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

Isaac Alberto Monteiro Felizardo

Dissertation to obtain the Master of Science Degree in

Industrial Engineering and Management

Supervisor:

Prof. António Sérgio Constantino Folgado Ribeiro

Examination Committee

Chairperson: Prof. Maria Margarida Martelo Catalão Lopes de Oliveira Pires Pina

Supervisor: Prof. António Sérgio Constantino Folgado Ribeiro

Member of the Committee: Prof. Fernando Martins

November 2022

Declaração

Declaro que o presente documento é um trabalho original da minha autoria e que cumpre todos os requisitos do Código de Conduta e Boas Práticas da Universidade de Lisboa.

Declaration

I declare that this document is an original work of my own authorship and that it fulfils all the requirements of the Code of Conduct and Good Practices of the Universidade de Lisboa.

Acknowledgments

First and foremost, I would like to thank Professor António Sérgio Constantino Folgado Ribeiro for his help, discussion sparking tips and final revision of the study.

Secondly, I would like to thank my mum and dad for giving me the opportunity to study and conclude my superior education – a privilege not all of us have.

I would also like to thank my dear university colleagues Dinis, José and Miguel for the support, study hours and final opinions regarding this work. Albeit indirectly, many people who are part of my life contributed to this work. Thank you, Luca, Alex, Julian, Miguel, Julia, Fábio, Mia, Julia, Avô Mário and Mariana.

I could not forget to thank Mafalda, Inês, Farrusca, Tina, Nico, Max, Bebé, Jejé, Belinha and Preto.

To Tatá and Mamã

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, team's 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.

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

Resumo

Futebol de associação, ou simplesmente futebol, é o desporto mais popular do mundo e de Portugal. Leva anualmente milhões de fãs aos estádios, ecrãs de televisão e rádios. É uma indústria milionário que chama a atenção pelas gigantescas quantias gastas em salários de jogadores e transferências.

Esta dissertação estuda, usando Análise por Envoltória de Dados, quão eficientemente estes recursos foram usados nas épocas de 2008/2009 até 2020/2021 da Primeira Liga – escalão máximo do futebol profissional português. Tenta também explicar a eficiência, os resultados desportivos e a assistência nos estádios.

Os resultados mostram que os clubes da Primeira Liga operaram sob retornos variáveis à escala. Enquanto todos os clubes que se qualificaram para competições europeias operavam com retornos decrescentes à escala, todos os despromovidos operavam com retornos crescentes à escala. O salário mais alto da equipa está altamente negativamente correlacionado com eficiência, enquanto o salário mais baixo está positivamente relacionado com esta. O coeficiente de variação dos salários de uma equipa está negativamente correlacionado com a eficiência, e, para dois dos três modelos conceptualizados, verifica-se com significância elevada a Teoria da Coesão. Para maximizar os resultados desportivos, as equipas devem investir em salários de jogadores e nos seus valores de mercado (i.e., valores de transferência) mas, dado um orçamento limitado, as equipas maximizam os seus resultados desportivos ao maximizarem a eficiência. A assistência aos estádios está relacionada com elevados salários, principalmente o salário do jogador mais bem pago, grandes valores de mercado, grandes diferenças de golo, vitórias e ineficiência das equipas.

Palavras-chave: Análise por Envoltória de Dados; Futebol; Eficiência; Salários; Teoria dos Torneios; Teoria da Coesão; Valores de Mercado Estimados.

Contents

List of Figures

List of Tables

List of Abbreviations and Acronyms

CrdR	Red Cards
CrdY	Yellow Cards
CRS	Constant Returns to Scale
D	Draws
DEA	Data Envelopment Analysis
DFA	Deterministic Frontier Analysis
DMU	Decision-making Unit
FIFA	Fédération Internationale de Football Association
GA	Goals Against
GD	Goal Difference
GF	Goals For
G-PK	Goals minus Penalty Kicks
IFAB	International Football Association Board
L	Losses
MLR	Multiple Linear Regression
MP	Matches Played
MV	Market Value
PG	Per Game
PK	Penalty Kicks
Pkatt	Penalty Kicks Attempted
PPG	Points per Game
QP	Quadros de Pessoal
Rk	Ranking
RTS	Returns to Scale
SFA	Stochastic Frontier Analysis
StoNED	The Stochastic Non-smooth Envelopment of Data
TotalAst	Total Assists

UEFA	Union of European Football Associations
VRS	Variable Returns to Scale
W	Wins

Chapter 1 – Introduction

This chapter aims at giving the reader an idea of what this dissertation is about and how it is structured. It has three sections. The Motivation provides the reader with what we are studying and why. The Objectives informs what questions we aim to answer. The Structure takes the reader through all the chapters of the dissertation and what to expect from each.

1.1 Motivation

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, its methodology and its results aim to add a small contribution to the existing literature in the field, as well as helping sports decision makers to make even better decisions.

1.2 Objectives

This dissertation 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?
- Are there variable returns to scale? If so, with what structure?
- How salary disparities influence team efficiency and performance?
- How can a team better leverage its resources to achieve maximum performance and efficiency?
- How is stadium attendance related to team performance and efficiency?

These objectives will be achieved by evaluating simplified versions of teams competing in Primeira Liga – our models.

1.3 Structure

The present document is comprised of six chapters plus the initial abstract, contents, list of figures, list of tables, list of abbreviations and, in the end, literature references and an appendix.

In the current and first chapter, the reader was given a brief summary of what the document is about and how it is structured. In the second chapter "Football and the Portuguese Primeira Liga", the reader is given a contextualisation of what football is, what football this thesis is about and what is the "Primeira Liga", whose teams we shall study. In the "Literature Review" chapter, the theoretical foundations are laid out as well as other examples of the same types of studies found in scientific literature. In the "Methodology" chapter, decisions are made and justified of what data to use, what models were conceived, what methodologies from the literature review were chosen and, most importantly, limitations of these choices are explained. Then, follows the "Results and Discussion", where the research questions are answered in a story-like manner to both the scientific reader and the sports decision-maker. Last, but not least, conclusions in "Conclusions and Future Work" the main findings originating from results discussions are synthetized and more ideas, many coming from the own limitations of the study, are suggested for further research in the future.

Chapter 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

This section explains the reader what football is, how many types of football exist and what type of football we are going to study and to elucidate to elucidate the reader on what Primeira Liga is and why it is relevant to our study

What is football and what type of football is this work about? Football is a group of team sports played with a ball that usually involve contact of the feet with the ball. There is proof that various types of football have been played since thousands of years ago with examples like the Chinese "Tsu' Chu", the Japanese "Kemari", the Greek "Episkyros" and the Roman "Harpastum". The romans eventually took the game to Britain where several types of football later emerged (FIFA, 2013b). In Britain and throughout the British Empire, several football codes emerged into what we nowadays call rugby union, rugby league, Australian rules football, gridiron football (American football), Gaelic football and association football (soccer) – the type of football this work is about (FIFA, 2013a) (Reilly & Gilbourne, 2008).

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)

In association football, Fédération Internationale de Football Association (FIFA), is the entity with the responsibility to "govern football and to develop the game around the world" (FIFA, 2022). It organizes, most notably, FIFA Men's World Cup, FIFA Women's World Cup plus other international competitions.

International Football Association Board (IFAB), composed by The Football Association, Wales Football Association, Scottish Football Association, Northern Ireland Football Association and FIFA, is the body responsible for the set of rules of the sport – known as Laws (IFAB, 2022). These Laws are specified in the rule book called The Laws of the Game (The International Football Association Board, 2022).

In this work, we will refer the sport “association football” merely as “football” for the sake of simplicity.

Even though half-century ago scientists were not welcomed within the football world, currently the connection between the academia and the effective practice of football is well established with several articles being written and notable literature revisions being made regarding several fields within football, such as players' characteristics, fitness, match demands, notation analysis, training, sport psychology and talent identification and development. This fact allows for this work to be enriched with “excellent literature” (Reilly & Gilbourne, 2008).

Not only has the football become more scientifically backed, but also more technological. In an era of game-related statistics, where data is king, several studies have tried to answer which are the most important sport specific game-related data points. One study of relevance (Zambom-Ferraresi, Rios, et al., 2018) analysed the five biggest European football leagues (Premier League, La Liga, Serie A, Bundesliga and Ligue 1), commonly known as the “Big Five”, and found out which technical and tactical actions had more explicative power over sports performance. Perhaps not surprisingly, different leagues had different major explanatory variables, but overall assists, shots conceded, saves made by the goalkeeper, passing accuracy, shots on target and shots conceded had the greatest power. Offensive actions explained the performance differentials more than defensive ones, and accuracy explained more than absolute number of attempts.

This however relates to team data, but player-specific data is also interesting to analyse (Memmert & Raabe, 2018).

For example, (Rajesh et al., 2020) analyse player data in depth in order to classify players according to position and recommend how to build a team from scratch using machine learning algorithms and clustering techniques.

Another great example, in (Weimar & Wicker, 2014), the authors applied the same logic behind Moneyball – that the baseball transfer market was undervaluing some skills – to football and found out that the total distance covered per match of a player was undervalued by football clubs. However, it is interesting to compare this study, which was confined to the German Bundesliga, and other studies about Serie A (Rampinini et al., 2009), Premier League (di Salvo et al., 2009) and La Liga (Clemente et al., 2019) that show, on a team-level and not player specific like (Weimar & Wicker, 2014), that the aggregated running distances do not have an impact in the league's final standings.

This takes us to the league under study – the Portuguese Primeira Liga. Primeira Liga, also known as Liga Portugal Bwin for sponsorship reasons, is the competition and whose teams this work will be studying. It is organized by Liga Portugal, previously known as Liga Portuguesa de Futebol Profissional – an organization created by the Portuguese football federation to manage and organize professional football in Portugal (Liga Portugal, 2022b). It is the highest league in the Portuguese league system and, currently, a total of 18 professional teams compete in it every year (until the 2014/2015 season, only 16 competed). Teams compete not only for the title of champion, but also for access to European competitions. Primeira Liga works with a system of relegation/promotion with the second-tier league "Liga Portugal SABSEG" (Liga Portugal, 2022a).

(Gomes Rocha, 2016) studied the Primeira Liga, namely the correlation between financial performance and sports performance of the biggest clubs in the league, concluding, among other things, that there was a positive relationship between sports results and financial results and that the league's clubs' revenue structure was converging with what was observable in other bigger European leagues.

On a European level, Primeira Liga is the 6th best league according to the official UEFA coefficients (UEFA, 2022), making it a relevant landscape for our study.

2.2 Problem Definition

This section aims at characterizing the problem after having a contextualisation of both the sport and the competitive setting being analysed.

To start explaining what systems (i.e., teams) we are studying, and, as summarized by (Espitia-Escuer & García-Cebrián, 2014), the works of (Carmichael et al., 2000) and of (Carmichael & Thomas, 2006) justify that the production function of a football team can be decomposed in two phases. In the first one, players' individual skills put together with the coaching staff's skills produce a certain level of performance during games in their attacking and defensive actions. In the second phase, these actions are translated into the level of success against the opposing teams (Figure 1).

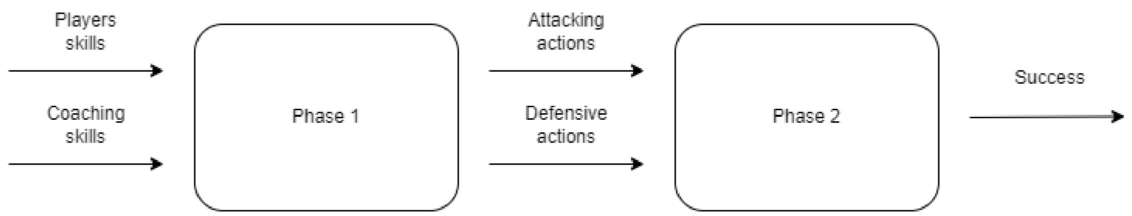


Figure 1: Production process of football teams by (Espitia-Escuer & García-Cebrián, 2014)

In this study, we are not looking exactly to what the team does on the pitch, but the sporting success of the team on a season level. One important aspect to consider in the case of football clubs is the dual objective their managers try to maximize, which (Zambom-Ferraresi, Lera-López, et al., 2017) explained best: there is a dichotomy in football clubs' objectives. Although sports results are the uncontested output of football clubs, their ability to generate income is also important, so there are both sportive and financial objectives – this should be taken into consideration when evaluating the clubs' efficiencies. For our study, we will be considering the above two phases as a black box as, using designed models, we will try to measure the efficiency of each team in transforming their salaries expenditures and squad market in sports success in the league (Figure 2).

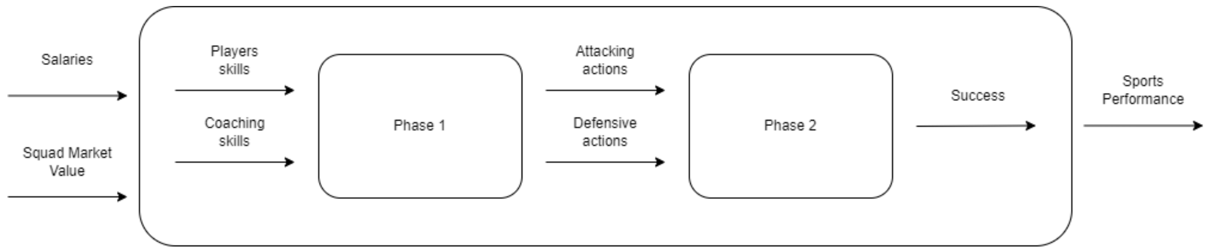


Figure 2: Process under study. Adapted from (Espitia-Escuer & García-Cebrián, 2014)

Now that it is clear what systems we are studying, we may proceed to develop and explain this study.

Chapter 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

This section introduces the reader to the concept of efficiency, how to estimate it and what studies have been made in the sports and football fields.

3.1.1 Efficiency definition

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". The author went on to define three types of efficiency:

- Technical Efficiency – "[a firm's] success in producing maximum output from a given set of inputs";
- Price Efficiency, also referred to as allocative efficiency – "a firm's success in choosing an optimal set of inputs";
- Overall Efficiency – "the product of the technical and price efficiencies"

Furthermore, (Leibenstein, 1966) proposed a distinction between allocative efficiency and "X-efficiency". While the first only concerns price-quantity inefficiencies due to monopolies, subsidies and other market distortions, the second is internal to the company and depends on many dimensions such as motivation and interpersonal relationships. The author explained that "firms and economies do not operate on an outer-bound production possibility surface consistent with their resources" for many reasons on the individual-level not considered by economists at the time.

- X-Efficiency (Frantz, 2019) – "inefficiency within the firm. It is shown as cost above the estimated cost frontier and output below the estimated output frontier"

Albeit used interchangeably (Button & Weyman-Jones, 1994), X-efficiency and technical efficiency are not the same. While the first has as unit the individual himself and the interactions of individuals to explain why firms do not operate optimally and a consistent methodology is still lacking, technical efficiency is only worried with the measurable deficiencies of transforming inputs into outputs considering the firm as a black box.

3.1.2 Efficiency measurement methods

3.1.2.1 Data Envelopment Analysis

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 input-oriented models try to minimize the number of resources needed for a given level of production. The output-oriented models look for the maximization of production for a given level of resources. Under the CRS assumption, both input and output-oriented models yield the same results, but the same is not true for the VRS assumption. Managers should opt for the version of the model which they feel they control best. If the manager has greater control over the resources used, he should opt for the input-oriented model; on the other hand, if the manager has greater control over the production, he should use the output-oriented one. (Cooper et al., 2004)

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:

DEA CRS Input-Oriented Model

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

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

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

DEA VRS Input-Oriented Model

$$\min \Theta \tag{10}$$

$$\sum_{j} \lambda_j x_{ij} \leq \Theta x_{i0}, i = 1,2,\dots,m; \tag{11}$$

$$\sum_{j} \lambda_j y_{rj} \geq y_{r0}, r = 1,2,\dots,s; \tag{12}$$

$$\lambda_j \geq 0 , j \in 1,2, \ldots , n \quad (9)$$

$$\sum_j \lambda_j = 1 \quad (13)$$

$$\lambda_j \geq 0 , j \in 1,2, \ldots , n \quad (14)$$

On the right, the VRS version is presented. The essential mathematical difference between the two models is the following constraint present in the VRS model:

$$\sum_j \lambda_j = 1 \quad (15)$$

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).

As with everything, DEA is also criticized by many authors for various reasons. The two most raised objections are, firstly, relatively to the (limited) available information regarding the production efficiency because efficiencies are computed in relation to the best performers because the DEA frontier is “the piecewise linear combination that connects the set of ‘best-practice observations’ ”(Casu et al., 2010), that is, the outermost observations on the production curve, and secondly, to the deterministic view or, said differently, the nonstochastic approach of how inputs are turned into outputs. Regarding the first objection, if DEA is used as a technique accounting for relative efficiency and not necessarily as a method for estimating production functions correspondences”, these worries fade away, which is what we aim to do in this study regarding football clubs. And relatively to the second objection, it is rather hard to distinguish random noise from inefficiency, but sensitivity analysis could provide a vehicle to test the model (Stolp, 1990).

One way of tackling DEA objections is by having Bootstrapped models being performed alongside it. These bootstrapping techniques resample the initial available data through several iterations and analyse the sensitivity of results to variations in the data. (Simar & Wilson, 1998)

As noted by the original author of DEA, the DEA methodology developed could be applied to any type of productive organization, from a “workshop to a whole economy” (Farrell, 1957). Fast-forward several years, and DEA has not only been thoroughly studied theoretically, but also applied and evolved in different directions such as network DEA (Lewis & Sexton, 2004) and two-stage DEA (Yang, 2006) to name a few. A survey of DEA applications (J. S. 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) and many more examples could be provided with several variations of DEA methodologies, using different inputs and outputs and different objectives of study.

3.1.2.2 Other frontier methods

Stochastic Frontier Analysis (SFA) is a method simultaneously proposed by (Aigner et al., 1977) and (Meeusen & van den Broeck, 1977) to study the efficiency of DMUs in transforming inputs to outputs. The difference regarding DEA is that it assumes a parametric function can be found that describes the production of outputs from inputs. Furthermore, it considers the error term to be composed of inefficiency and a symmetric component of statistical noise and/or outside shocks that could not be controlled by the DMU.

To start understanding the production, one should go back to the start and use the following ideal production function model for the i-th firm, as done in (Afriat, 1972):

$$y_i = f(x_i, \beta) \tag{16}$$

, where y_i is the maximum output obtainable with a non-stochastic vector x_i of inputs with β parameter vector to be estimated. Now, to allow for firms to operate inside the ideal production function, one must add an error term (Lovell, 1993):

$$y_i = f(x_i, \beta) + \varepsilon_i \tag{17}$$

, which, according to the works of (Aigner et al., 1977) and (Meeusen & van den Broeck, 1977), can be decomposed in two parts independently distributed. One is a symmetric error resulting from poor measurements or input quantities (v_i) while the other is one-sided error resulting from bad practices and the non-optimality of the production processes used (u_i).

$$y_i = f(x_i, \beta) + v_i + u_i \tag{18}$$

, where a normal distribution with zero mean $v_i \sim N(0, \delta_{v_i}^2)$ is assumed for v_i and an independently normal distribution truncated above at zero $u_i \sim N^+(0, \delta_{u_i}^2)$ is assumed for u_i.

$$y_i = f(x_i, \beta) + v_i + u_i \tag{19}$$

According to (Lovell, 1993), assuming the Cobb-Douglas technology is common for most econometric production frontier applications. The biggest danger of this is "confusing a nonconstant scale elasticity and nonunitary substitution elasticities with inefficiency". Therefore, our models will obey the following structure for the i-th club on the t period:

$$\ln(y_{it}) = \beta * \ln(x_{it}) + v_{it} + u_{it} \tag{20}$$

The biggest danger with the SFA methodology is the possibility of confounding the inefficiency present with the effects of misspecified functional forms for the production functions. (Lovell, 1993). Other weaknesses include the fact that SFA can only deal with multiple inputs to one output or vice-versa**,** although rankings are not very sensitive, the absolute technical efficiency values are, SFA requires many DMU's, and other weaknesses present in DEA also apply albeit in various degrees (Bezat et al., 2009).

