Unit was always considered to be the pair team-season. In all of the models the output "PPG" was used. In the first model (M1), we take as input the total estimated market value of the squad:

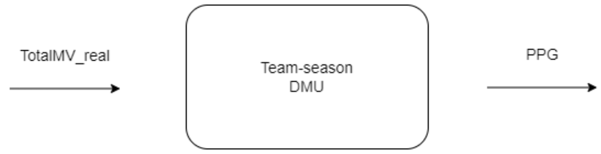


Figure 1: Schematic representation of model M1

In the second model (M2), the input is the total salary expenditure with players (real values):

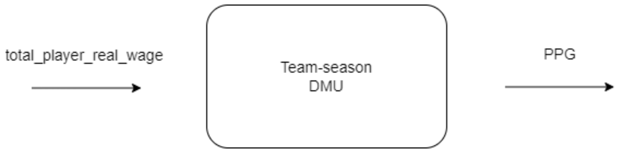


Figure 2: Schematic representation of model M2

Lastly, one third model was developed (M3) which combines both inputs:

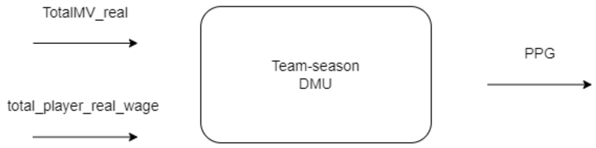


Figure 3: Schematic representation of model M3

With the models conceptually defined, we proceeded with the calculations. For the implementation of the calculations regarding DEA technical efficiencies estimations, we opted to use the "Benchmarking" R package (Bogetoft et al., 2022) (*CRAN - Package Benchmarking*, 2022) because it was cited in at least 26 publications (swMATH, 2022). For implementing the bootstrap sensitivity analysis on DEA, the same library was used, with the 2000 iterations, as recommended in (Simar & Wilson, 1998). For rendering the DEA plots, another library had to be picked – "deaR" R package (Vicente Coll-Serrano et al., 2022). The same package was used to retrieve the returns to scale of each DMU.

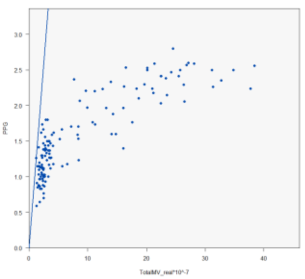
Because it is important to sports decision-makers in order to drive profits, we will also investigate how stadium attendance, performance and efficiency might be correlated. The Covid-19 pandemic negatively affected stadium attendance, creating outliers in our data. It was chosen to remove 2019/2020 and 2020/2021 season data from the Attendance analysis to eliminate these outliers.

5. Results and Discussion

This section will present the results of the in a logical and chronological format resembling a story.

5.1 Data Envelopment Analysis

The results of the three different models, both for CRS and VRS DEA calculations (always input-oriented) are more easily understood when seen graphically in the usual DEA plots. For M1 and M2 models, we retrieved the estimated efficient frontiers.



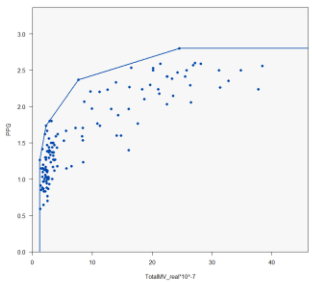
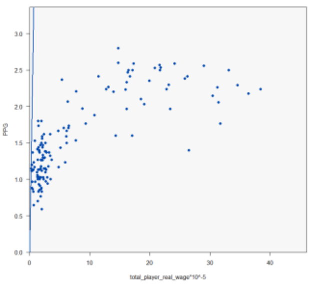


Figure 4: Efficient frontier for M1 (assuming CRS)

Figure 5: Efficient frontier for M1 (assuming VRS)



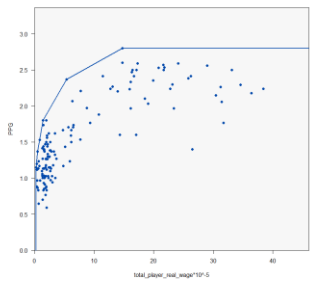
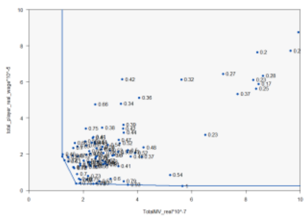


Figure 6: Efficient frontier for M2 (assuming CRS)

Figure 7: Efficient frontier for M2 (assuming VRS)

Even though the R software package "deaR" offered the calculations of RTS for each DMU, it is interesting to see the plots above in the DEA-CRS models (Figure 4 and Figure 6), the constant returns to scale assumption completely prohibits top spending teams of ever achieving efficiency. The DEA-VRS models (**Error! Reference source not found.** and REF _Ref117952888 \h * MERGEFORMAT **Error! Reference source not found.**) created the convex hull that looked for the next successive efficient DMU. Whilst the DEA-CRS M1 and M2 only find one efficient DMU – 2019 Gil Vicente FC and 2018 Portimonense, respectively – DEA-VRS models find multiple efficient DMUs. As a curiosity for the reader, the 2019-2020 season of Gil Vicente FC is particularly interesting concerning this study because the team was administratively promoted to Primeira Liga following a judicial decision. At the time, the squad had to be built from scratch with several players being hired to prepare the team for the first tier of the Portuguese professional football. This is an indication that those decision makers tried to maximize the sports performance of the team with the available budget; the team was considered to have a technical efficiency of 1 for DEA-CRS in M1 and M3 and for DEA-VRS in M1 and M3.



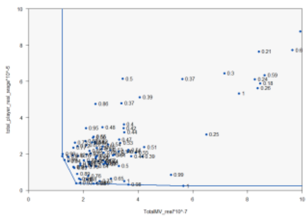


Figure 8: Production isoquant for M3 (assuming CRS)

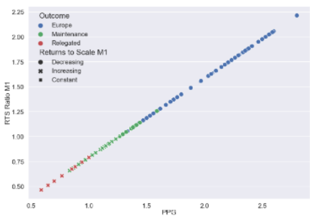
Figure 9: Production isoquant for M3 (assuming CRS)

The bootstrapped DEA bias-corrected efficiencies had excellent results, achieving correlations close to 1 (subject to software roundings). As expected, with better performances in the Pearson and Spearman correlation measures with a slightly worse performance in the Kendall tau for the recommended 2000 iterations. Thus, highlighting the robustness of our estimates of efficiency (see Table 1):

Table 1: Correlation measures of DEA results with Bootstrapped DEA version

	Correlation with Bias-corrected efficiencies		
	Pearson's p	Spearman's ρ	Kendall's τ
CRS1	1	1	1
VRS1	0.9870674	0.9923762	0.9375613
CRS2	1	1	1
VRS2	0.9905475	0.9968178	0.9582563
CRS3	0.9907693	0.9949215	0.9604446
VRS3	0.9832753	0.9861485	0.9172687

Let us take a closer look on the nature of the returns to scale and its relationship with the outcome of the team on the league. In the one input to one output models (M1 and M2), there is a linear relationship between the scale at which the club operates and its returns to scale level (see Figure 10 and Figure 11):



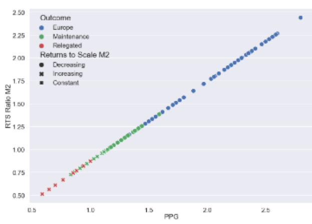


Figure 10: PPG and scale efficiency of M1

Figure 11: PPG and scale efficiency of M2

However, when there is more than one input to output, like it is the case for model M3, the relationship is not as simple and linear as it involves a combination of both inputs and outputs (see Figure 12):

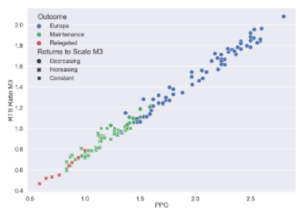


Figure 12: PPG and scale efficiency of M3

Furthermore, one of the things that is possible to graphically notice from the previous charts is the relationship between scale, outcome of the league and efficiency. Whereas all clubs that were awarded a place in an European competition are operating in a decreasing returns to scale frontier, all those relegated are working in an increasing returns to scale environment, meaning a small investment (their inputs) from their side would have a big impact on their outputs. It is in the group that managed to stay in the league without attaining a European spot that lies the efficient scale – some are operating at in the increasing RTS, others in the decreasing segment and a few at an optimal scale.

The most attentive of readers will also notice some of the clubs which did not attain a European spot had higher PPG than some of those who did. There are two independent explanations for this phenomenon. The first is that European (and relegation for that matter) spots are awarded on a final ranking basis and different seasons will end up in different PPG values for each position in the ranking. The second explanation is the fact that the number of European places Portuguese teams have available varied in time according to the moving average of performances of Portuguese teams in these same competitions in prior seasons.

Following the same rationale, the same reader would find suspicious how some teams who were relegated had significantly more PPG than some of those who stayed in the league. Again, there is a twofold explanation. The antepenultimate team in the league has access to a play-off against the third best team of the “Liga Portugal 2” (the second tier of Portuguese professional football). Depending on the outcome of this playoff, it is decided if the team is relegated or not. The second part of the explanation are the administrative relegations due to unlawful actions or financial problems, which create outliers in term of final ranking versus PPG at which they were relegated.

From this point onward, only results related with efficiencies obtained with VRS DEA calculations are shown because the CRS assumption does not hold.