Just as with DEA, SFA has the same wide range of applications. Some non-exhaustive examples include the healthcare sector (Rosko & Mutter, 2011), energy sector (Hatrori, 2002), agricultural sector (Y. Liu & Myers, 2009) (Mailena et al., 2014), container ports (Cullinane & Song, 2006).

Specifically, SFA has also been used in football (Barros et al., 2007; Dawson et al., 2016).

Several other frontier methods for estimating efficiencies could and should be mentioned as well, despite not being so popular as the aforementioned. It is interesting to mention two other methods – one for its legacy and the other for its novelty.

Deterministic Frontier Analysis (DFA), mainly replaced by DEA in literature, assumed a deterministic and parametric production function to measure the efficiency of units. (Hjalmarsson et al., 1996; Odeck, 2007).

The Stochastic Non-smooth Envelopment of Data (StoNED) aims at the best of both worlds between DEA nonparametric nature and SFA stochastic error form (Andor & Hesse, 2013; Kuosmanen & Kortelainen, 2012). It computes "the shape of the frontier without any assumptions about its functional form or smoothness", tackling SFA's greatest disadvantage and then looks for the potential error in efficiency estimation, tackling DEA's greatest weakness of incorporating errors into the efficiency estimates.

3.1.2.3 Ratio Analysis

Although mostly used in finance (Horrigan, 1968), Ratio Analysis is one the oldest ways of measuring efficiencies. It lies mostly on the assumption that one can observe a unit's efficiency by diving its output by its input:

$$Efficiency = \frac{Output}{Input} \tag{21}$$

In football specifically, it is also mostly used for doing the financial assessment of football clubs (Dimitropoulos, 2010) (Ecer et al., 2010), with some authors incorporating sports results in aggregated indices (Plumley et al., 2014).

3.2 Correlation measures and tools for explaining efficiency differences

This section presents tools used to analyse relationships in the data, namely regarding efficiency estimates.

Pearson correlation coefficient, also known as Pearson's *r*, is one of the most common correlation measurements in statistics (Freedman et al., 2007). It has a very simple formula for computing between an x and a y variable, with n observations (which variable is x or y is irrelevant to the result):

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \tag{22}$$

It is considered a parametric method, as it relies on a few assumptions such as the continuity of the data, linearity of the relationship between variables, data is normally distributed and no outliers are present.

Also notorious is the Spearman correlation, or Spearman's ρ, coefficient (Zar, 2005), which relies on the same assumptions of Pearson's *r*, except for the normality assumption and data can be ordinal as well – making it a non-parametric coefficient. It is also less sensitive to outliers present in the data. Instead of looking for linearity in the data, it looks for the monotonicity in it, by comparing rankings of data of each variable instead of the raw data itself.

$$\rho = 1 - \frac{6\sum d_i^2}{(n^3 - n)} \tag{23}$$

Where n is the sample size and di is the difference between the ranking of x_i and the ranking of y_i.

Spearman's correlation coefficient is preferable to the Pearson correlation coefficient when:

i. Data is ordinal
ii. Data is not normally distributed
iii. Outliers are present
iv. Data is non-linearly correlated, but has a monotonic relationship

One other suggested way to check for dependence between variables and compare rankings, is to use Kendall's Tau - τ coefficient (Kendall, 1938). Also known as the Kendall rank correlation coefficient, it is used to measure rank correlation. It is a non-parametric hypothesis test for statistical dependence. Again, it is non-parametric as it does not rely on assumptions regarding the distributions of both rankings. The τ coefficient can be defined as the ratio between the actual score – number of concordant pairs *c* minus discordant pairs *d*, denoted by *S*, and the maximum possible score of association between the rankings. (Nelsen, 2011)

S can be computed as:

$$S = c - d \tag{24}$$

And the maximum possible score for the two rankings as:

$$maximum\ possible\ score = c + d = \binom{n}{2} = \frac{n(n-1)}{2} \tag{25}$$

Finally, the formula for τ can comprehensibly be written like:

$$\tau = \frac{actual\ score}{maximum\ possible\ score} = \frac{2S}{n(n-1)} \tag{26}$$

One very important tool to analyse how one independent variable may be explained by two or more other variables is the Multiple Linear Regression (MLR) (Rubinfeld, 2011). Assuming:

i. A linear relationship between independent and dependent variables
ii. Low correlation among dependent variables
iii. Residuals have constant variance

iv. Residuals are independent
v. Residuals are normally distributed

One can find how and how much each independent variable explains the variation in the dependent variable. The generic formula is:

$$y_i = \beta_0 + \beta_1\, x_{i1} + \beta_2\, x_{i2} + \cdots + \beta_p\, x_{ip}\ + \epsilon \tag{27}$$

where y_i is the dependent variable, β_0 is the intercept, and β_p are the slopes for each x_{ip} independent variable. ϵ is the residual (model's error).

Multiple linear regression is great to understand how well a group of variables explains a dependent one, but what if we want to know how only one variable impacts the dependent one, while controlling for the others' effects? That is where partial correlation can be helpful (Rummel, 1976) – it allows for measuring the strength of the relationship between two variables controlling for a third (or more) that is (or are) numerically related to them. It is formalized as the correlation of the residuals of the linear regression of the two variables with the third that is being controlled and can be computed with the following formula:

$$r_{xy\cdot z} = \frac{r_{xy} - (r_{xy} r_{yz})}{\sqrt{(1 - r_{xz}^2)(1 - r_{yz}^2)}} \tag{28}$$

Where $r_{xy\cdot z}$ is the partial correlation between x and y, while controlling for z, and, generically, r_{ij} is the correlation between variable i and j.

3.3 Salaries and Market Value of Football Players

This section goes deep into the literature to explain the differences between players' salaries, their market values, the sources for both measures and what possible conclusions can be drawn from their distributions

3.3.1 Players Wages

Players, just like every other profession, earn salaries. Studies around sports players 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.

(Wilson et al., 2003) proved that players from South America and Western Europe increased ticket demand, which is interesting because foreign players are also paid more (Frick, 2006) and this proves that there is no space for discrimination in a highly competitive labour market.

Coming to European football, (Frick, 2006) found interesting salary-performance relationships. There was both a positive correlation between total team wage bill and sports performance and between head coach's salary and the team's performance. Probably surprisingly, the relative coach salary had a "linear positive and statistically significant impact on team performance".

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.

3.3.2 Tournament Theory versus Cohesive Theory

(Torgler & Schmidt, 2011) discovered that wage dispersion and relative income also played an important role in sports performance. This takes us back to the efficiency of football clubs in managing their wages because it is a perfect setting to test the tournament theory. 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). Since its foundational paper stating the superiority of structuring labor compensation schemes based on worker rank, in opposition to individual outputs (Lazear & Rosen, 1981), the theory has gained momentum. The theory helps to explain "the long right tails of wage distributions within firms"(Neal & Rosen, 2000), which is especially relevant for the case of executive compensation schemes and their ratio in relation to workers (Business et al., 2001; Chen et al., 2011; Eriksson, 2015). The tournament theory, or contest theory, was also empirically tested in sports as is the case of (Melton & Zorn, 2000b, 2000a). Although perhaps its effects are mostly felt in individual sports disciplines, such as golf or tennis (Szymanski, 2003), it is also possible to test in team sports.

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. This makes sense in team sports as “when work groups support the goals of the firm, firms will want to increase group cohesiveness”(Levine, 1991).

Different studies have had different results in different sports, in part related with how necessary cohesion is for the team success and the size of teams (Katayama & Nuch, 2009). 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. MLS is, nevertheless, a football league with characteristic specificities, such as the team salary cap which may help explain these results. 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. Also, the investigators found the pay structure affect the teams’ playing styles, with hierarchical pay structures having more individualistic actions. The authors concluded that football teams “should either have a strong culture of individualism and personal rewards or a culture of cooperation, teamwork and team-based rewards”.

3.3.3 Transfer Fees and Estimated Market Values

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). Several studies aimed at predicting which factors mattered the most for determining a given player transfer fee (Carmichael & Thomas, 1993; Coates & Parshakov, 2022; Dobson et al., 2010; Dobson & Gerrard, 1999; Garcia-del-Barrio & Pujol, 2020; Majewski, 2016; Ruijg & van Ophem, 2014). However, a study that is extremely noticeable for data used was done by the Football Observatory (Poli et al., 2021), in which the authors not only determined the most relevant factors impacting transfer fees but also successfully estimated multiple linear regression models for predicting them. They concluded a player's age, contract duration, international status, career progression, performance, the selling club's sporting level and economic level and inflation were the most important independent factors. The authors also point some limitations to their study such as the fact that it does not cover nonquantifiable factors such as leadership skills, or the selling "club's particular economic situation" or disagreements "between a player and his coach or fellow team members" or even "the superstar effect for very popular players" (Lucifora & Simmons, 2016). One other possible problem with using transfer fees in a study is the lack of transparency involved in such transactions as the values are not always verifiable.

One way to eliminate the previous limitations 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).

One of the biggest gaps between the concept of market value and transfer fee comes from the consequences of the European Court of Justice decision on the Bosman case – the ability of out-of-contract players to sign for a different club without the prior club receiving any fee (Antonioni & Cubbin, 2000).

Two sources of market values estimations worth mentioning are the German magazine Kicker and the consulting company KPMG. Despite being widely used in scientific literature, Kicker's estimates are only available for Bundesliga (German premier division) clubs (Coates & Parshakov, 2022). KPMG's Football Benchmark tool dataset covers only around 8300 players (Football Benchmark, 2022), as of the 25th of August 2022, and the full data is only available for premium users, i.e., a fee.

However, 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. Wisdom of the crowds is the “collective intelligence that arises when our imperfect judgments are aggregated” (Surowiecki, 2004) and it has been applied to sports scientific literature with impressive results, like predicting matches results based on ignorant crowds’ judgements (Herzog & Hertwig, 2011; Williams & Reade, 2020), betting tips (Brown & Reade, 2019; Brown & Yang, 2019), increasing team performance (Kim & Kim, 2019) and even to prepare video games features (Coates & Parshakov, 2022). In the specific case of *Transfermarkt.com*, players are evaluated in a forum and then pass through several levels of assessment, finally reaching senior members who validate predictions (Coates & Parshakov, 2022).

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) and in other studies as inputs, namely of DEA and SFA models (Pelloneová & Tomíček, 2022; Pyatunin et al., 2016; Zambom-Ferraresi, Iráizoz, et al., 2018; Zambom-Ferraresi, Lera-López, et al., 2017).

Chapter 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.

The first section explains methodological choices made. The second section explains what data was chosen, what were its sources and how it was processed. The third section goes into detail for the reader to understand how our calculations were implemented.

4.1 Methodological choices

This section aims to clarify the reader on the choices that were made and the limitations of these.

Because this study focuses on the efficiency of football teams from what is observable from the outside, we shall use the terms “technical efficiency” (Farrell, 1957) and “efficiency” interchangeably, while leaving deeper and meaningful studies regarding “X-efficiency” in football teams for future work to be done in the field.

We opted to use Portuguese first division teams due to the easiness of obtaining player statistics, team statistics and their obvious socio-economic relevance in comparison to lower tier leagues.

We chose to study the seasons between 2008/2009 and 2020/2021 for two reasons. Firstly, we started in the 2008/2009 season to pick up from the work already done for seasons 2002/03 to 2008/09 by (Ribeiro & Lima, 2012). Secondly, we tried to cover the longest temporal span possible to increase the robustness and relevance of the conclusions to be taken.

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. All these methods have been thoroughly explained in the literature revision chapter. SFA and Least-squares regression were not chosen due to the implications of assuming the same parametric production function for every football team in the league, despite all the heterogeneity this league presents. Ratio analysis would not be complex enough to develop models with more than one output and input, which would impoverish our study. 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.

We also needed to choose our input(s) and output(s) whose choices are explained and justified below.

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.

Other example found in the literature to tackle this sort of problem is using the ratio between points won and available points to be won (Forrest & Simmons, 2002). The interpretation of the results does not change, but it was preferred to use values closer to the reality of football (between 0 points for a loss and 3 points for a win) than a generic ratio value falling in the [0;1] interval.

A main limitation of the choice of this output for football teams in the Portuguese Primeira Liga is twofold. For one side, we are not considering the success of teams in attaining a spot in a European competition (positive output) or even their relegation to the second tier of professional football in Portugal (negative output). Another limitation is not considering the financial performance of teams in their success, which is partly correlated with the possible outputs referred above.

To study the relationship between variables and results, we chose to compute correlation and partial correlation computations where necessary. To avoid not satisfying one or more of the Pearson's *r* required assumptions and drawing the wrong, Kendall's *τ* and Spearman's *ρ* were the chosen coefficients to accompany the first to determine the correlation between variables, whenever possible. These will be of uttermost importance in understanding the importance of certain relationships in the data as a whole and to compare our results with past studies mentioned in the literature review chapter.

Multiple linear regressions were performed whenever we wanted to explain one variable/result using the other variables/results of our study, due to their simplicity and easiness of understanding.

4.2 Data Collection and Processing

This section explains the reader what data was fetched and how. It also delves into the limitations of the data and its gathering process.

To every type of data collected and its respective source, potential errors may be present, despite the excellence for which they are known scientifically and commercially.

4.2.1 Wages

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. Football clubs competing in the Primeira Liga are part of the three hundred and fifty thousand employers available and their football players and coaches are among the more than three million employees tracked. 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).

Wages data were retrieved for the seasons between and including 2008/2009 and 2020/2021. In total we could correctly collect data from 136 different season-club pairs out of 222 possible. It is a limitation that our dataset does not cover the entire universe of teams playing the Primeira Liga in the time period under study and that the most easily identifiable teams are the bigger ones, thus creating a data imbalance regarding team size.

Only values adjusted to 2020 salary levels were used because, as explained afterwards, the goal was to achieve maximum comparability of data in time.

Table 1: Players' wages variables

Variable	Description
max_player_real_wage	Maximum player salary of the team-season
max_coach_real_wage	Maximum coach salary of the team-season
mean_player_real_wage	Mean player salary of the team-season
mean_coach_real_wage	Mean coach salary of the team-season
mdev_player_real_wage	Mean deviation player salary of the team-season
mdev_coach_real_wage	Mean deviation coach salary of the team-season
min_player_real_wage	Minimum player salary of the team-season
min_coach_real_wage	Minimum coach salary of the team-season
p10_player_real_wage	10^{th} percentile of players' salaries of the team-season
p25_player_real_wage	25^{th} percentile of players' salaries of the team-season
p50_player_real_wage	50^{th} percentile of players' salaries of the team-season
p75_player_real_wage	75^{th} percentile of players' salaries of the team-season
p90_player_real_wage	90^{th} percentile of players' salaries of the team-season
total_player_real_wage	Total players' salaries of the team-season
total_coach_real_wage	Total coaches' salaries of the team-season

These data were collected during the month of August 2022.

4.2.2 Estimated market values

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). The only variable taken from this source was the team's total market value (TotalMV). Not having access to the total amount spent in fees and contract premiums, it was decided to use crowd-sourced estimated market values to have a proxy of how much the squads of each team were worth for each season. Using estimated player market values as an indicator for the squad as an asset may not correspond to the actual amount invested in transfer fees or the ability to convert these players' contracts into transfer fees because, although "comparable", crowdsourced estimates overlook important factors such as contract length and other contextual aspects (Serna Rodríguez et al., 2018). Furthermore, the "crowd-sourced metric tends to underestimate the value of the player" (Coates & Parshakov, 2022).

Table 2: Players' Market Values variable

Variable	Description
TotalMV	Total estimated market value for the team-season

Coherently, team market value data were retrieved, using a python webscraping library called "BeautifulSoup" (Richardson, 2019), for the seasons between and including 2008/2009 and 2020/2021. These data were collected on the 4th of April 2022.

4.2.3 Team sports statistics

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).

Table 3: Team sports statistics variables

Variable	Description
Rk	Final Ranking of the team in the season *Removed due to having a proxy of the normalized value
MP	Number of Matches Played in the season *Removed after normalizing other variables
W	Number of Wins in the season
D	Number of Draws in the season
L	Number of Wins in the season
GF	Goals for the team *Removed due to having a normalized value
GA	Number of goals against during the season *Removed due to having a normalized value
GD	Total goal difference during the season
Attendance	Average attendance at the team's venue in the season *Not considered in 2020 because Attendance was 0 due to Covid-19 pandemic restrictions.
#Pl	Number of players used in the season
Age	Average age of players used
TotalAst	Total number of Assists of the team-season *Removed due to having a normalized value
G-PK	Goals minus penalty kicks *Removed for high correlation with GF that also has a normalized value
PK	Number of penalty kicks converted during the season
PKatt	Number of penalty kicks attempted during the season *Removed for high correlation with PK
CrdY	Number of yellow cards attributed during the season *Removed due too many outliers being present and having too many missing values
CrdR	Number of yellow cards attributed during the season

	*Removed due too many outliers being present and having too many missing values
Gls	Goals per game
Ast	Assists per game
GoalkeepersPl	Number of goalkeepers used in the season
GAPG	Goals Against Per Game
SoTA	Shots on Target Against
Saves	Number of Saves in the Season
Save%	Save % during the season
CS	Number of clean sheets in the season *Removed due to having a normalized value
CS%	Clean sheet % in the season
SoT	Shots on Target during the season *Removed due to having a normalized value
SoTPG	Shots on Target per game
G/SoT	Goals per Shots on Target
Points	Total Points achieved during the season
Outcome	Outcome of the league in the case the team went to an European competition, maintained itself in the league or was relegated

Once again, sports statistics by team were retrieved for the seasons between and including 2008/2009 and 2020/2021. These data were collected on the 7th of April 2022.

4.2.4 Data Profiling and Preparation

Data manipulation and cleaning was done both in the Microsoft Office Excel software and using the famous data manipulation python library “pandas” (McKinney, 2011).

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 “TotalMV” 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.

As previously stated, to be able to compare the outputs of teams from when Primeira Liga had 16 teams instead of 18 (starting on the 2014/2015 season), a new feature was created – Points Per Game (PPG), this way normalizing teams' sport success throughout the studied seasons:

$$PPG = \frac{Points}{MP} \tag{29}$$

Other variables were also normalized to the number of matches played in order to become comparable in time. "W", "D", "L", "GD", "#Pl", "PK", "GoalkeepersPl", "SoTA" and "Saves" were all transformed into a "per game" basis, respectively yielding "WPG", "DPG", "LPG", "GDPG", "PlayersUsedPG", "PKPG", "GoalkeepersPlPG", "SoTAPG" and "SavesPG", respectively.

In order to compare our results and data with (Franck & Nüesch, 2010) findings, we needed to create a measure of salary disparity. In their study, they used both the intra-team Gini Index and coefficient of variation. Because we only have available the standard deviation and mean of the salaries but not individual salaries, we created a new variable "coefficient_of_variation_player_real_wage" that corresponds to the coefficient of variation of players' salaries in a certain team:

$$coefficient\_of\_variation\_player\_real\_wage = \frac{mdev\_player\_real\_wage}{mean\_player\_real\_wage} \tag{30}$$

The same coefficient was computed for the coaching staff, creating the variable "coefficient_of_variation_coach_real_wage".

As a comparison with the study of (Frick, 2006), a new variable "total_coach_to_total_player_real_wage" was created to verify if there is any relationship between the coaches' salaries and players' salaries.

$$total\_coach\_to\_total\_player\_real\_wage = \frac{total\_coach\_real\_wage}{total\_player\_real\_wage} \quad (31)$$

A variable "TotalMV_to_total_player_wage_real" was also created to understand how a ratio between the estimated market values and players total salaries may impact team efficiency.

$$TotalMV\_to\_total\_player\_real\_wage = \frac{TotalMV\_real}{total\_player\_real\_wage} \quad (32)$$

For preparing the salaries and estimated market values of players for the DEA models, we referred to (Sarkis, 2007) proposed methodology. To tackle possible problems with imbalance in data magnitudes, we transformed real squad market values and real total player wages in order to be closer to PPG magnitude. So, "TotalMV_real" was multiplied by 10^{-7} and "total_player_real_wage" by 10^{-5}, thus creating new variables for use in the DEA models:

$$TotalMV\_real\_10\_7 = TotalMV\_real * 10^{-7} \quad (33)$$

$$total\_player\_real\_wage\_10\_5 = total\_player\_real\_wage * 10^{-5} \quad (34)$$

As explained by the author, results remain unchanged, but possible rounding errors are avoided at a software level. We confirmed only strictly positive values were present – no negative or zero values. Models were run without missing any missing data on the input or output side.

To sum up, the following variables were created:

Table 4: Variables created as part of data processing

Variable	Description
TotalMV_real	TotalMV values adjusted for 2020 levels
PPG	Points per game
WPG	W per game

DPG	D per game
LPG	L per game
GDPG	GD per game
PlayersUsedPG	#Pl per game
PKPG	PK per game
GoalkeepersPlPG	GoalkeepersPl per game
SoTAPG	SoTA per game
SavesPG	Saves per game
coefficient_of_variation_player_real_wage	Intra-team coefficient of variation of players' salaries
coefficient_of_variation_coach_real_wage	Intra-team coefficient of variation of coaches' salaries
total_coach_to_total_player_real_wage	Ratio of total coaches' salaries versus total players' salaries
TotalMV_to_total_player_real_wage	Ratio between
TotalMV_real_10_7	TotalMV_real with magnitude adjusted for DEA models
total_player_real_wage_10_5	total_player_real_wage with magnitude adjusted for DEA models

Non-normalized variables were removed when normalized ones were created as these new ones better represented teams' performances.

Decision-making units were identified by their team names and first year of the season.

4.3 Implementation details

This section guides the reader through all the implementation stages in a way that, if wanted, the reader could reproduce exactly the study performed.

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:

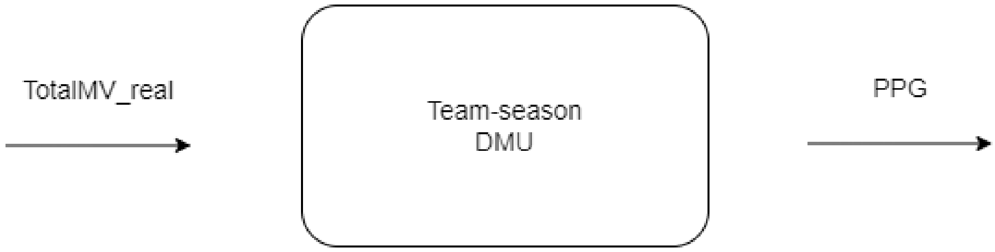


Figure 3: Schematic representation of model M1

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

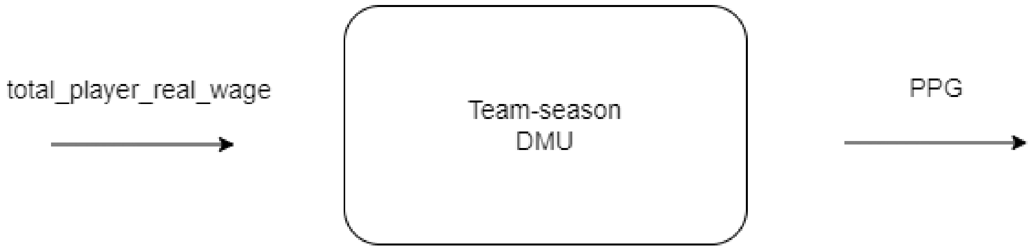


Figure 4: Schematic representation of model M2

Lastly, one third model was developed (M3) which combines both the total estimated market value of the squad and the total wages of players (real values) as inputs:

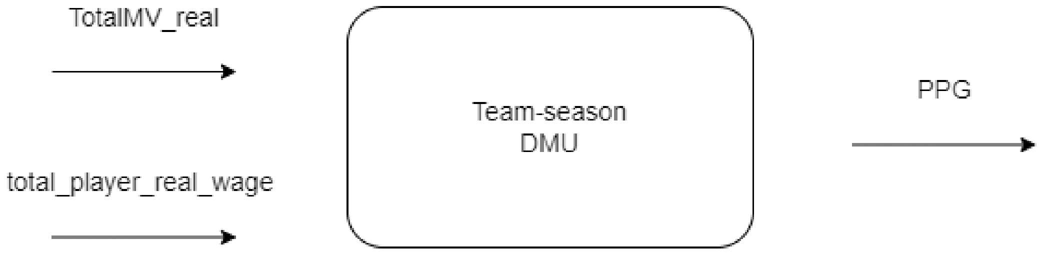


Figure 5: Schematic representation of model M3

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.

After having computed and tested the robustness of the efficiencies, we investigated if there are variable returns to scale and, if so, how the returns to scale vary.

Then, it was time to find the drivers of efficiency and try to explain it as much as possible, for which we will investigate correlations between variables and perform multiple linear regressions. For correlations, multiple linear regressions and other mathematical operations with variables and results, python libraries NumPy (Harris et al., 2020), SciPy (Virtanen et al., 2020) and Pingouin (Vallat, 2022) libraries were used.

To determine which variables to include in the MLRs both an analysis on variables correlation with each efficiency result and stepwise linear regression methods were used. Estimated market values and salaries had to be logarithmized to reach acceptable significant levels. Different combinations of variables were tested to reach the most robust regressions possible.

Next, we will delve into the (possible) relationships between intra-team salary inequality and how it may affect teams' performance and efficiency.

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. Again, resorting to studies on correlations and multiple linear regressions. As seen in Figure 6, 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.

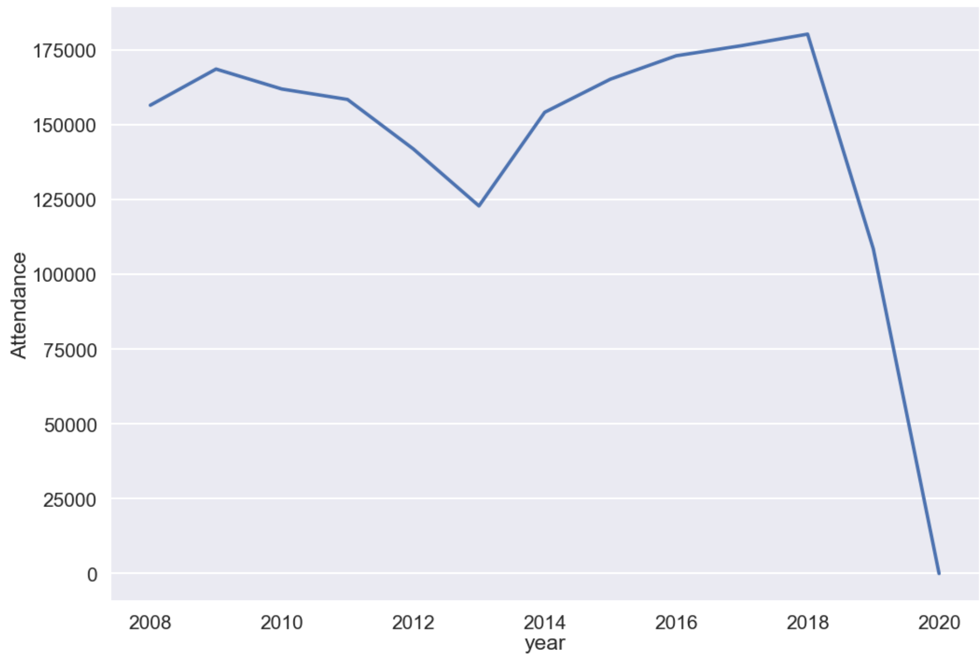


Figure 6: Attendance evolution from 2008/2009 to 2020/2021

Lastly, but most importantly, we will investigate how teams can maximize their average points won per game and how this metric is correlated with efficiency.

Chapter 5 – Results and Discussion

The results of the study are organized in a logical and chronological format (resembling a story) in order to help football decision-makers – especially of the Portuguese Primeira Liga – to understand how they can improve their team's management and options in order to achieve maximum efficiency. Many interesting results addressed in the literature review are revisited and their validity tested for the current data. Most importantly, no research questions are left unanswered, and several hypotheses are tested.

5.1 Data Envelopment Analysis

This section shows, describes and comments the results of the DEA calculations in both CRS and VRS assumptions

The results of the three different models, both for CRS and VRS DEA calculations (always input-oriented) can be found in the Appendix in tabular form for the eager and analytical reader, where results starting with "dea" represent technical efficiencies for the respective model, those starting with "RTS" an indication of where on the returns to scale curve each DMU is and the "RTS ratio" how far they are from optimal scale efficiency.

These results are more easily understood when seen graphically in the usual DEA plots. For M1 and M2 models, we retrieved the estimated efficient frontiers.

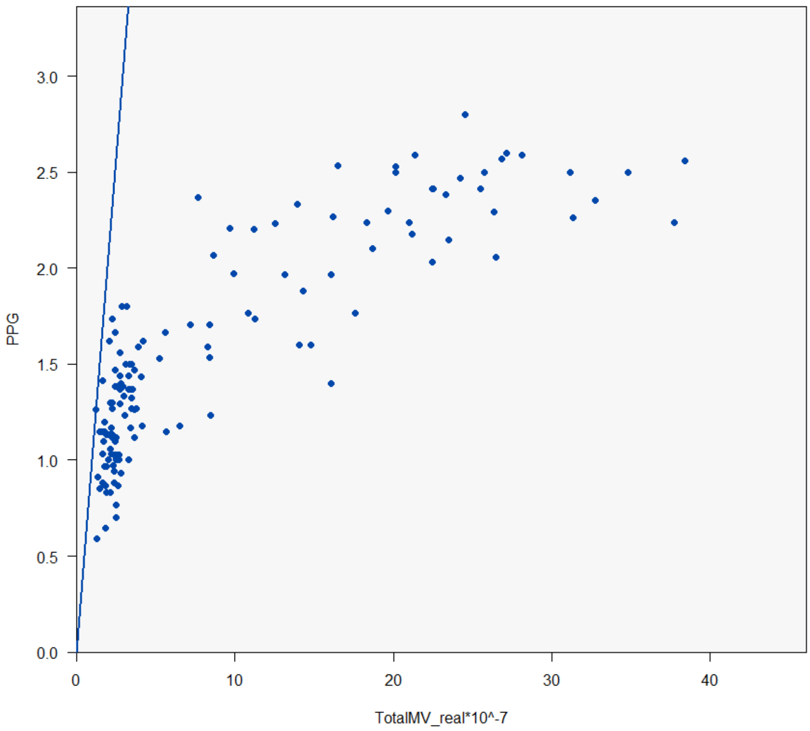


Figure 7: Efficient frontier for M1 (assuming CRS)

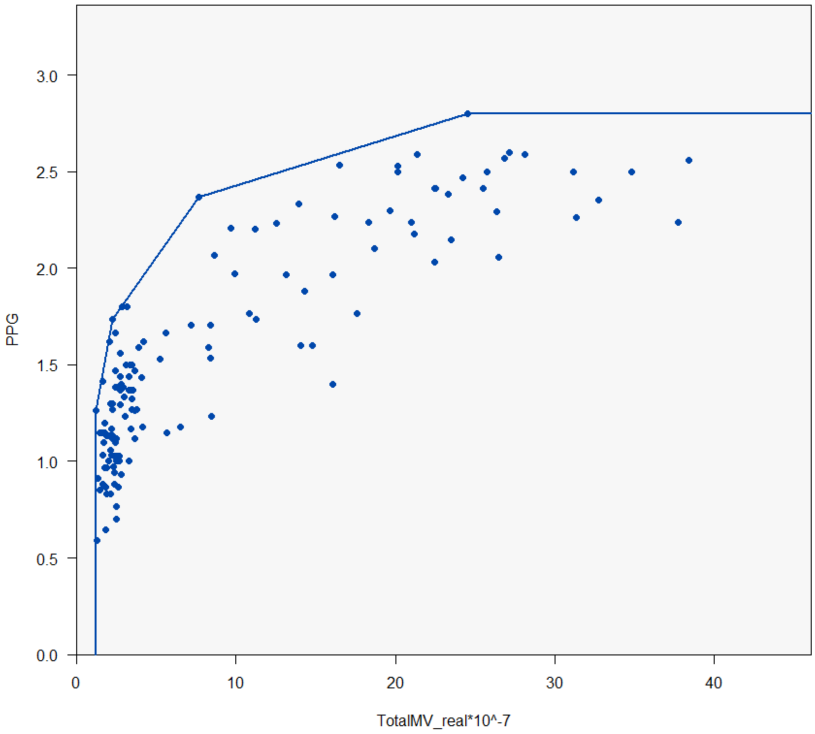


Figure 8:Efficient frontier for M1 (assuming VRS)

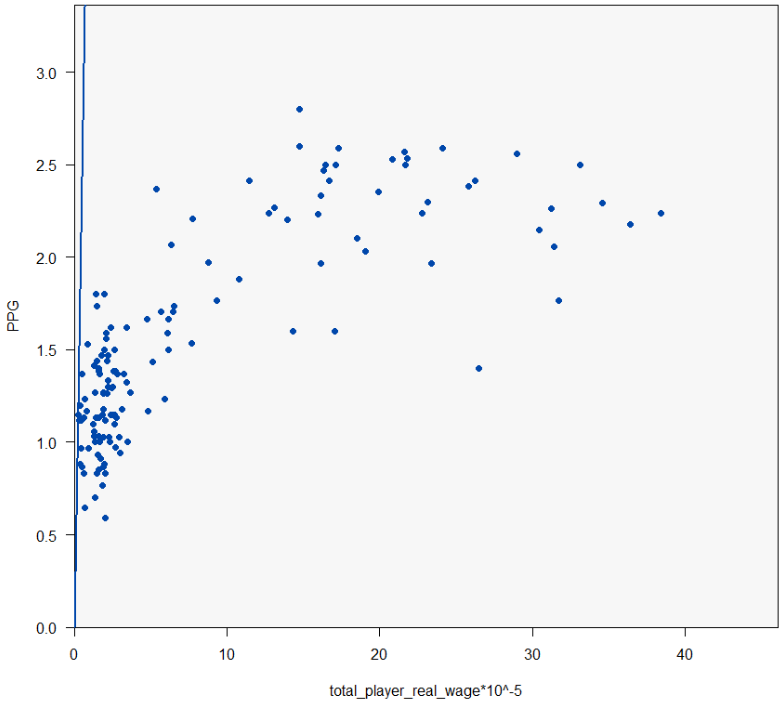


Figure 9: Efficient frontier for M2 (assuming CRS)

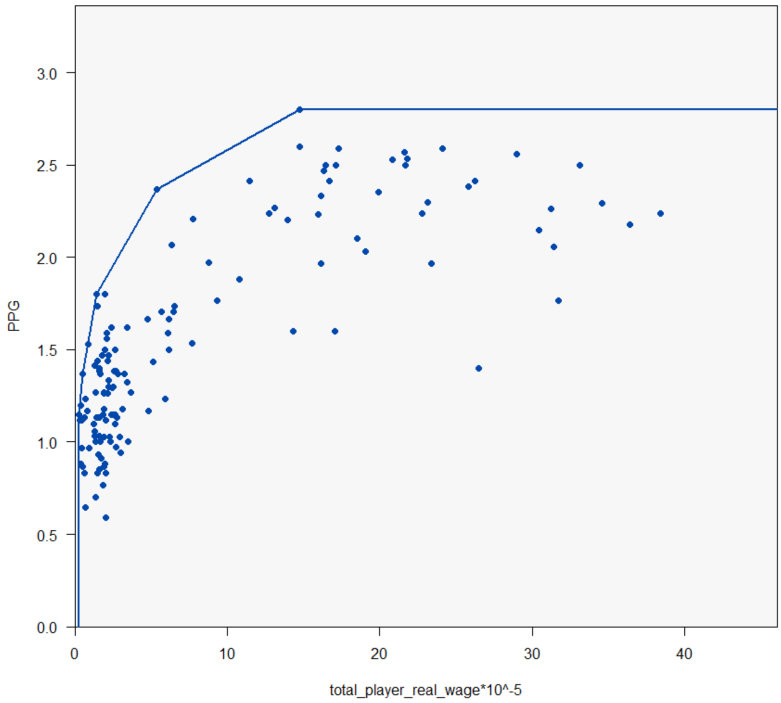


Figure 10: Efficient frontier for M2 (assuming VRS)

As easily seen by the plots above in the DEA-CRS models (Figure 7 and Figure 9), the constant returns to scale assumption completely prohibits top spending teams of ever achieving efficiency. The DEA-VRS models (Figure 8 and Figure 10) 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.