5.2 Explaining Efficiency

This study would be uninteresting if we did not try to explain efficiency for sports decision makers to have actionable insights to improve or to reflect upon.

As explained in the Methodology chapter, different combinations of variables were used to look for the best-fitting multiple linear regressions to each of the efficiency estimates. You can find in the Appendix the Spearman’s rho correlations between the variables considered and the technical efficiency estimates “deavrs1”, “deavrs2” and “deavrs3”.

In front of each variable, there is the coefficient of variation of the variable, and, beneath it, there is the standard error between parentheses.

MLR on Model 1

Table 2: Multiple Linear Regression for M1

	Dependent variable:
	deavrs1
coefficient of variation player real wage	-0.297^{***}
	(0.077)
PlayersUsedPG	-0.435^{***}
	(0.147)
mean player real wage	-0.114^{***}
	(0.024)
WPG	0.929^{***}
	(0.180)
Outcome: Maintenance	0.076
	(0.049)
Outcome: Relegated	0.270^{***}
	(0.083)
Age	0.028^{**}
	(0.013)
total coach real wage	-0.070^{***}
	(0.025)
coefficient of variation coach real wage	0.116
	(0.078)
Constant	1.634^{***}
	(0.446)
Observations	136
R^2	0.476
Adjusted R^2	0.439
Residual Std. Error	0.160 (df = 126)
F Statistic	12.738^{***} (df = 9; 126)
Note:	$^{*}p<0.1$; $^{**}p<0.05$; $^{***}p<0.01$

In explaining the efficiency of clubs using the estimated market values of players, a MLR was achieved with r^2 = 0,476.

One can attest with a high significance level that the salary disparity of players, the number of players used, the mean level of players’ salaries and the total spent on coaches negatively affects the efficiency. Perhaps unsurprisingly, the number of wins increases efficiency with high significance, as it adds three points versus one point for a draw. The unitary variance in the number of wins per game has an effect of 0.929 in the dependent variable, ceteris paribus.

Players’ average age benefits teams with significance (p-value < 0.05), probably because players’ potential, when they are young, is incorporated in their market value, but that does not translate in on-field performance. Relegated teams appear to be more efficient, with high significance.

MLR on Model 2

In explaining the efficiency of clubs using their total amount spent on players’ salaries, a MLR managed to have an r^2 = 0,767. Regarding players’ salaries, more variables show their relevance. The ratio between money spent on the coaching staff and players, wins per game and shots on target per goal has a positive impact on team efficiency, with high significance. The % of saves, the minimum player wage and the mean coach wage also impact efficiency positively with a p-value < 0.05. The disparity of the coaching staff salary and the maximum paid to a player have a negative relationship with efficiency with a high significance level. However, the coefficient of variation of players wages has a slightly positive impact in efficiency.

Table 3: Multiple Linear Regression for M2

	Dependent variable:
	deavrs2
total coach to total player real wage	0.114***
	(0.032)
LPG	-0.433**
	(0.179)
log(max player real wage)	-0.250***
	(0.024)
WPG	1.252***
	(0.210)
Outcome: Maintenance	-0.042
	(0.037)
Outcome: Relegated	0.082
	(0.063)
coefficient of variation coach real wage	-0.164***
	(0.044)
SoTPG	0.057***
	(0.019)
log(mean coach real wage)	0.046**
	(0.018)
coefficient of variation player real wage	0.141*
	(0.082)
log(min player real wage)	0.050**
	(0.020)
GDPG	-0.110
	(0.068)
Save%	0.007**
	(0.003)
Constant	1.022***
	(0.337)
Observations	136
R^2	0.767
Adjusted R^2	0.742
Residual Std. Error	0.117 (df = 122)
F Statistic	30.885*** (df = 13; 122)
Note:	*p<0.1; **p<0.05; ***p<0.01

It is expected to see wins favour efficiency and losses damaging it, but less intuitive is to see the average goal difference contribute negatively to efficiency. To understand this, one must go back to the model definition and understand our output is the average points per game. Even though goals ultimately lead to wins, goals are not wins. So, a team winning a match by a great margin versus a team winning a match by the minimal margin (1 goal) makes the first less efficient in comparison, as it probably had to spend more money for the same number of points – 3. Although it contributed to the explanatory value of the MLR, the Outcome of the teams did not reach a p-value of 0.1 in this regression.

MLR on Model 3

The maximum player wage negatively affected team efficiency, just like the total spent on coaches, even though the mean of coaches' salaries affected it positively. Wins, again, impacted positively teams' efficiency. All these variables have high significance. The number of players used negatively impacted efficiency again, while players' age positively impacted it again. Relegated teams were again more efficient. The minimum player wage had positive impacts on teams' efficiencies.