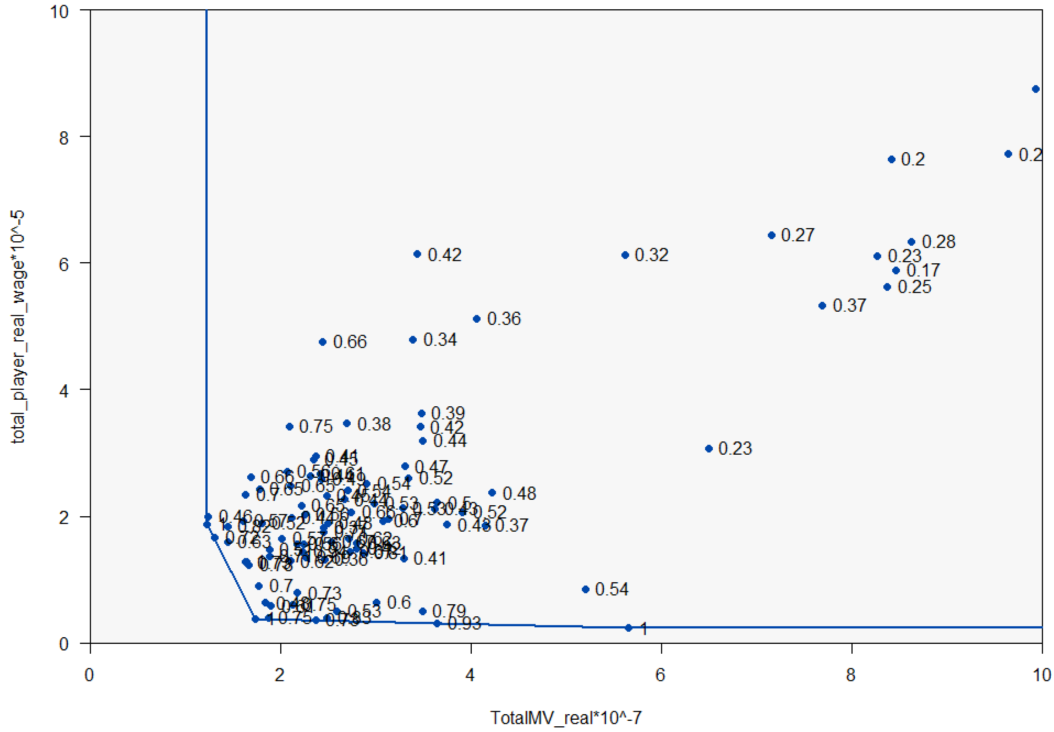


Figure 11: Production isoquant for M3 (assuming CRS)

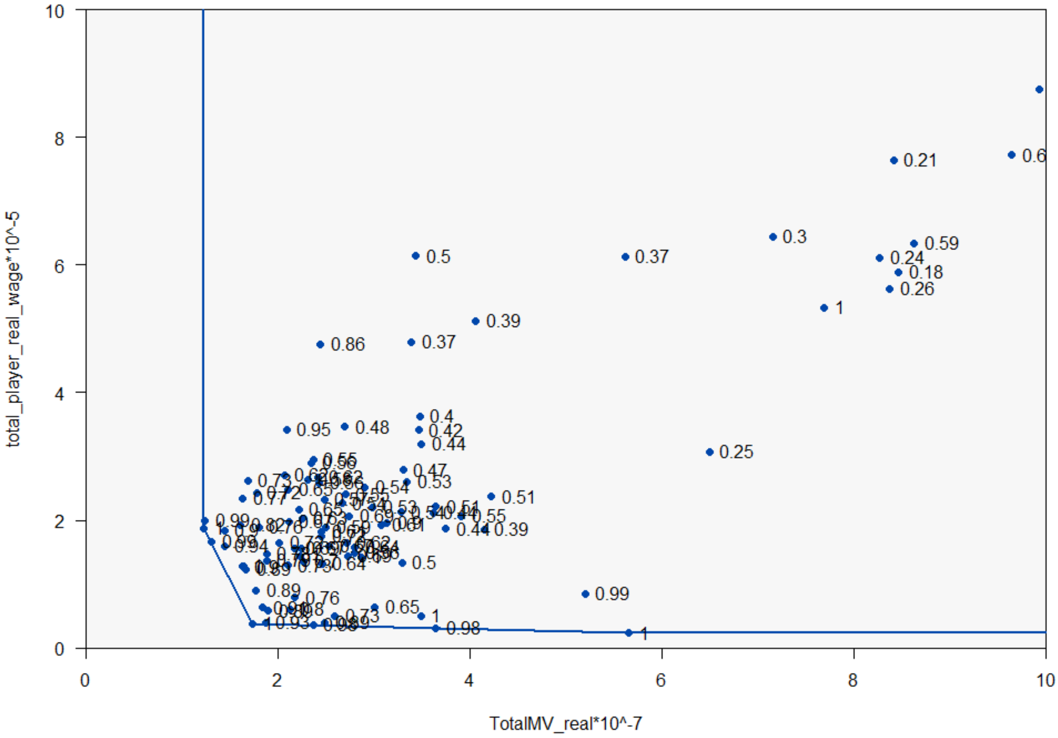


Figure 12: Production isoquant for M3 (assuming CRS)

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 5):

Table 5: 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

5.2 Returns to Scale

This section discusses the existence of variable returns to scale and the consequences for the remaining study.

As promised in the Methodology chapter, efficiencies were computed for constant and variable returns to scale to enrich the study and discussion. Even though the R software package “deaR” offered the calculations of RTS for each DMU, it is interesting to see, in the charts below, how biased the Constant Returns to Scale results are, making it impossible for any top-spending team to achieve efficiency. The inputs scales were logarithmized to make it easier for the reader to understand the distributions.

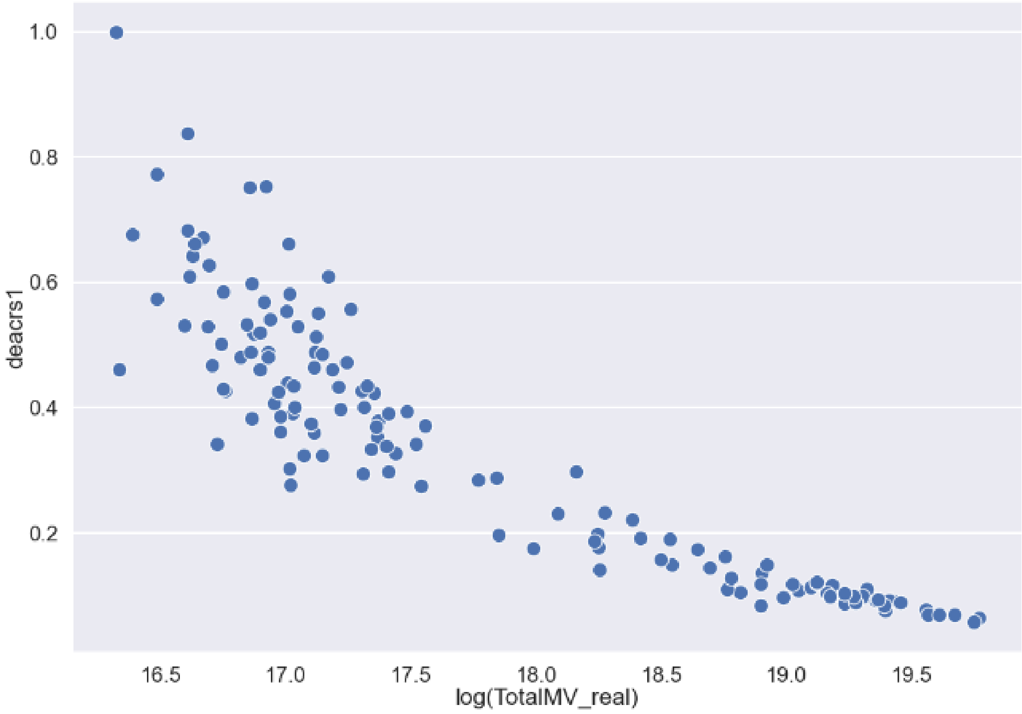


Figure 13: Graphical representation of the distribution between the logarithmized total real market values and deacrs1

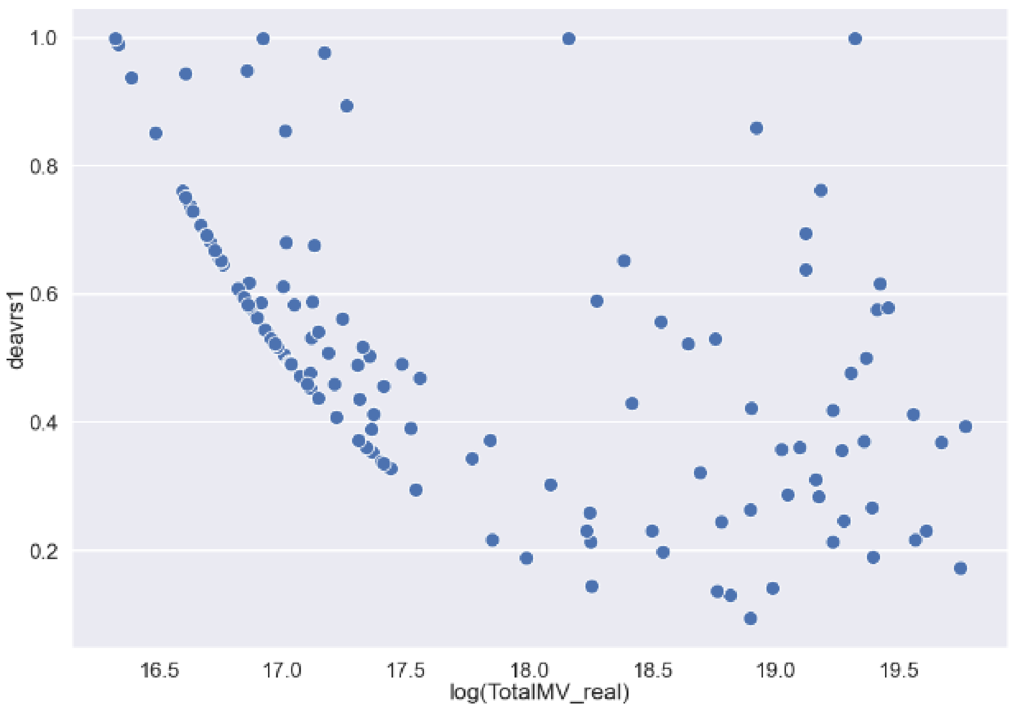


Figure 14: Graphical representation of the distribution between the logarithmized total real market values and deavrs1

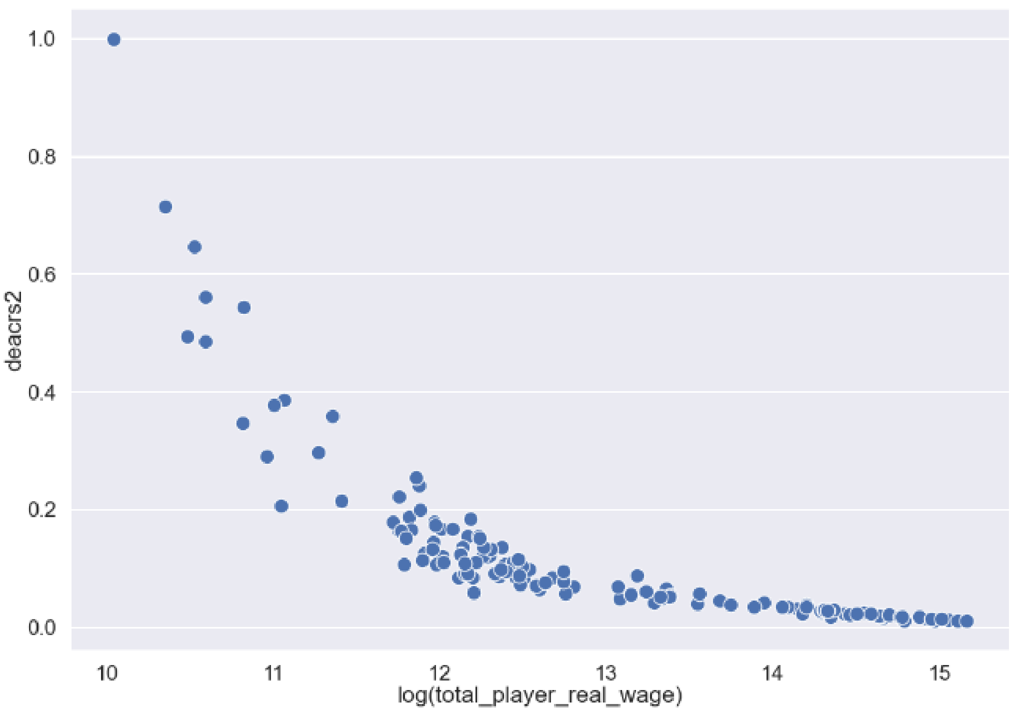


Figure 15: Graphical representation of the distribution between the logarithmized total player real wages and deacrs2

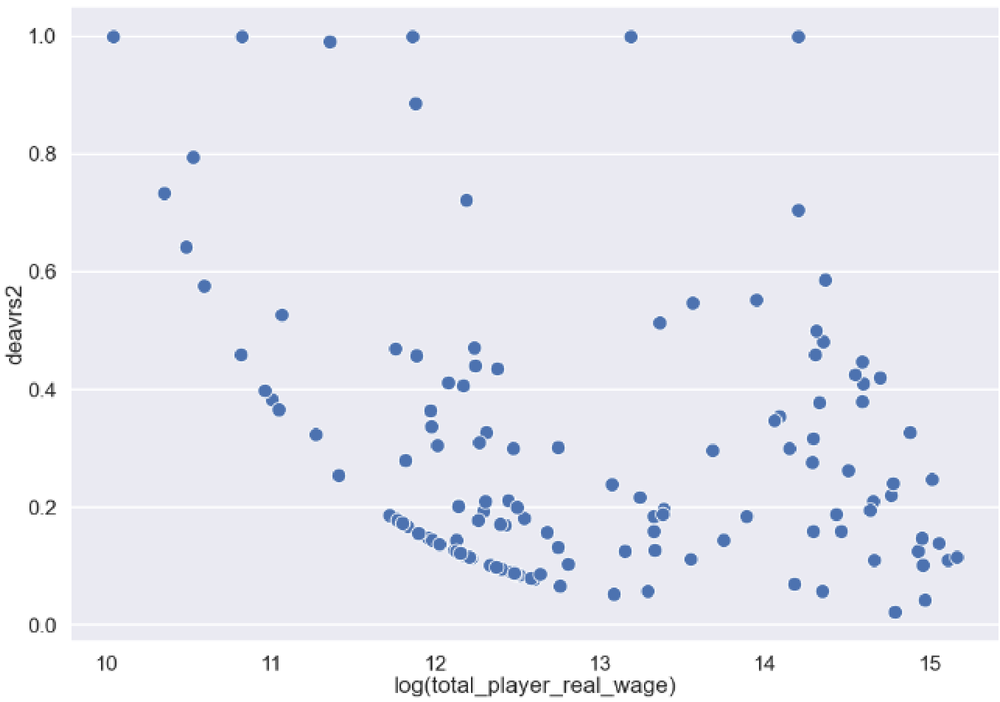


Figure 16: Graphical representation of the distribution between the logarithmized total player real wages and deavrs2

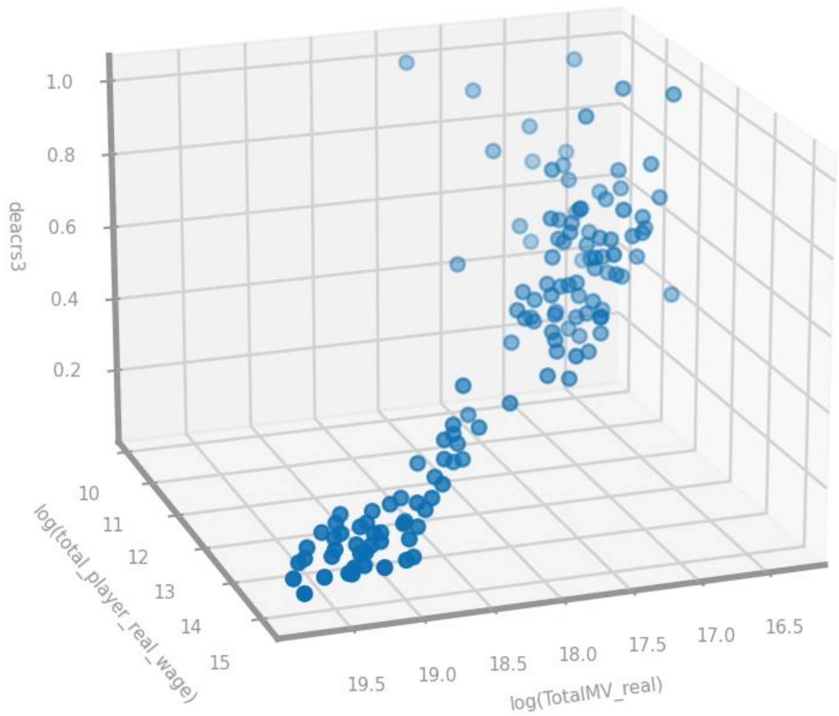


Figure 17: Graphical representation of the distribution between the logarithmized total real market values, the logarithmized total player real wages and deacrs3

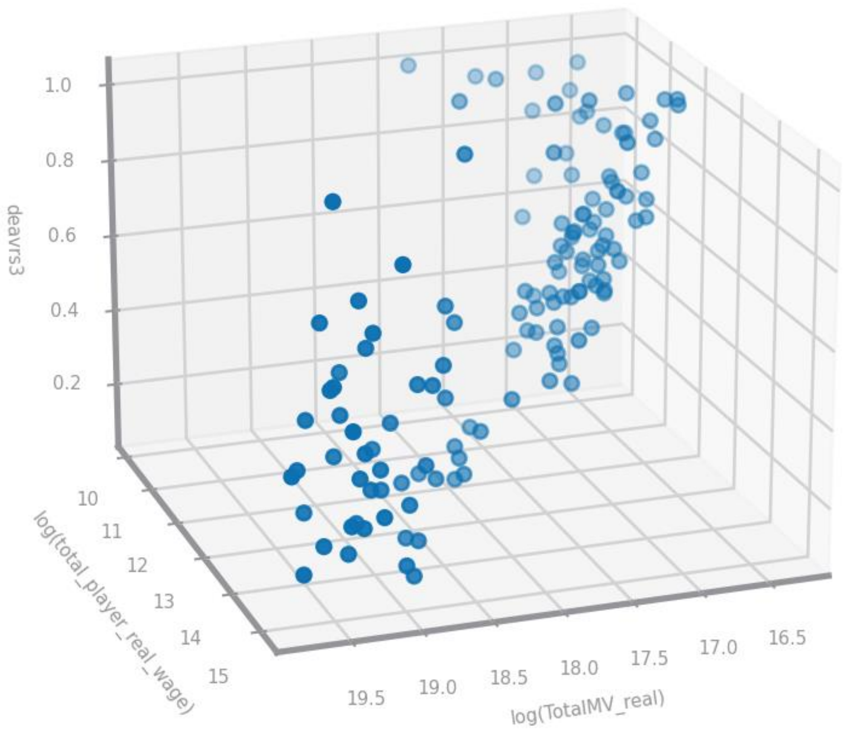


Figure 18: Graphical representation of the distribution between the logarithmized total real market values, the logarithmized total player real wages and deavrs3

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 19 and Figure 20):