Table 4: Multiple Linear Regression for M3

	Dependent variable:
	deavrs3
log(max player real wage)	-0.152***
	(0.045)
WPG	0.928***
	(0.175)
Outcome: Maintenance	0.067
	(0.045)
Outcome: Relegated	0.245***
	(0.078)
PlayersUsedPG	-0.355**
	(0.137)
Attendance	0.00000*
	(0.00000)
log(total coach real wage)	-0.087***
	(0.023)
log(mean player real wage)	-0.088
	(0.055)
total coach to total player real wage	0.058
	(0.042)
Constant	2.058***
	(0.416)
Observations	136
R^2	0.672
Adjusted R^2	0.640
Residual Std. Error	0.148 (df = 123)
F Statistic	21.030*** (df = 12; 123)
Note:	*p<0.1; **p<0.05; ***p<0.01

General considerations

The maximum amount spent on a player wage seems to be one major source of inefficiency of teams. Interestingly, the minimum spent of players' wages always had a positive relationship with efficiency. Even though one cannot assume, from the above models how the coefficient of variation of salaries impacts the overall efficiency of teams, their decision-makers should investigate raising their poorly paid players and look for potential squandering with maximum player wage. The number of players used per season also dents on efficiency. It is also interesting to note how the coefficient of variation of players' wages contributes positively to the team's efficiency using salaries, but negatively to their efficiency using their players' market values. There would not be a better introduction for the next subchapter.

5.3 Salary Inequality and Efficiency

This section delves into how salary inequality is related to team efficiency and if the Tournament theory or the Cohesive theory holds true in this scenario. To evaluate which salary disparity theory – tournament or cohesive – was verified during the studied seasons, regressions were made to try to predict teams' efficiency estimates solely based on the intra-team salary inequality.

The linear regression on the efficiency of teams in transforming their players' values into points, the coefficient of variation of their salaries had a $r^2 = 0.187$, negatively sloped (-0.388) and with a p-value = $1.487*10^{-7}$. The intercept was of 0.756.

Strangely, the same was not verified for the efficiency with salaries, where, although negatively sloped (-0.128), the p-value of the regression was (slightly) bigger than 0.1 (0.124). The intercept was of

0.374 and the $r^2 = 0.017$. For the third model, the best r^2 (0.295) and p-value ($8.098*10^{-12}$) were achieved. The slope was of -0.564 and the intercept was 0.944. As shown in the charts above, (Franck & Nüesch, 2010)'s conclusion about Bundesliga that team performance is best for very high or very low levels of salary inequality does not apply to our variable returns to scale models. In our models, team efficiency is highest when the coefficient of variation of players' salaries is lowest, even though we could not prove it with a significant level for the M2 model. This falls short of verifying the Cohesive Theory for our studied scenario in the M2 model but proves it with high significance levels for M1 and M3.

Even when considering PPG as the de facto performance in the league, the same U-shaped relationship does not hold, as the chart shows. In the Portuguese Primeira Liga, teams wanting to go to Europe seem to increase their coefficient of variation. It partly verifies (Coates et al., 2016) findings on MLS teams' salary disparity.

We also wanted to investigate if we could find (Frick, 2006) findings on the ratio between coaches' salaries and players' salaries.

Table 5: Correlation measures of the ratio total coach to total player real wages and efficiency measures

	Correlation with Total Coach/Total Player wages		
	Pearson's ρ	Spearman's ρ	Kendall's τ
VRS1	0.046 p-val = 0.59	0.067 p-val = 0.44	0.046 p-val = 0.42
VRS2	0.377 p-val < 0.001	0.336 p-val < 0.001	0.237 p-val < 0.001
VRS3	0.326 p-val < 0.001	0.266 p-val = 0.002	0.180 p-val = 0.002
PPG	-0.186 p-val = 0.03	-0.170 p-val = 0.05	-0.114 p-val = 0.05

According to Table 5, and just like in the previous analysis, the ratio between coach and player wages is positively correlated with efficiency measures, but negatively correlated with sports performance measured by points per game. It's worth noting that the correlations between efficiencies of M1 and the ratio being analysed did not reach any significance threshold. Top spending teams can thus afford to spend more on the better and star players while poorer teams try to be as efficient as possible with the players, they have by hiring the best coach possible. Further research could be done both on this relationship and on the sports management rationale on the field.