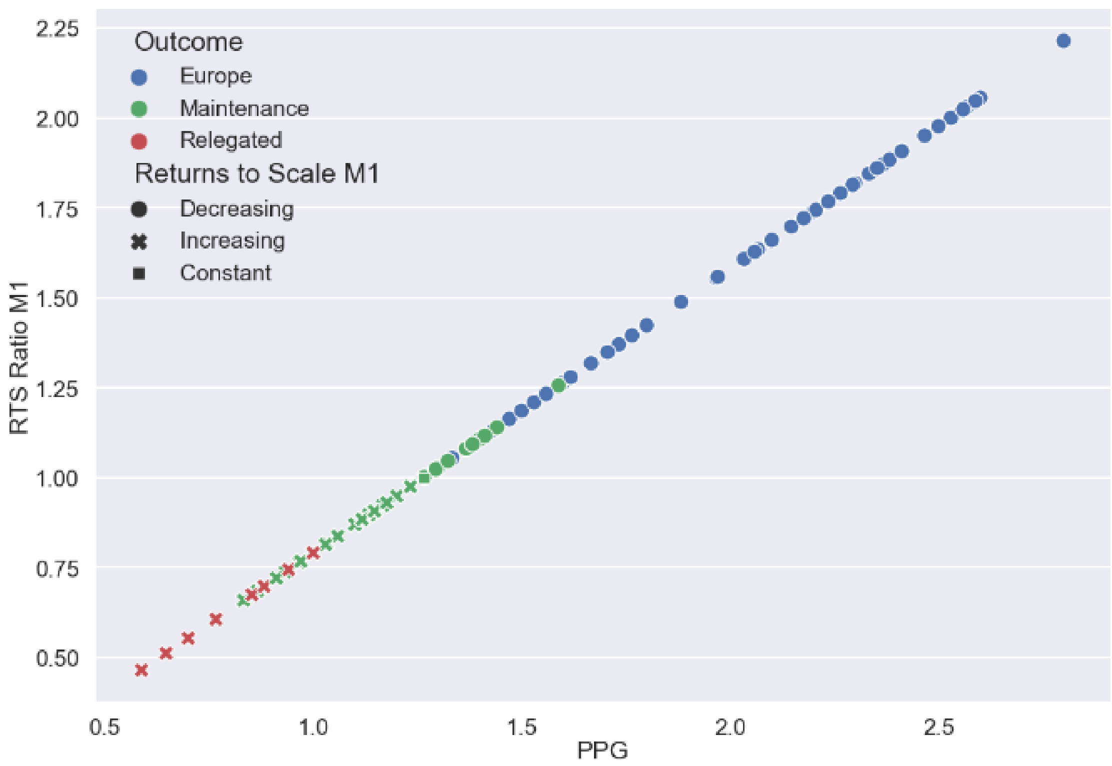


Figure 19: PPG and scale efficiency of M1

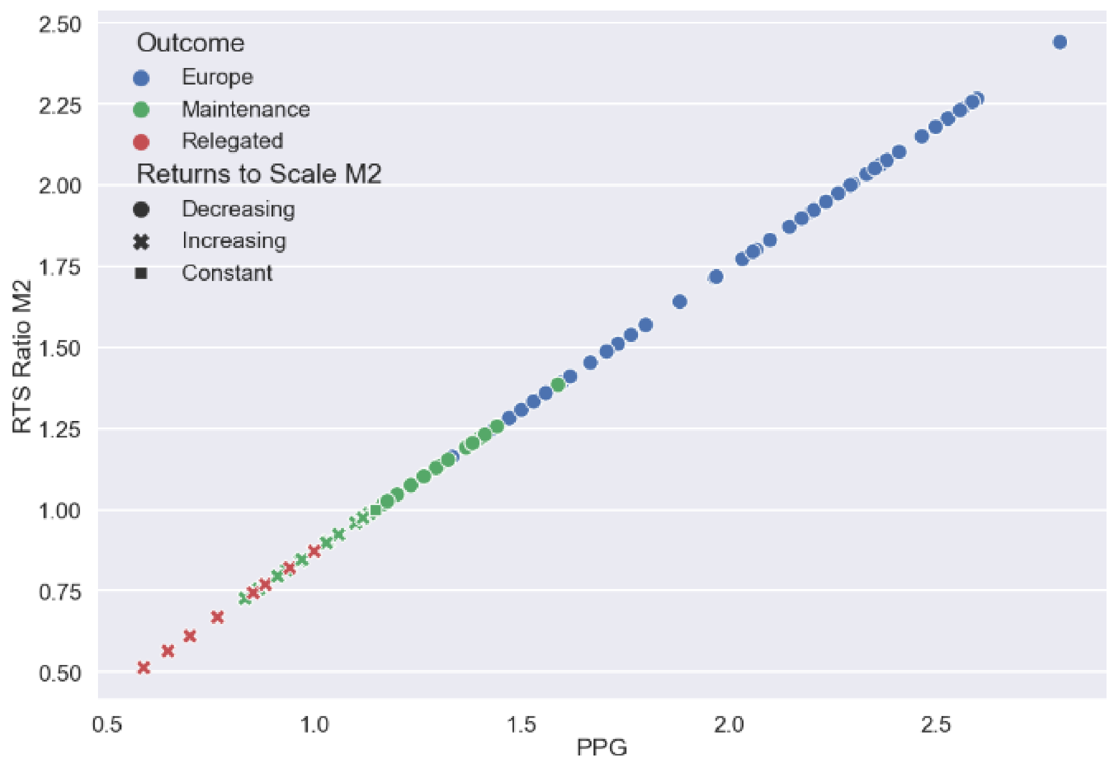


Figure 20: PPG and scale efficiency of M2

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 21):

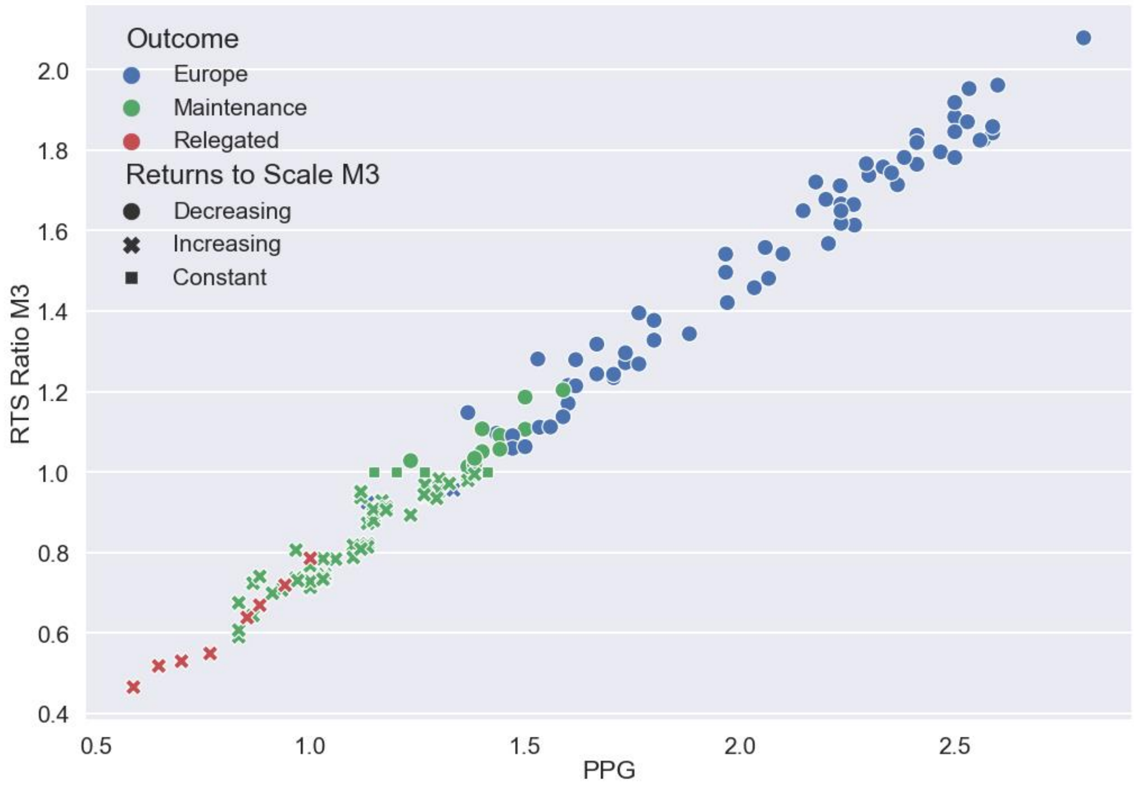


Figure 21: PPG and scale efficiency of M3

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.

5.3 Explaining Efficiency

This section displays the best-fitting MLR regressions to each model efficiency estimates and draws conclusions from its interpretation.

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.

5.3.1 MLR on Model 1

	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. The standard error of coefficient is between parentheses.

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.

5.3.2 MLR on Model 2

	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

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. The unitary variance in the coefficient of variation of players wages has an effect of 0.141 in "deavrs2", ceteris paribus. The standard error of coefficient is between parentheses.

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.

5.3.3 MLR on Model 3

	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 coach real wage)	0.069***
	(0.026)
Age	0.027**
	(0.012)
log(min player real wage)	0.052**
	(0.025)
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

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 minimum player wage had positive impacts on teams' efficiencies, while players' age positively impacted it again. Relegated teams were again more efficient.

The average number of players used negatively impacted efficiency again. The unitary variance in the average number of players used had an effect of -0.355 in the dependent variable, ceteris paribus. The standard error of coefficient is between parentheses.

5.3.4 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.

5.4 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.

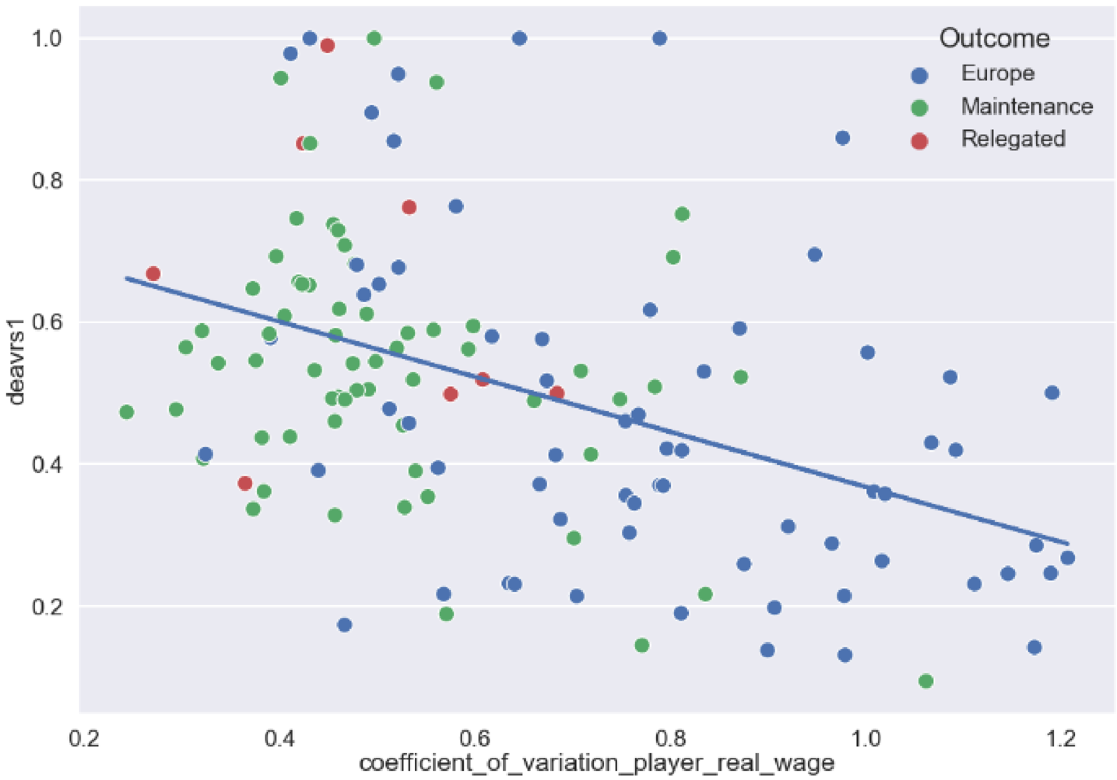


Figure 22: Linear regression between coefficient of variation of players' real wages and deavrs1

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.

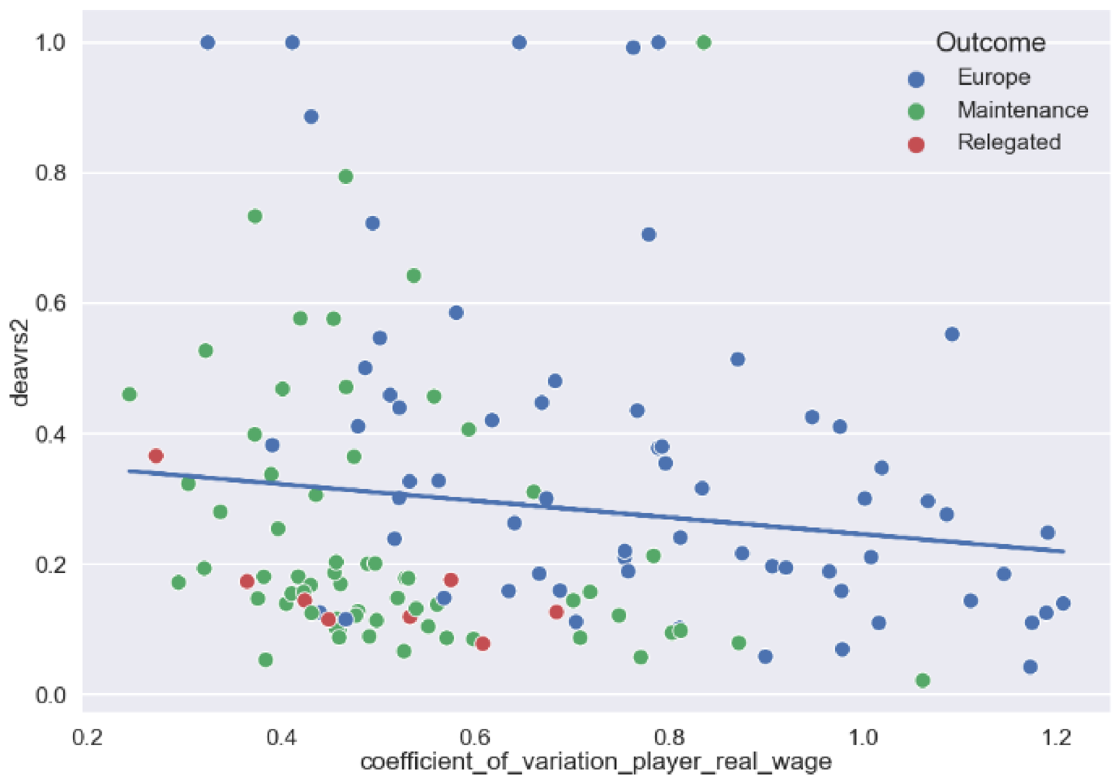


Figure 23: Linear regression between the coefficient of variation of players' real wages and deavrs2

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 $r^2 = 0.017$.

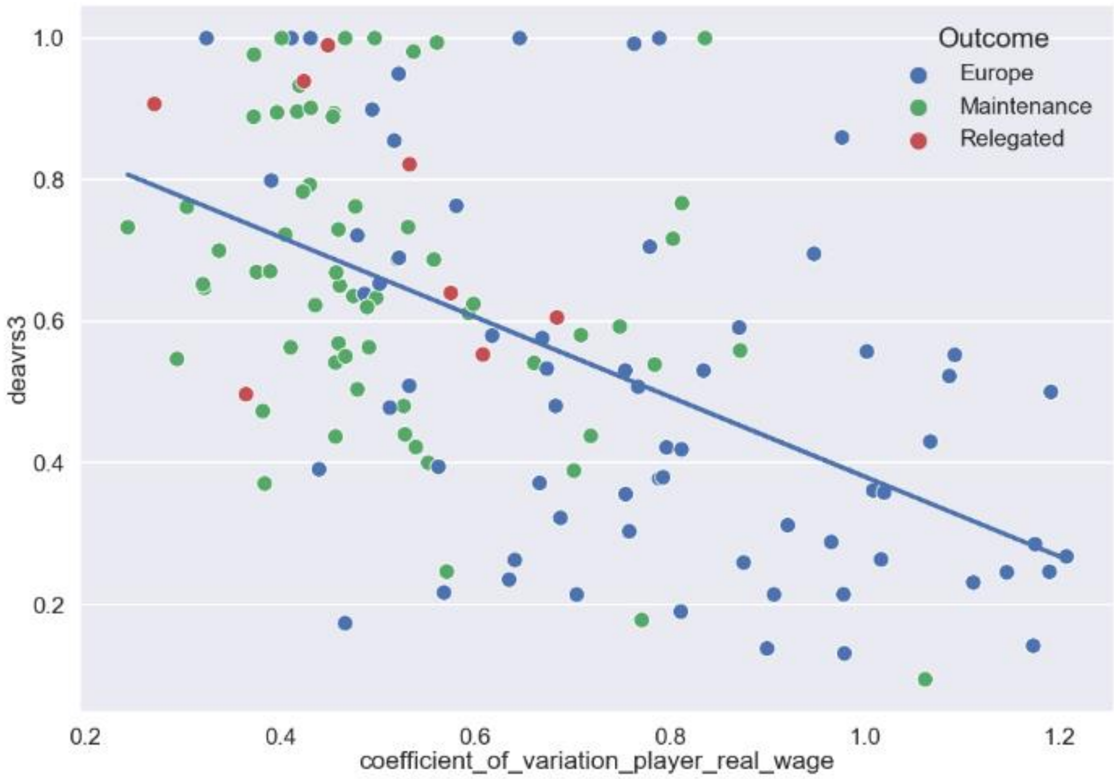


Figure 24: Linear regression between coefficient of variation of players' real wages and deavrs3

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.

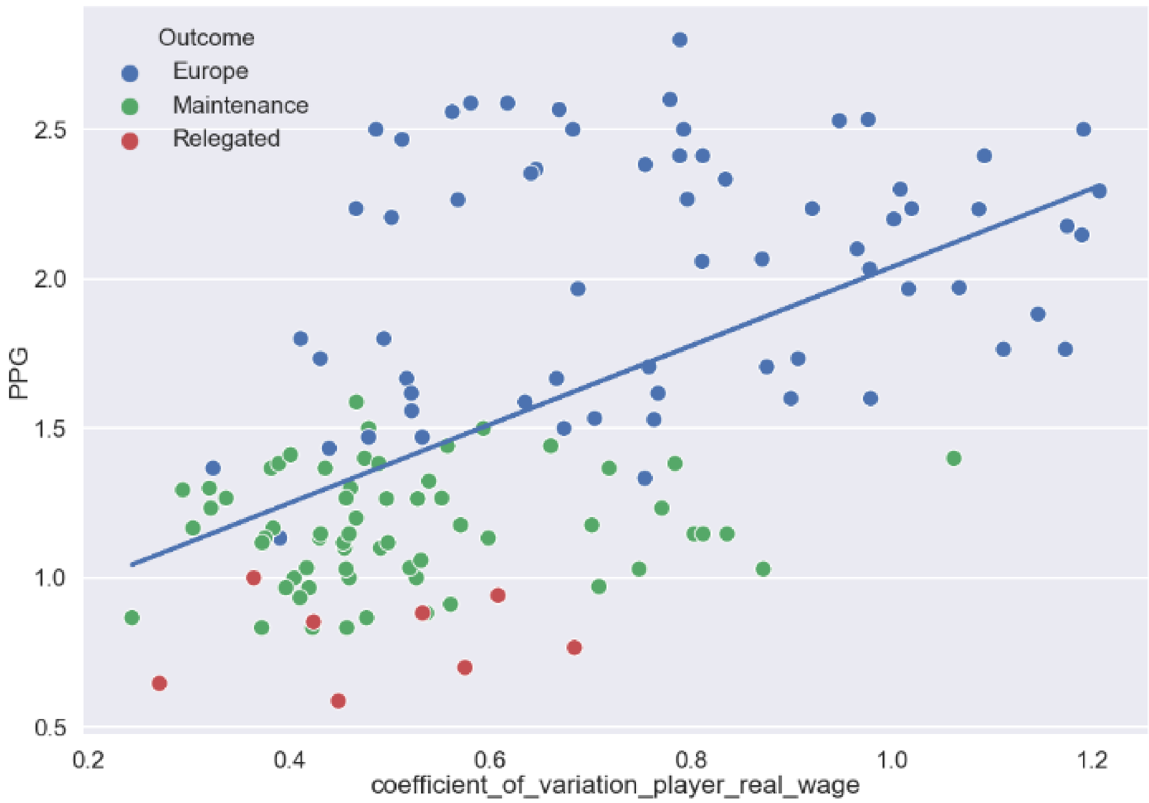


Figure 25: Linear regression between coefficient of variation of players' real wages and PPG

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. However, in Primeira Liga there is not a salary cap and teams seek to increase players' salaries in order to rank higher in the league, which sometimes leads to higher coefficient of variation. It is, however, an interesting topic for further research in the future.

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

Table 6: Correlation measures of the ratio total coach to total player real wages and efficiency measures

	Correlation with Total Coach/Total Player wages		
	Pearson's p	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 the table above, 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.

5.5 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 following 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.

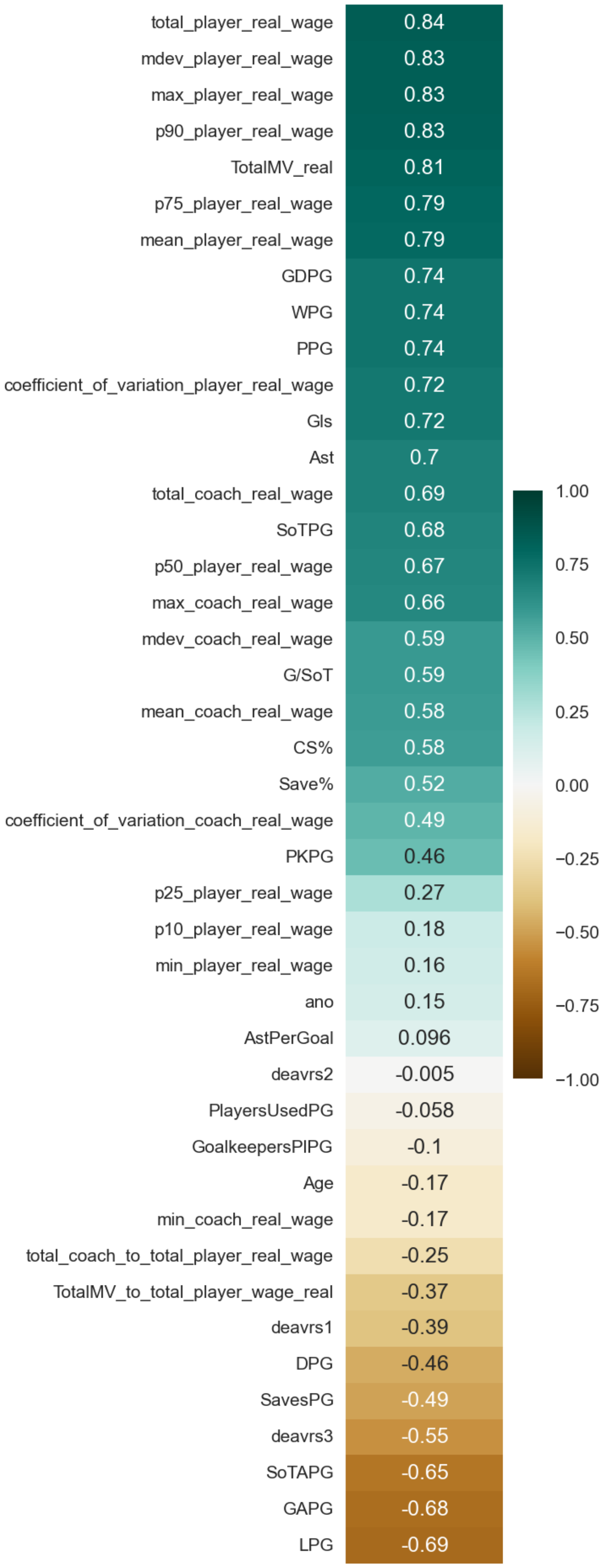


Figure 26: Correlation heatmap of Attendance and other variables

The rationale on the previous chart 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, a chart with the Spearman partial correlations of Attendance while controlling for the total player wage expenditure and total market value of the team was plotted. On a qualitative level, anyone can immediately see that variables lost a lot of their explanatory value.

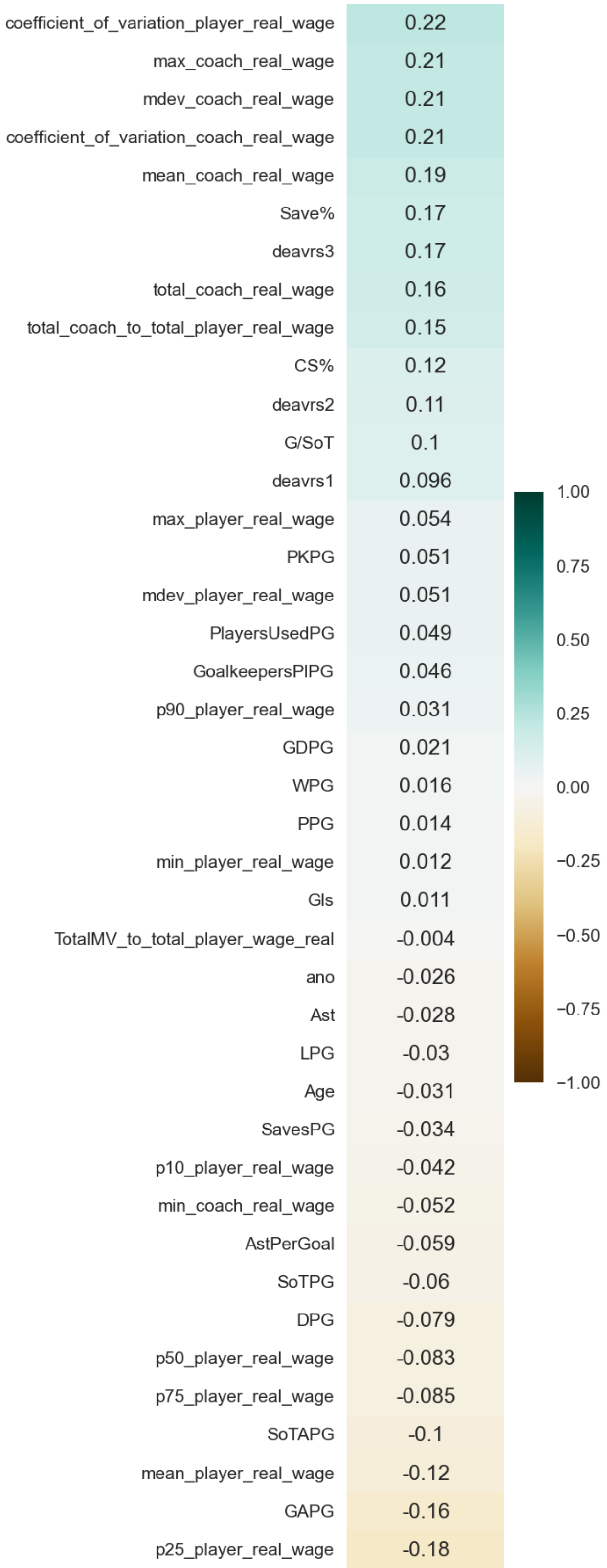


Figure 27: Partial correlation heatmap of Attendance and other variables (controlling for total player real wages and total real market values)

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.

5.6 Maximizing Points Per Game and Efficiency

This section, especially aimed at sports decision-makers, shows the relationship between sports results and efficiency.

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

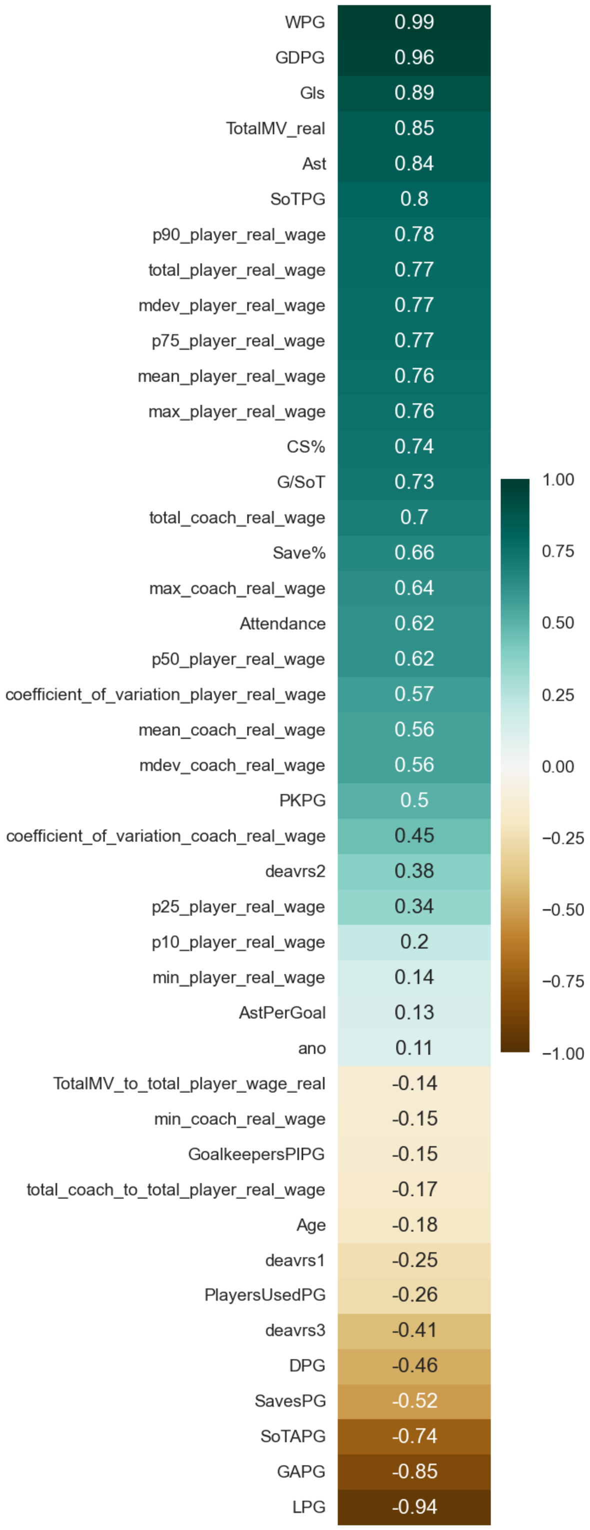


Figure 28: Correlation heatmap of PPG and other variables

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.

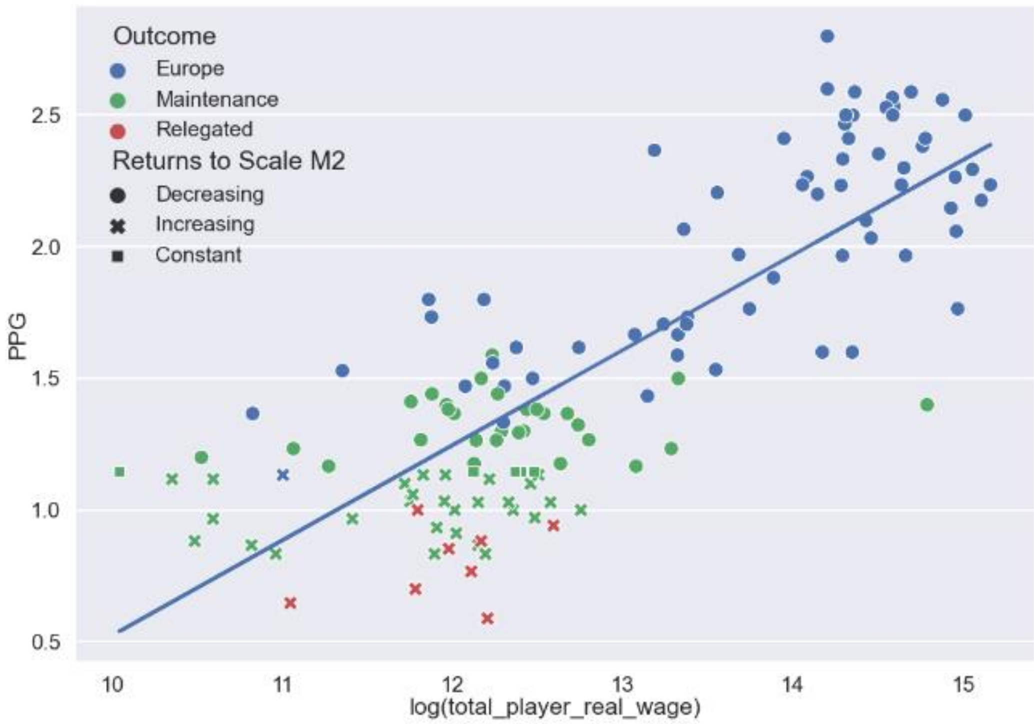


Figure 29: Linear regression between the logarithmized total player real wages and PPG

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.

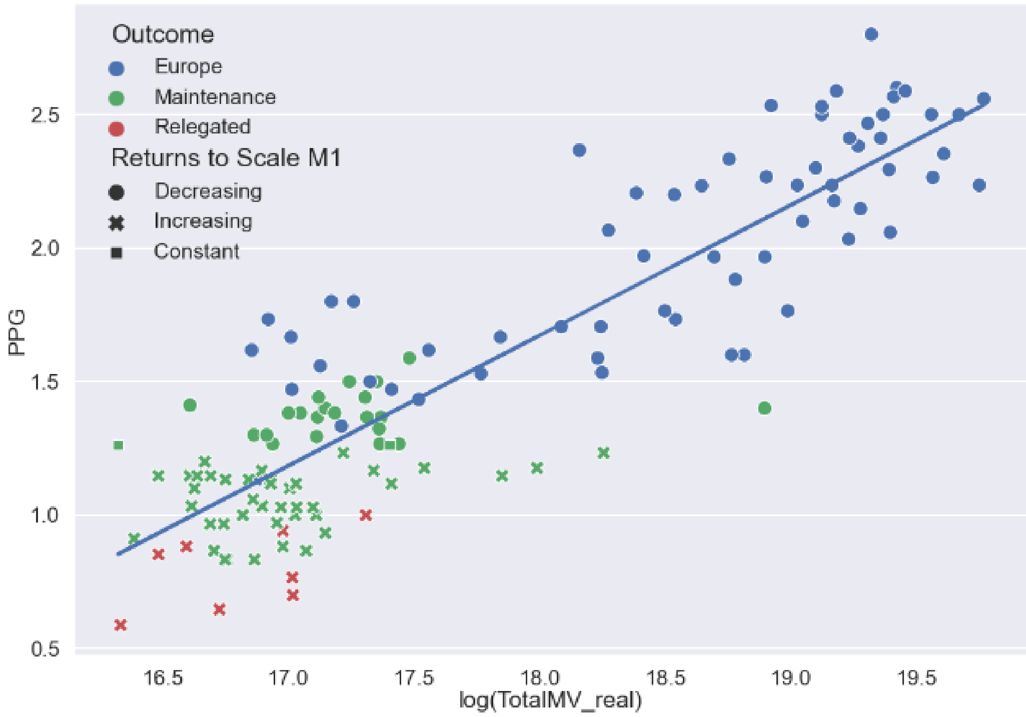


Figure 30: Linear regression between the logarithmized total real market values and PPG

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.

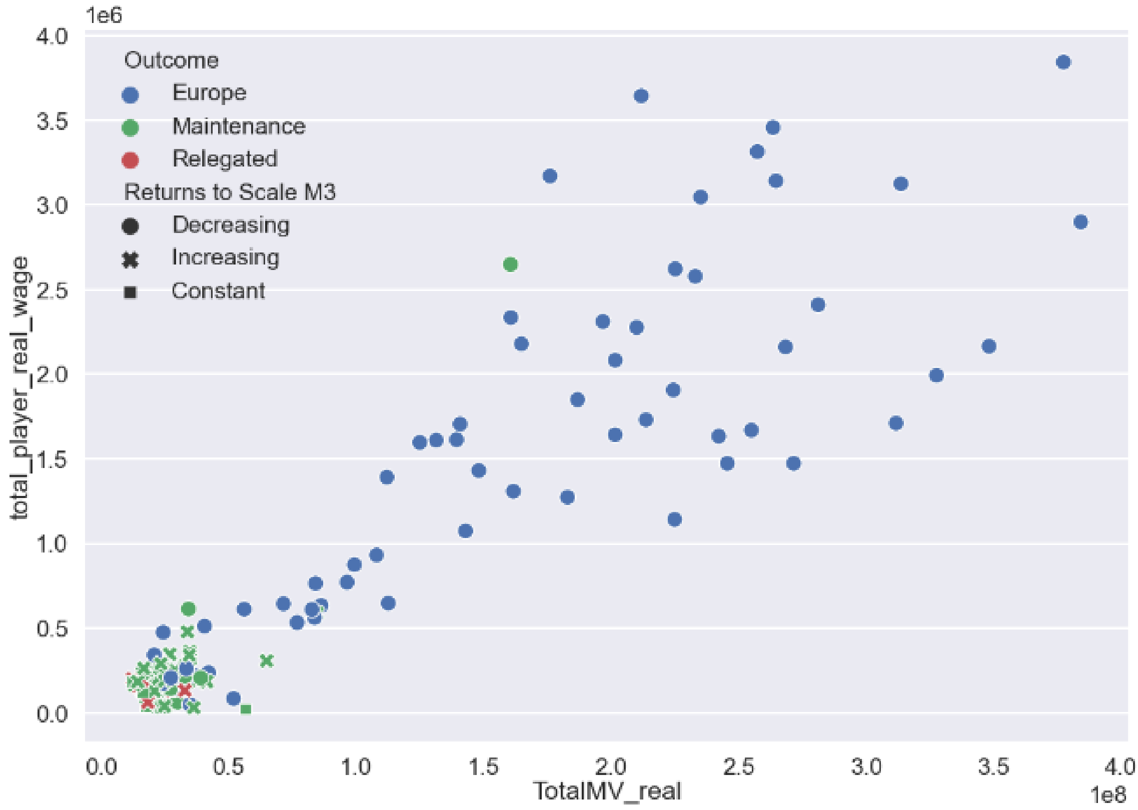


Figure 31: Graphical representation of the distribution of the total real market values and the total player real wages

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.

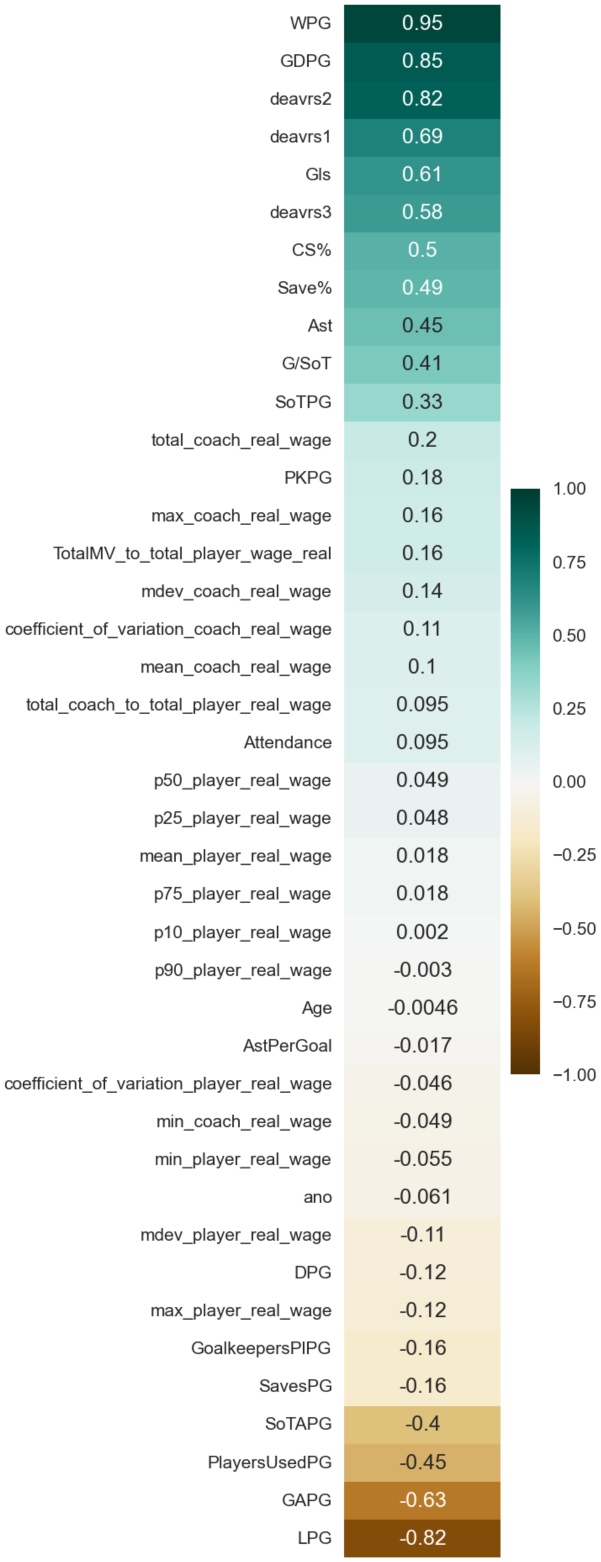


Figure 32: Partial correlation heatmap of PPG and other variables (controlling for total player real wages and total real market values)

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, Rios, 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.