5.4 Attendance and Efficiency

This section aims to explain the relationship between stadium attendance and efficiency, as well as with sports results. After dropping the 2019/2020 and 2020/2021 seasons, we found the relationships between our study variables and Attendance. It is very interesting to see that the variables most correlated with Attendance are not efficiency estimates, wins and not even goals per match, but total amount spent on player wages and related variables, the maximum paid to a player and the total market value of the team. This is very interesting because it shows the star players have a bigger attraction effect on the fans than wins and goals. Fans love when their team scores, with a big goal difference preferably, but also their team's goalkeeper skills in terms of saves % and clean sheet %. Apparently, fans hate losing and suffering goals even more. This part was expected. Efficiency of teams, from the three models, also has a negative effect on Attendance. With the ratio variables built, it is also possible to see they prefer highly paid players to highly "valuable" players and prefer highly paid players to highly paid coaches. The rationale on the previous calculations is interesting as it verses on reality, but it is not actionable information for the typical sports decision-maker that, in the short term, has budget constraints that would not allow for hiring the highest paid stars. For that reason, we analyzed the Spearman partial correlations of Attendance while controlling for the total player wage expenditure and total market values. The most interesting change, given the study at hand, is that after controlling for the amount spent on players, efficiency measures are positively correlated with Attendance. The coefficient of variation of players' salaries becomes the most positively correlated variables. The coaches' wages also gain relative relevance.

5.5 Maximizing Points Per Game and Efficiency

The most anticipated section is how performance relates with efficiency and how maximizing it would impact the latter.

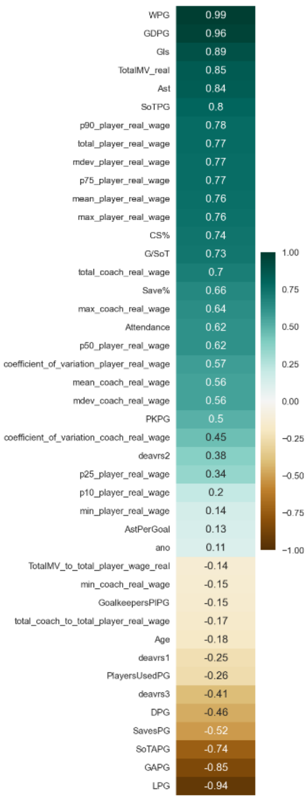


Figure 13: Correlation heatmap of PPG and other variables

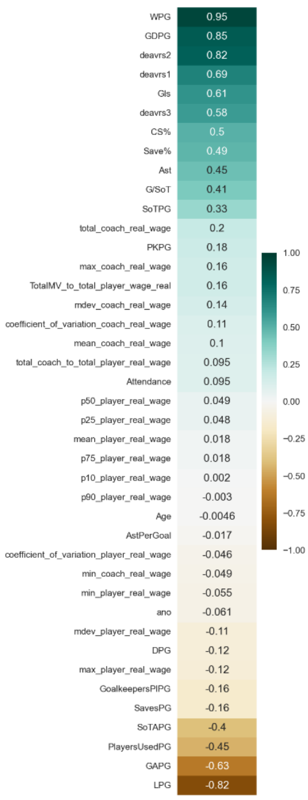


Figure 14: Partial correlation heatmap of PPG and other variables (controlling for total player real wages and total real market values)

Perhaps unsurprisingly, wins, goal difference and goals are all highly positively correlated with points per game. Clean sheets % and saves are once again correlated with points. Efficiencies found in M2 are positively correlated with PPG, while M1 and M3 efficiencies are negatively correlated with it. Losing goals, suffering goals, having shots on target on their goal, the absolute number of saves per game and ties are all things negatively correlated with points per game. It is important to note that the absolute number of saves is negatively correlated with points per game because it is also correlated with the number of opponents shots on target and hence offensive pressure. In the opposite direction, the goalkeeper's ability to secure the goal measured as the save % is positively correlated. Still, it seems the best way to win points is to get the players with the highest market value and pay them well. As will be shown after, players' wages and market values explain most of the team's performance. To show how important the total expenditures with player wages is and how important the market value of players is, the linear regressions underneath were plotted.

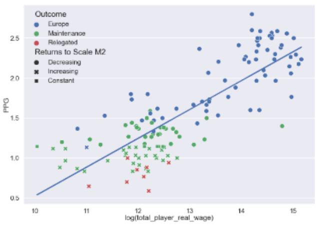


Figure 15: Linear regression between the logarithmized total player real wages and PPG

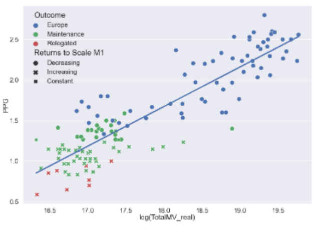


Figure 16: Linear regression between the logarithmized total real market values and PPG

Starting with the linear regression using the logarithmized total expenditure on players' wages to predict PPG, one reached a slope of 0.361 with an r^2 of 0.676 and a p-value < 0.001. The intercept was of -3.088 and the standard error was 0.022. This high significance regression is in line with the studies mentioned in the literature review chapter on the importance of the total players' wages expenditures and performance.

The linear regression between the logarithmized market value of the squad and PPG had a positive slope of 0.488 and a $r^2 = 0.800$, with a p-value < 0.001. The intercept was of -7.11 and the standard error was of 0.021. This shows how much explanatory power the market value of players has with high significance level. It is also interesting to note that the crowdsourced estimates of market value have more explanatory value than players' salaries, which probably confirms that the wisdom of crowds beat the experienced decision-makers, who decided the salaries, during the seasons under study in Primeira Liga. Qualitatively, from the regression studies and from the plot between total squad market value and total player wages, one can see how disproportionate the investment is for clubs who want to go to Europe, and that are probably competing at the same time in European competitions. Of course, all of this is important in order to understand the reality being studied, but, in the short term, is not attainable for the typical sports decision-maker. Hence, a chart with the partial Spearman correlations was plotted. Obviously, wins and goal difference are still the most important variables to positively influence PPG, but what is interesting is that efficiency measures, especially regarding players' wages, are highly correlated with PPG. So, given a limited budget, decision-makers should try to maximize their teams' efficiency. As shown previously, the ratio between coaches' and players' salaries was positively correlated with efficiency and, when controlling for players' salaries and market values, the total spent on coaches gains prominence as a tool to increase PPG. Defensive capabilities like the clean sheets and saves % are positively, while the number of shots on target against and total number of saves are negatively correlated. This indicates keeping the goal safe from opponents' shots is of uttermost importance. The average number of players used also dents on PPG. Therefore, teams should try to remain compact in order to achieve higher PPG. In a similar fashion to what (Zambom-Ferraresi et al., 2018) concluded, the average goals per game have more explanatory power than the absolute number of shots on target, and the percentage of saves has more explanatory power than the absolute number of saves – one can conclude, for the data available, that the accuracy is more important than the total number of actions.

Conclusions and Future Work

Our models were successful in estimating a nonparametric production frontier and in computing technical efficiencies estimates for teams competing in Primeira Liga for the seasons between 2008/2009 and 2020/2021. Teams in this league operated with variable returns to scale as shown mathematically and graphically.

Teams who achieved European qualification were all operating in the decreasing returns to scale part of the production curve and were often inefficient. The ideal scale was found among clubs who avoided relegation but who could not reach European spots. All the relegated clubs in the study were still operating in the increasing returns to scale part of the curve.

The maximum wage of a team's player was a source of inefficiency in Primeira Liga for the seasons studied. In the inverse direction, a higher minimum salary of a player in the league was connected to higher efficiency levels. The intra-team salary inequality worsened team efficiency, even though it was associated with higher points per game. Interestingly, total coach to total player wage ratio had a positive correlation with efficiency in using players' salaries, but a negative one with points per game – suggesting poorer teams try to hire relatively better coaches to make the most out of their players. Teams who field less players in Primeira Liga are more efficient, and, controlling for players' salaries and market values, also get more points in the season. In absolute terms, maximizing points per game is achieved by "buying" wins, goals, assists and clean sheets, which means paying high salaries to valuable players. However, for a limited budget in terms of players' market values or salaries, maximizing points per game is directly related to maximizing efficiency. Stadium attendance is maximized by paying high salaries, especially the maximum salary of a player, investing in big estimated market values, achieving big goal differences, wins, goals, assists, shots on target and being inefficient. Crowdsourced estimates of squad market value, using the wisdom of the crowds, had a greater explanatory value on team results than player wages decided by clubs' decision-makers.

In the future, more research on Primeira Liga teams' efficiency could be done including financial performance indicators as outputs of the study. As was shown, a behavioural economics study on the impacts of players and staff salary structures on the "X-efficiency" of the club, as well as other internal factors. Further studies on intra-team wages variations and how they affect performance and cooperation between team members with more extensive data.

References

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.

Behravan, I., Seyed, •, & Razavi, M. (2020). A novel machine learning method for estimating football players' value in the transfer market. *Soft Computing*, *25*. https://doi.org/10.1007/s00500-020-05319-3

Bogetoft, P., Maintainer, L. O., & Otto, L. (2022). *Package "Benchmarking" - Benchmark and Frontier Analysis Using DEA and SFA*.