Chapter 6 – Conclusions and Future Work

This chapter synthetizes the conclusions drawn during the results discussion and, based on these, other sources found in the literature review, and limitations of the present study, presents further work to be the researched in the field.

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.

In Primeira Liga, 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

Afriat, S. N. (1972). Efficiency Estimation of Production Functions. *International Economic Review*, *13*(3), 568. https://doi.org/10.2307/2525845

Aigner, D., Lovell, C. A. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models. *Journal of Econometrics*, *6*(1), 21–37. https://doi.org/10.1016/0304-4076(77)90052-5

Andor, M., & Hesse, F. (2013). The StoNED age: the departure into a new era of efficiency analysis? A monte carlo comparison of StoNED and the "oldies" (SFA and DEA). *Journal of Productivity Analysis 2013 41:1*, *41*(1), 85–109. https://doi.org/10.1007/S11123-013-0354-Y

Antonioni, P., & Cubbin, J. (2000). The Bosman Ruling and the Emergence of a Single Market in Soccer Talent. *European Journal of Law and Economics 2000 9:2*, *9*(2), 157–173. https://doi.org/10.1023/A:1018778718514

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.

Barros, C. P., del Corral, J., & Garcia-del-Barrio, P. (2007). Identification of Segments of Soccer Clubs in the Spanish League First Division With a Latent Class Model. *Http://Dx.Doi.Org/10.1177/1527002508316980*, *9*(5), 451–469. https://doi.org/10.1177/1527002508316980

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

Bezat, A., Rolnictwa, E., Stosunków, M., & Sggw, G. (2009). Comparison of the deterministic and stochastic approaches for estimating technical efficiency on the example of non-parametric DEA and parametric SFA methods. *Qme.Sggw.Edu.Pl*. https://qme.sggw.edu.pl/article/view/3019

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

Brown, A., & Reade, J. J. (2019). The wisdom of amateur crowds: Evidence from an online community of sports tipsters. *European Journal of Operational Research*, *272*(3), 1073–1081. https://doi.org/10.1016/J.EJOR.2018.07.015

Brown, A., & Yang, F. (2019). The wisdom of large and small crowds: Evidence from repeated natural experiments in sports betting. *International Journal of Forecasting*, *35*(1), 288–296. https://doi.org/10.1016/J.IJFORECAST.2018.06.002

Business, W., Conyon, M. J., & Sadler, G. v. (2001). *Executive pay, tournaments and corporate performance in UK firms 1*.

Button, K. J., & Weyman-Jones, T. G. (1994). X-Efficiency and Technical Efficiency. *Public Choice*, *80*(1), 83–104.

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

Carmichael, F., & Thomas, D. (2006). Production and efficiency in team sports: an investigation of rugby league football. *Http://Dx.Doi.Org/10.1080/00036849500000039*, *27*(9), 859–869. https://doi.org/10.1080/00036849500000039

Carmichael, F., Thomas, D., & Ward, R. (2000). Team performance: the case of English premiership football. *Wiley Online Library*. https://onlinelibrary.wiley.com/doi/abs/10.1002/1099-1468(200001/02)21:1%3C31::AID-MDE963%3E3.0.CO;2-Q?casa_token=AwmKGPF2jRcAAAAA:myIhqEZx1YG4XgSCp7lXcw6K8kHVnlKyD5aVJXhqpjOyk-iuB6HbkLMQRllPXNAJ5eOk6VWan1BRLZhy

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

Casu, B., Molyneux, P., & Molyneuxy, P. (2010). *A comparative study of efficiency in European banking*. https://doi.org/10.1080/0003684032000158109

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

Chen, J., Ezzamel, M., & Cai, Z. (2011). Managerial power theory, tournament theory, and executive pay in China. *Journal of Corporate Finance*, *17*(4), 1176–1199. https://doi.org/10.1016/J.JCORPFIN.2011.04.008

Clemente, J. A. A., Requena, B., Jukic, I., Nayler, J., Hernández, A. S., & Carling, C. (2019). Is Physical Performance a Differentiating Element between More or Less Successful Football Teams? *Sports 2019, Vol. 7, Page 216*, *7*(10), 216. https://doi.org/10.3390/SPORTS7100216

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*.

Cooper, W. W., Seiford, L. M., & Zhu, J. (2004). *Data Envelopment Analysis BT - Handbook on Data Envelopment Analysis* (W. W. Cooper, L. M. Seiford, & J. Zhu, Eds.; pp. 1–39). Springer US. https://doi.org/10.1007/1-4020-7798-X_1

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

Cullinane, K., & Song, D. W. (2006). Estimating the Relative Efficiency of European Container Ports: A Stochastic Frontier Analysis. *Research in Transportation Economics*, *16*, 85–115. https://doi.org/10.1016/S0739-8859(06)16005-9

Dawson, P., Dobson, S., & Gerrard, B. (2016). Stochastic Frontiers and the Temporal Structure of Managerial Efficiency in English Soccer. *Http://Dx.Doi.Org/10.1177/152700250000100402*, *1*(4), 341–362. https://doi.org/10.1177/152700250000100402

di Salvo, V., Gregson, W., Atkinson, G., Tordoff, P., & Drust, B. (2009). Analysis of high intensity activity in premier league soccer. *International Journal of Sports Medicine*, *30*(3), 205–212. https://doi.org/10.1055/S-0028-1105950/ID/29

Dimitropoulos, P. (2010). The Financial Performance of the Greek Football Clubs. *Choregia*, *6*(1).

Dobson, S., & Gerrard, B. (1999). The Determination of Player Transfer Fees in English Professional Soccer. *Journal of Sport Management*, *13*(4), 259–279. https://doi.org/10.1123/JSM.13.4.259

Dobson, S., Gerrard, B., & Howe, S. (2010). The determination of transfer fees in English nonleague football. *Http://Dx.Doi.Org/10.1080/000368400404281*, *32*(9), 1145–1152. https://doi.org/10.1080/000368400404281

Ecer, F., Economics, A. B.-A. J. of, & 2014, undefined. (2010). Measuring performances of football clubs using financial ratios: The gray relational analysis approach. *Researchgate.Net*, *5*, 52–66. https://www.researchgate.net/profile/Adem-Boeyuekaslan/publication/259799919_Measuring_Performances_Of_Football_Clubs_Using_Financial_Ratios_The_Gray_Relational_Analysis_Approach/links/0deec52deae390724f000000/Measuring-Performances-Of-Football-Clubs-Using-Financial-Ratios-The-Gray-Relational-Analysis-Approach.pdf

Eriksson, T. (2015). Executive Compensation and Tournament Theory: Empirical Tests on Danish Data. *Https://Doi.Org/10.1086/209920*, *17*(2), 262–280. https://doi.org/10.1086/209920

Espitia-Escuer, M., & García-Cebrián, L. I. (2014). Comparison of efficiency measures for Spanish first division football teams using data envelopment and stochastic frontier analyses. *Atlantic Review of Economics*, *2*, 1–28. https://www.econstor.eu/handle/10419/146588

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

FIFA. (2013a, September 22). *History of Football - Britain, the home of Football*. https://web.archive.org/web/20130922082153/http://www.fifa.com/classicfootball/history/the-game/Britain-home-of-football.html

FIFA. (2013b, September 22). *History of Football - The Origins*. https://web.archive.org/web/20130424153755/http://www.fifa.com/classicfootball/history/the-game/origins.html

FIFA. (2022). *About FIFA*. https://www.fifa.com/about-fifa

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

Football Benchmark. (2022). *Football Benchmark - Home of football business data*. https://www.footballbenchmark.com/home

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

Forrest, D., & Simmons, R. (2002). *Team Salaries and Playing Success in Sports: A Comparative Perspective* (pp. 221–238). https://doi.org/10.1007/978-3-663-07711-4_12

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

Frantz, R. (2019). The beginnings of behavioral economics: Katona, simon, and leibenstein's x-efficiency theory. *The Beginnings of Behavioral Economics: Katona, Simon, and Leibenstein's X-Efficiency Theory*, 1–240. https://doi.org/10.1016/C2017-0-02119-5

Freedman, D., Pisani, R., & Purves, R. (2007). Statistics (international student edition). *Pisani, R. Purves, 4th Edn. WW Norton & Company, New York*.

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

Garcia-del-Barrio, P., & Pujol, F. (2020). Recruiting talent in a global sports market: appraisals of soccer players' transfer fees. *Managerial Finance*, *47*(6), 789–811. https://doi.org/10.1108/MF-04-2020-0213/FULL/XML

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

Gomes Rocha, R. M. (2016). *A correlação entre a performance económico-financeira e a performance desportiva dos clubes de futebol: O caso dos três grandes em Portugal*.

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

Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., del Río, J. F., Wiebe, M., Peterson, P., … Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*(7825), 357–362. https://doi.org/10.1038/S41586-020-2649-2

Hatrori, T. (2002). Relative Performance of U.S. and Japanese Electricity Distribution: An Application of Stochastic Frontier Analysis. *Journal of Productivity Analysis 2002 18:3*, *18*(3), 269–284. https://doi.org/10.1023/A:1020695709797

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

Herzog, S. M., & Hertwig, R. (2011). The wisdom of ignorant crowds: Predicting sport outcomes by mere recognition. *Judgment and Decision Making*, *6*(1), 58–72.

Hjalmarsson, L., Kumbhakar, S. C., & Heshmati, A. (1996). DEA, DFA and SFA: A comparison. *Journal of Productivity Analysis 1996 7:2*, *7*(2), 303–327. https://doi.org/10.1007/BF00157046

Horrigan, J. O. (1968). A short history of financial ratio analysis. *The Accounting Review*, *43*(2), 284–294.

IFAB. (2022). *The IFAB Organisation*. https://www.theifab.com/organisation/

Katayama, H., & Nuch, H. (2009). *A game-level analysis of salary dispersion and team performance in the national basketball association*. 1194–1206. https://doi.org/10.1080/00036840802600335

Kendall, M. G. (1938). A New Measure of Rank Correlation. *Biometrika*, *30*(1/2), 81. https://doi.org/10.2307/2332226

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

Kim, Y. S., & Kim, M. (2019). “A wisdom of crowds”: Social media mining for soccer match analysis. *IEEE Access*, *7*, 52634–52639. https://doi.org/10.1109/ACCESS.2019.2912009

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

Kuosmanen, T., & Kortelainen, M. (2012). Stochastic non-smooth envelopment of data: Semi-parametric frontier estimation subject to shape constraints. *Journal of Productivity Analysis*, *38*(1), 11–28. https://doi.org/10.1007/S11123-010-0201-3/FIGURES/1

Lazear, E. P., & Rosen, S. (1981). Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy*, *89*(5), 841–864. https://doi.org/10.1086/261010

Leibenstein, H. (1966). Allocative Efficiency vs. “X-Efficiency.” *The American Economic Review*, *56*(3), 392–415. http://www.jstor.org/stable/1823775

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

Lewis, H. F., & Sexton, T. R. (2004). Network DEA: eeciency analysis of organizations with complex internal structure. *Computers & Operations Research*, *31*, 1365–1410. https://doi.org/10.1016/S0305-0548(03)00095-9

Liga Portugal. (2022a). *Apresentação da Liga*. https://www.ligaportugal.pt/pt/paginas/conteudos/apresentacao-da-liga/

Liga Portugal. (2022b). *Liga Portugal - 40 Anos*. https://www.ligaportugal.pt/pt/40anos/

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

Liu, Y., & Myers, R. (2009). Model selection in stochastic frontier analysis with an application to maize production in Kenya. *Journal of Productivity Analysis*, *31*(1), 33–46. https://doi.org/10.1007/S11123-008-0111-9/FIGURES/2

Lovell, C. A. K. (1993). Production frontiers and productive efficiency. *The Measurement of Productive Efficiency: Techniques and Applications*, *3*, 67.

Lucifora, C., & Simmons, R. (2016). Superstar Effects in Sport: Evidence From Italian Soccer. *Http://Dx.Doi.Org/10.1177/1527002502239657*, *4*(1), 35–55. https://doi.org/10.1177/1527002502239657

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

Mailena, L., Shamsudin, M., & Radam, A. (2014). Efficiency of rice farms and its determinants: Application of stochastic frontier analysis. *Trends in Applied Sciences Research*. https://search.proquest.com/openview/a5676f801cc67705c787287837ca7d23/1.pdf?pq-origsite=gscholar&cbl=276280

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

McKinney, W. (2011). pandas: a foundational Python library for data analysis and statistics. In *dlr.de*. https://www.dlr.de/sc/portaldata/15/resources/dokumente/pyhpc2011/submissions/pyhpc2011_submission_9.pdf

Meeusen, W., & van den Broeck, J. (1977). Efficiency Estimation from Cobb-Douglas Production Functions with Composed Error. *International Economic Review*, *18*(2), 435. https://doi.org/10.2307/2525757

Melton, M., & Zorn, T. S. (2000a). An empirical test of tournament theory: The senior PGA tour. *Managerial Finance*, *26*(7), 16–31. https://doi.org/10.1108/03074350010766765/FULL/XML

Melton, M., & Zorn, T. S. (2000b). Risk taking in tournaments. *Managerial Finance*, *26*(7), 52–61. https://doi.org/10.1108/03074350010766792/FULL/XML

Memmert, D., & Raabe, D. (2018). Data Analytics in Football : Positional Data Collection, Modelling and Analysis. *Data Analytics in Football*. https://doi.org/10.4324/9781351210164

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

Neal, D., & Rosen, S. (2000). Chapter 7 Theories of the distribution of earnings. *Handbook of Income Distribution*, *1*, 379–427. https://doi.org/10.1016/S1574-0056(00)80010-X

Nelsen, R. B. (2011). *Kendall tau metric*. Encyclopaedia of Mathematics. https://encyclopediaofmath.org/index.php?title=Kendall_tau_metric

Odeck, J. (2007). Comparison of data envelopment analysis and deterministic parametric frontier approaches: an application in the Norwegian road construction sector. *Http://Dx.Doi.Org/10.1080/03081060108717663*, *24*(2), 111–134. https://doi.org/10.1080/03081060108717663

Pelloneová, N., & Tomíček, M. (2022). Ranking Players by DEA: An analysis of Czech and Danish football. *Studia Sportiva*, *16*(1), 76–90. https://doi.org/10.5817/STS2022-1-8

Plumley, D., Wilson, R., & Ramchandani, G. (2014). Towards a model for measuring holistic performance of professional Football clubs. *Http://Dx.Doi.Org/10.1080/14660970.2014.980737*, *18*(1), 16–29. https://doi.org/10.1080/14660970.2014.980737

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

Pyatunin, A. v., Vishnyakova, A. B., Sherstneva, N. L., Mironova, S. P., Dneprov, S. A., & Grabozdin, Y. P. (2016). The Economic Efficiency of European Football Clubs--Data Envelopment Analysis (DEA) Approach. *International Journal of Environmental and Science Education*, *11*(15), 7515–7534. http://creativecommons.org/licenses/by/4.0/

Rajesh, P., Bharadwaj, Alam, M., & Tahernezhadi, M. (2020). A Data Science Approach to Football Team Player Selection. *IEEE International Conference on Electro Information Technology*, *2020-July*, 175–183. https://doi.org/10.1109/EIT48999.2020.9208331

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

Rampinini, E., Impellizzeri, F. M., Castagna, C., Coutts, A. J., & Wisløff, U. (2009). Technical performance during soccer matches of the Italian Serie A league: Effect of fatigue and competitive level. *Journal of Science and Medicine in Sport*, *12*(1), 227–233. https://doi.org/10.1016/J.JSAMS.2007.10.002

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

Richardson, L. (2019). Beautiful soup documentation. In *media.readthedocs.org*. https://media.readthedocs.org/pdf/beautiful-soup-4/latest/beautiful-soup-4.pdf

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

Rosko, M. D., & Mutter, R. L. (2011). What have we learned from the application of stochastic frontier analysis to U.S. hospitals? *Medical Care Research and Review*, *68*(1), 75–100. https://doi.org/10.1177/1077558710370686

Rubinfeld, D. L. (2011). Reference Guide on Multiple Regression. In *Reference Manual on Scientific Evidence*. National Academies Press. https://doi.org/10.17226/13163

Ruijg, J., & van Ophem, H. (2014). Determinants of football transfers. *Http://Dx.Doi.Org/10.1080/13504851.2014.892192*, *22*(1), 12–19. https://doi.org/10.1080/13504851.2014.892192

Rummel, R. J. (1976). *Understanding Correlation*. http://www.hawaii.edu/powerkills/UC.HTM

Sarkis, J. (2007). Preparing your data for DEA. In *Modeling Data Irregularities and Structural Complexities in Data Envelopment Analysis* (pp. 305–320). Springer US. https://doi.org/10.1007/978-0-387-71607-7_17/COVER

Serna Rodríguez, M., Ramírez Hassan, A., & Coad, A. (2018). Uncovering Value Drivers of High Performance Soccer Players. *Journal of Sports Economics*. https://doi.org/10.1177/1527002518808344

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

Surowiecki, J. (2004). The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations. In *The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business, economies, societies, and nations.* Doubleday & Co.

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

Szymanski, S. (2003). Incentives and competitive balance in team sports. *European Sport Management Quarterly*, *3*(1), 11–30. https://doi.org/10.1080/16184740308721937

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

The International Football Association Board. (2022). *Laws of the Game 2022/23*.

Torgler, B., & Schmidt, S. L. (2011). What shapes player performance in soccer? Empirical findings from a panel analysis. *Https://Doi.Org/10.1080/00036840600660739*, *39*(18), 2355–2369. https://doi.org/10.1080/00036840600660739

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

Vallat, R. (2022). *Functions — pingouin 0.5.2 documentation*. https://pingouin-stats.org/api.html

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

Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., … Vázquez-Baeza, Y. (2020). SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*, *17*(3), 261–272. https://doi.org/10.1038/S41592-019-0686-2

Weimar, D., & Wicker, P. (2014). Moneyball Revisited: Effort and Team Performance in Professional Soccer. *Http://Dx.Doi.Org/10.1177/1527002514561789*, *18*(2), 140–161. https://doi.org/10.1177/1527002514561789

Williams, L. V., & Reade, J. J. (2020). Sports Forecasting. *Wiley StatsRef: Statistics Reference Online*, 1–7. https://doi.org/10.1002/9781118445112.STAT08137

Wilson, D., Ying, Y.-H., Wilson, D., & Ying, Y.-H. (2003). Nationality preferences for labour in the international football industry. *Applied Economics*, *35*(14), 1551–1559. https://doi.org/10.1080/0003684032000010048

Yang, Z. (2006). A two-stage DEA model to evaluate the overall performance of Canadian life and health insurance companies. *Mathematical and Computer Modelling*, *43*(7–8), 910–919. https://doi.org/10.1016/J.MCM.2005.12.011

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., Iráizoz, B., & Lera-López, F. (2018). Are football managers as efficient as coaches? Performance analysis with ex ante and ex post inputs in the Premier league. *Https://Doi.Org/10.1080/00036846.2018.1495821*, *51*(3), 303–314. https://doi.org/10.1080/00036846.2018.1495821

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

Zar, J. H. (2005). Spearman rank correlation. *Encyclopedia of Biostatistics*, *7*.