Carmichael, F., & Thomas, D. (1993). Bargaining in the transfer market: Theory and evidence. *Applied Economics*, *25*(12), 1467–1476. https://doi.org/10.1080/00036849300000150

Carvalho, R. (2019). *A relação pay-performance no futebol português*. https://www.repository.utl.pt/handle/10400.5/17747

Charnes, A. ; W. W. C. E. R. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research*, 429–444. https://doi.org/10.1016/j.rser.2016.12.030

Coates, D., Frick, B., & Jewell, T. (2016). Superstar Salaries and Soccer Success: The Impact of Designated Players in Major League Soccer. *Journal of Sports Economics*, *17*(7), 716–735. https://doi.org/10.1177/1527002514547297

Coates, D., & Parshakov, P. (2022). The wisdom of crowds and transfer market values. *European Journal of Operational Research*, *301*(2), 523–534. https://doi.org/10.1016/J.EJOR.2021.10.046

Connelly, B. L., Tihanyi, L., Crook, T. R., & Gangloff, K. A. (2014). Tournament Theory: Thirty Years of Contests and Competitions. *Journal of Management*, *40*(1), 16–47. https://doi.org/10.1177/0149206313498902

Cooper, W. W., Seiford, L. M., & Tone, K. (2007). *Data envelopment analysis: a comprehensive text with models, applications, references and DEA-solver software*.

CRAN - Package Benchmarking. (2022). https://cran.r-project.org/web/packages/Benchmarking/

Farrell, M. J. (1957). The Measurement of Productive Efficiency. *Journal of the Royal Statistical Society. Series A (General)*, *120*(3), 253–290.

FIFA TMS. (2020). *Global Transfer Market Report 2020. A Review of International Football Transfers Worldwide. Zurich: FIFA*.

Forrest, D., & Simmons, R. (2000). *The relationship between pay and performance: Team salaries and playing success from a comparative perspective*.

Franck, E., & Nüesch, S. (2010). The effect of wage dispersion on team outcome and the way team outcome is produced. *Https://Doi.Org/10.1080/00036840903427224*, *43*(23), 3037–3049. https://doi.org/10.1080/00036840903427224

Frick, B. (2006). Salary determination and the pay-performance relationship in professional soccer: Evidence from Germany. *Sports Economics After Fifty Years: Essays in Honour of Simon Rottenberg. Oviedo: Ediciones de La Universidad de Oviedo*, 125–146.

García-Cebrián, L. I., Zambom-Ferraresi, F., & Lera-López, F. (2018). Efficiency in European football teams using WindowDEA: analysis and evolution. *International Journal of Productivity and Performance Management*, *67*(9), 2126–2148. https://doi.org/10.1108/IJPPM-02-2018-0053

García-Sánchez, I. M. (2007). Efficiency and effectiveness of Spanish football teams: a three-stage-DEA approach. *CEJOR*, *15*, 21–45. https://doi.org/10.1007/s10100-006-0017-4

Halkos, G. E., & Tzeremes, N. G. (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

He, M., Cachucho, R., & Knobbe, A. (2015). *Football Player's Performance and Market Value*.

Herm, S., Callsen-Bracker, H. M., & Kreis, H. (2014). When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community. *Sport Management Review*, *17*(4), 484–492. https://doi.org/10.1016/J.SMR.2013.12.006

Kidwell, J. (2008, November 15). 226 countries can't be wrong. *Bleacher Report*. https://bleacherreport.com/articles/82112-226-countries-cant-be-wrong

Kirschstein, T., Statistics, S. L.-J. of A., & 2019, undefined. (2019). Assessing the market values of soccer players–a robust analysis of data from German 1. and 2. Bundesliga. *Taylor & Francis*, *46*(7), 1336–1349. https://doi.org/10.1080/02664763.2018.1540689

Levine, D. I. (1991). Cohesiveness, productivity, and wage dispersion. *Journal of Economic Behavior & Organization*, *15*(2), 237–255. https://doi.org/10.1016/0167-2681(91)90031-R

Liu, J. S., Lu, L. Y. Y., Lu, W. M., & Lin, B. J. Y. (2013). A survey of DEA applications. *Omega*, *41*(5), 893–902. https://doi.org/10.1016/J.OMEGA.2012.11.004

Luo, X. (2003). Evaluating the profitability and marketability efficiency of large banks: An application of data envelopment analysis. *Journal of Business Research*, *56*(8), 627–635. https://doi.org/10.1016/S0148-2963(01)00293-4

Majewski, S. (2016). Identification of factors determining market value of the most valuable football players. *Journal of Management and Business Administration. Central Europe*, *24*(3), 91–104. https://doi.org/10.7206/JMBA.CE.2450-7814.177

Most Popular Sport by Country. (2022). World Population Review. https://worldpopulationreview.com/country-rankings/most-popular-sport-by-country

Mourao, P., & de Cima, C. (2015). Studying the Golden Generations' effects and the changes in the competitive balance of the Portuguese Soccer League. *International Journal of Sport Finance*, *10*, 42–61.

Mourao, P. R. (2016). Soccer transfers, team efficiency and the sports cycle in the most valued European soccer leagues – have European soccer teams been efficient in trading players? *Http://Dx.Doi.Org/10.1080/00036846.2016.1178851*, *48*(56), 5513–5524. https://doi.org/10.1080/00036846.2016.1178851

Müller, O., Simons, A., & Weinmann, M. (2017). Beyond crowd judgments: Data-driven estimation of market value in association football. *European Journal of Operational Research*, *263*(2), 611–624. https://doi.org/10.1016/J.EJOR.2017.05.005

Poli, R., Besson, R., & Ravenel, L. (2021). Econometric Approach to Assessing the Transfer Fees and Values of Professional Football Players. *Economies 2022, Vol. 10, Page 4*, *10*(1), 4. https://doi.org/10.3390/ECONOMIES10010004

Ramos, A. S., Hammerschmidt, J., Ribeiro, A. S., Lima, F., & Kraus, S. (2022). Rethinking dual careers: success factors for career transition of professional football players and the role of sport entrepreneurship. *International Journal of Sports Marketing and Sponsorship*, *23*(5), 881–900. https://doi.org/10.1108/IJSMS-02-2021-0029/FULL/XML

Reilly, T., & Gilbourne, D. (2008). *Science and football: a review of applied research in the football codes*. https://doi.org/10.1080/0264041031000102105

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

Ribeiro, A. S., Lima, F., Kraus, S., & Calabuig, F. (2021). Tournaments within football teams: players' performance and wages. *Http://Www.Tandfonline.Com/Action/AuthorSubmission?JournalCode=rero20&page=instructions*, *35*(1), 4884–4901. https://doi.org/10.1080/1331677X.2021.2019595

Ribeiro, A. S., Lima, F., Kraus, S., Calabuig, F., & Onio S Ergio Ribeiro, A. (2021). *Tournaments within football teams: players' performance and wages*. https://doi.org/10.1080/1331677X.2021.2019595

Romann, M., Javet, M., Cobley, S., & Born, D. P. (2021). How relative age effects associate with football players' market values: Indicators of losing talent and wasting money. *Sports*, *9*(7). https://doi.org/10.3390/SPORTS9070099

Serna Rodríguez, M., Ramírez Hassan, A., & Coad, A. (2019). Uncovering value drivers of high performance soccer players. *Journals.Sagepub.Com*, *20*(6), 819–849. https://doi.org/10.1177/1527002518808344

Simar, L., & Wilson, P. W. (1998). Sensitivity Analysis of Efficiency Scores: How to Bootstrap in Nonparametric Frontier Models. In *Source: Management Science* (Vol. 44, Issue 1). http://www.jstor.orgURL:http://www.jstor.org/stable/2634426

Singh, P., Sciences, P. L.-J. of D. M., & 2019, undefined. (2019). Influence of crowdsourcing, popularity and previous year statistics in market value estimation of football players. *Taylor & Francis*, *22*(2), 113–126. https://doi.org/10.1080/09720529.2019.1576333

Sports-Reference.com. (2022). *About Sports Reference*. https://www.sports-reference.com/about.html

Stolp, C. (1990). Strengths and weaknesses of data envelopment analysis: An urban and regional perspective. *Computers, Environment and Urban Systems*, *14*(2), 103–116. https://doi.org/10.1016/0198-9715(90)90016-M

swMATH. (2022). *Benchmarking - Mathematical software*. https://www.swmath.org/software/14630

Szymanski, Stefan., & Kuypers, Tim. (1999). *Winners and losers*. Viking.

UEFA. (2022). *Rankings do país | Coeficientes da UEFA*. https://pt.uefa.com/nationalassociations/uefarankings/country/#/yr/2022

Velema, T. A. (2019). Upward and downward job mobility and player market values in contemporary European professional football. *Sport Management Review*, *22*(2), 209–221. https://doi.org/10.1016/J.SMR.2018.02.004

Vicente Coll-Serrano, A., Bolos, V., Benitez Suarez, R., & Vicente Bolos, M. (2022). *Package "deaR" Type Package Title Conventional and Fuzzy Data Envelopment Analysis Version 1.2.6*. https://doi.org/10.1287/mnsc.30.9.1078

Zambom-Ferraresi, F., García-Cebrián, L. I., Lera-López, F., & Iráizoz, B. (2017). Performance Evaluation in the UEFA Champions League. *Journal of Sports Economics*, *18*(5), 448–470. https://doi.org/10.1177/1527002515588135

Zambom-Ferraresi, F., Lera-López, F., & Iráizoz, B. (2017). And if the ball does not cross the line? A comprehensive analysis of football clubs' performance. *Applied Economics Letters*, *24*(17), 1259–1262. https://doi.org/10.1080/13504851.2016.1270408

Zambom-Ferraresi, F., Rios, V., & Lera-López, F. (2018). Determinants of sport performance in European football: What can we learn from the data? *Decision Support Systems*, *114*(March), 18–28. https://doi.org/10.1016/j.dss.2018.08.006