Appendix

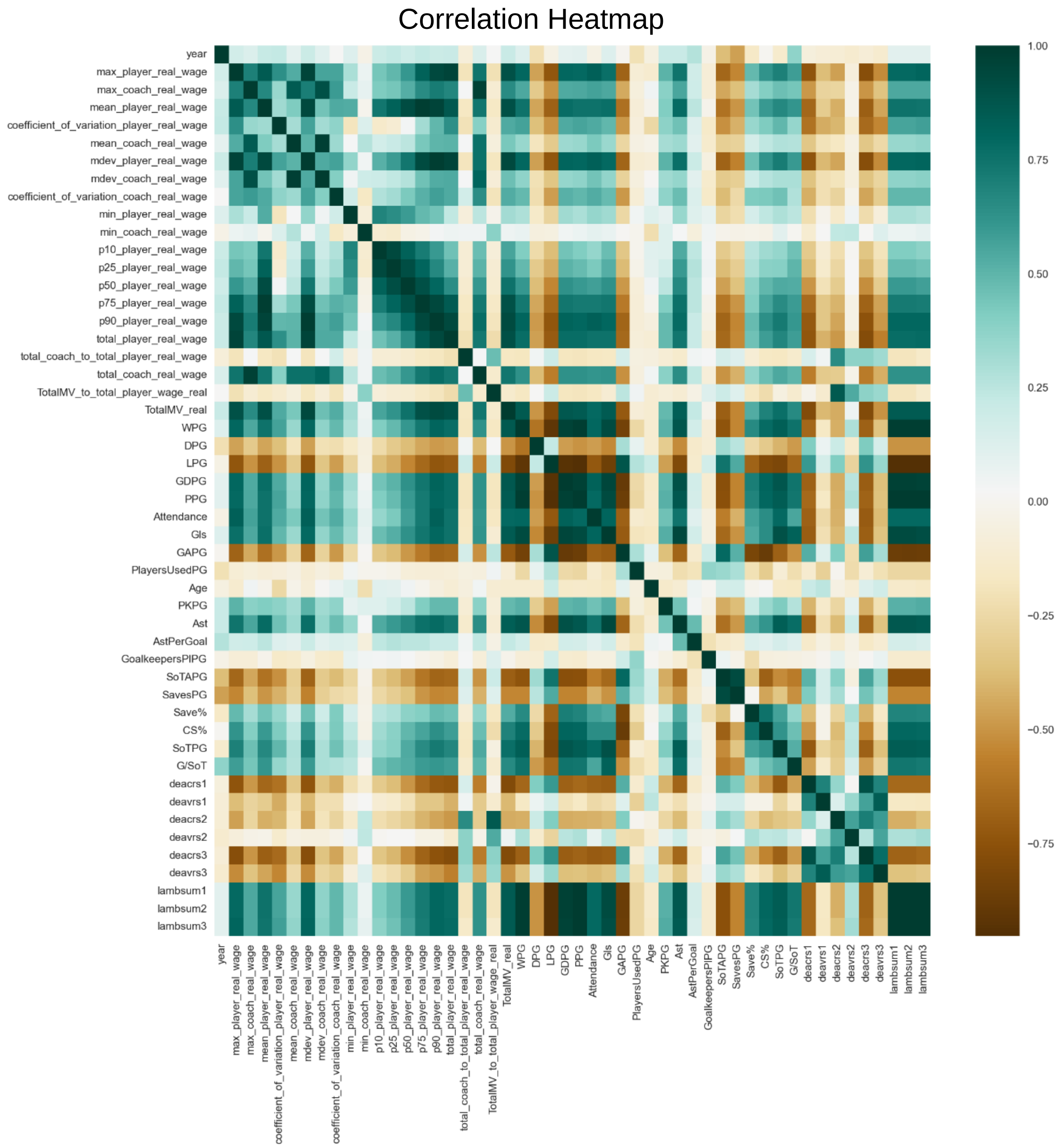
year	squad	deacrs1	deavrs1	deacrs2	deavrs2	deacrs3	deavrs3	RTS M1	RTS M2	RTS M3	RTS Ratio M1	RTS Ratio M2	RTS Ratio M3
2008	Sporting CP	0,1903	0,5574	0,0317	0,3005	0,2027	0,5574	Decreasing	Decreasing	Decreasing	1,7395	1,9179	1,6777
2008	Porto	0,1622	0,5306	0,0290	0,3163	0,1763	0,5306	Decreasing	Decreasing	Decreasing	1,845	2,0342	1,7582
2008	Braga	0,2880	0,3718	0,0546	0,1853	0,3179	0,3718	Decreasing	Decreasing	Decreasing	1,3178	1,453	1,2436
2008	Vitória Setúbal	0,3243	0,4733	0,3477	0,4602	0,5304	0,7330	Increasing	Increasing	Increasing	0,6853	0,7556	0,7235
2008	Benfica	0,1450	0,3225	0,0245	0,1595	0,1552	0,3225	Decreasing	Decreasing	Decreasing	1,555	1,7145	1,4964
2008	Leixões	0,4726	0,5621	0,1560	0,4066	0,5959	0,6119	Decreasing	Decreasing	Decreasing	1,186	1,3077	1,1061
2008	Marítimo	0,3981	0,4082	0,3868	0,5276	0,5979	0,6473	Increasing	Decreasing	Decreasing	0,9752	1,0752	1,028
2008	Naval	0,5021	0,6569	0,4861	0,5768	0,7518	0,9331	Increasing	Increasing	Increasing	0,7643	0,8427	0,8057
2008	Académica	0,5990	0,6186	0,1053	0,1697	0,6480	0,6506	Decreasing	Decreasing	Increasing	1,0279	1,1333	0,9823
2008	Vitória	0,3536	0,3543	0,0700	0,1044	0,3948	0,4004	Decreasing	Decreasing	Increasing	1,0016	1,1043	0,9384
2008	Rio Ave	0,4815	0,6089	0,1215	0,1394	0,5686	0,7225	Increasing	Increasing	Increasing	0,7907	0,8718	0,7132
2008	Paços	0,5180	0,5781	0,3779	0,3825	0,7484	0,7990	Increasing	Increasing	Increasing	0,8961	0,988	0,9237
2008	Nacional	0,7541	1,0000	0,2415	0,8863	0,9441	1,0000	Decreasing	Decreasing	Decreasing	1,3705	1,5111	1,2716
2009	Sporting CP	0,1047	0,1312	0,0224	0,0695	0,1192	0,1312	Decreasing	Decreasing	Decreasing	1,2651	1,3949	1,1705
2009	Porto	0,1359	0,4220	0,0347	0,3548	0,1609	0,4220	Decreasing	Decreasing	Decreasing	1,7922	1,9761	1,6137
2009	Braga	0,2986	1,0000	0,0889	1,0000	0,3673	1,0000	Decreasing	Decreasing	Decreasing	1,8713	2,0632	1,7142
2009	Vitória Setúbal	0,4266	0,6475	0,2897	0,3988	0,6109	0,8891	Increasing	Increasing	Increasing	0,6589	0,7265	0,6747
2009	Belenenses	0,3030	0,4998	0,0846	0,1265	0,3665	0,6055	Increasing	Increasing	Increasing	0,6062	0,6684	0,5486
2009	Benfica	0,1490	0,8598	0,0233	0,4107	0,1559	0,8598	Decreasing	Decreasing	Decreasing	2,0031	2,2085	1,9531
2009	Leixões	0,2760	0,4987	0,1071	0,1755	0,3606	0,6401	Increasing	Increasing	Increasing	0,5535	0,6103	0,5296
2009	Marítimo	0,3800	0,4142	0,5447	1,0000	0,7933	1,0000	Decreasing	Decreasing	Decreasing	1,0806	1,1915	1,1476
2009	Naval	0,6721	0,7083	0,6460	0,7943	1,0000	1,0000	Increasing	Decreasing	Constant	0,9488	1,0462	1
2009	Académica	0,4397	0,5055	0,0856	0,0892	0,4887	0,5635	Increasing	Increasing	Increasing	0,8698	0,959	0,8172
2009	Vitória	0,3797	0,4139	0,0857	0,1573	0,4375	0,4383	Decreasing	Decreasing	Increasing	1,0806	1,1915	0,9915
2009	Rio Ave	0,6098	0,7464	0,1630	0,1810	0,7292	0,8964	Increasing	Increasing	Increasing	0,8171	0,9009	0,7335
2009	Olhanense	0,5296	0,6929	0,2143	0,2543	0,6977	0,8950	Increasing	Increasing	Increasing	0,7643	0,8427	0,7359
2009	Paços	0,5207	0,5645	0,2974	0,3234	0,7280	0,7618	Increasing	Decreasing	Increasing	0,9225	1,0171	0,9278
2010	Sporting CP	0,1102	0,1380	0,0188	0,0583	0,1182	0,1380	Decreasing	Decreasing	Decreasing	1,2651	1,3949	1,2151
2010	Porto	0,1106	1,0000	0,0381	1,0000	0,1408	1,0000	Decreasing	Decreasing	Decreasing	2,214	2,441	2,0794
2010	Braga	0,1768	0,2142	0,0402	0,1117	0,2041	0,2142	Decreasing	Decreasing	Decreasing	1,2124	1,3368	1,1112

2010	Vitória Setúbal	0,5329	0,5947	0,0841	0,0852	0,5591	0,6247	Increasing	Increasing	Increasing	0,8961	0,988	0,8723
2010	Benfica	0,1090	0,2884	0,0228	0,1887	0,1233	0,2884	Decreasing	Decreasing	Decreasing	1,6605	1,8308	1,5422
2010	Marítimo	0,3337	0,3617	0,0488	0,0531	0,3420	0,3710	Increasing	Decreasing	Increasing	0,9225	1,0171	0,9099
2010	Académica	0,3912	0,4947	0,0862	0,0989	0,4482	0,5691	Increasing	Increasing	Increasing	0,7907	0,8718	0,7282
2010	Beira-Mar	0,6418	0,7379	0,1792	0,1869	0,7765	0,8941	Increasing	Increasing	Increasing	0,8698	0,959	0,7874
2010	Vitória	0,3425	0,3911	0,0561	0,1256	0,3632	0,3911	Decreasing	Decreasing	Decreasing	1,1333	1,2496	1,0962
2010	Rio Ave	0,5416	0,5426	0,1878	0,2801	0,6907	0,7001	Decreasing	Decreasing	Increasing	1,0016	1,1043	0,9416
2010	Olhanense	0,4894	0,5461	0,1454	0,1472	0,6016	0,6698	Increasing	Increasing	Increasing	0,8961	0,988	0,8205
2010	Paços	0,4886	0,5325	0,1666	0,3059	0,6207	0,6230	Decreasing	Decreasing	Decreasing	1,0806	1,1915	1,0132
2011	Sporting CP	0,1187	0,2638	0,0169	0,1100	0,1205	0,2638	Decreasing	Decreasing	Decreasing	1,555	1,7145	1,5417
2011	Porto	0,0777	0,4130	0,0293	0,4807	0,1009	0,4807	Decreasing	Decreasing	Decreasing	1,9767	2,1795	1,8829
2011	Braga	0,2323	0,5912	0,0653	0,5142	0,2816	0,5912	Decreasing	Decreasing	Decreasing	1,6341	1,8017	1,4813
2011	Vitória Setúbal	0,3595	0,4547	0,0579	0,0664	0,3794	0,4805	Increasing	Increasing	Increasing	0,7907	0,8718	0,7671
2011	Benfica	0,1133	0,3614	0,0199	0,2108	0,1227	0,3614	Decreasing	Decreasing	Decreasing	1,8186	2,0051	1,7375
2011	Marítimo	0,6625	0,8553	0,0703	0,2389	0,6625	0,8553	Decreasing	Decreasing	Decreasing	1,3178	1,453	1,3178
2011	Vitória	0,4239	0,5041	0,0490	0,1277	0,4239	0,5041	Decreasing	Decreasing	Decreasing	1,186	1,3077	1,186
2011	Rio Ave	0,3237	0,4387	0,1260	0,1548	0,4231	0,5633	Increasing	Increasing	Increasing	0,738	0,8137	0,7064
2011	Gil Vicente	0,5845	0,6523	0,1657	0,1677	0,7100	0,7927	Increasing	Increasing	Increasing	0,8961	0,988	0,8136
2011	Olhanense	0,5692	0,5879	0,1200	0,1935	0,6455	0,6523	Decreasing	Decreasing	Increasing	1,0279	1,1333	0,9534
2011	Paços	0,4604	0,5635	0,1334	0,1480	0,5624	0,6878	Increasing	Increasing	Increasing	0,8171	0,9009	0,7447
2012	Sporting CP	0,0845	0,0944	0,0106	0,0216	0,0845	0,0944	Decreasing	Decreasing	Decreasing	1,107	1,2205	1,107
2012	Porto	0,0928	0,6174	0,0354	0,7055	0,1208	0,7055	Decreasing	Decreasing	Decreasing	2,0558	2,2667	1,9617
2012	Braga	0,1491	0,1978	0,0536	0,1968	0,1917	0,2143	Decreasing	Decreasing	Decreasing	1,3705	1,5111	1,296
2012	Vitória Setúbal	0,4676	0,6824	0,0917	0,1214	0,5209	0,7624	Increasing	Increasing	Increasing	0,6853	0,7556	0,643
2012	Benfica	0,0927	0,5763	0,0238	0,4477	0,1099	0,5763	Decreasing	Decreasing	Decreasing	2,0295	2,2376	1,8258
2012	Marítimo	0,3277	0,3283	0,1363	0,2033	0,4341	0,4373	Decreasing	Decreasing	Increasing	1,0016	1,1043	0,9681
2012	Vitória	0,4337	0,4606	0,1218	0,2103	0,5254	0,5307	Decreasing	Decreasing	Increasing	1,0543	1,1624	0,9553
2012	Rio Ave	0,4855	0,5421	0,1786	0,3647	0,6271	0,6358	Decreasing	Decreasing	Decreasing	1,107	1,2205	1,0505
2012	Gil Vicente	0,4306	0,6535	0,1140	0,1569	0,5135	0,7833	Increasing	Increasing	Increasing	0,6589	0,7265	0,5903
2012	Olhanense	0,3831	0,5815	0,0845	0,1163	0,4391	0,6690	Increasing	Increasing	Increasing	0,6589	0,7265	0,6068
2012	Paços	0,6092	0,9785	0,2548	1,0000	0,8078	1,0000	Decreasing	Decreasing	Decreasing	1,4233	1,5692	1,3768

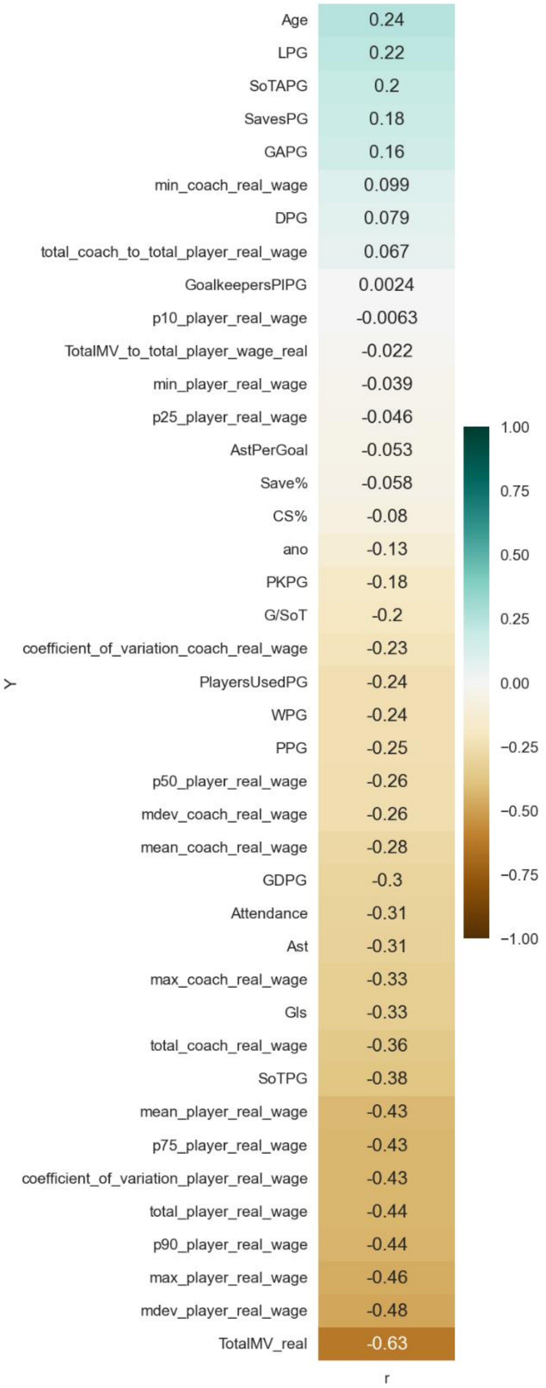
2013	Sporting CP	0,1732	0,5228	0,0280	0,2762	0,1830	0,5228	Decreasing	Decreasing	Decreasing	1,7659	1,947	1,7116
2013	Porto	0,0879	0,2145	0,0214	0,1589	0,1030	0,2145	Decreasing	Decreasing	Decreasing	1,6078	1,7726	1,4581
2013	Braga	0,1413	0,1449	0,0420	0,0572	0,1737	0,1781	Increasing	Decreasing	Increasing	0,9752	1,0752	0,8928
2013	Benfica	0,0988	0,4781	0,0303	0,4592	0,1223	0,4781	Decreasing	Decreasing	Decreasing	1,9504	2,1504	1,7957
2013	Estoril	0,5574	0,8952	0,1842	0,7227	0,7030	0,8992	Decreasing	Decreasing	Decreasing	1,4233	1,5692	1,3276
2013	Marítimo	0,4014	0,4376	0,0984	0,1807	0,4711	0,4733	Decreasing	Decreasing	Increasing	1,0806	1,1915	0,979
2014	Sporting CP	0,1185	0,3583	0,0352	0,3476	0,1456	0,3583	Decreasing	Decreasing	Decreasing	1,7674	1,9487	1,6182
2014	Porto	0,1040	0,4198	0,0423	0,5527	0,1371	0,5527	Decreasing	Decreasing	Decreasing	1,907	2,1026	1,8375
2014	Braga	0,2313	0,3035	0,0531	0,1890	0,2675	0,3035	Decreasing	Decreasing	Decreasing	1,3488	1,4872	1,2344
2014	Benfica	0,1203	0,6388	0,0305	0,5007	0,1422	0,6388	Decreasing	Decreasing	Decreasing	1,9767	2,1795	1,7818
2014	Marítimo	0,4645	0,4772	0,1080	0,1718	0,5388	0,5469	Decreasing	Decreasing	Increasing	1,0233	1,1282	0,9345
2014	Vitória	0,3718	0,4696	0,1369	0,4357	0,4804	0,5076	Decreasing	Decreasing	Decreasing	1,2791	1,4103	1,2141
2014	Penafiel	0,3419	0,6682	0,2065	0,3661	0,4818	0,9071	Increasing	Increasing	Increasing	0,5116	0,5641	0,5177
2014	Paços	0,5293	0,5836	0,1745	0,3376	0,6672	0,6707	Decreasing	Decreasing	Decreasing	1,093	1,2051	1,0191
2014	Rio Ave	0,3393	0,3393	0,1202	0,1784	0,4348	0,4405	Constant	Decreasing	Increasing	1	1,1026	0,9435
2015	Sporting CP	0,1216	0,6955	0,0244	0,4257	0,1362	0,6955	Decreasing	Decreasing	Decreasing	2	2,2051	1,8705
2015	Porto	0,0885	0,2464	0,0141	0,1252	0,0932	0,2464	Decreasing	Decreasing	Decreasing	1,6977	1,8718	1,6495
2015	Braga	0,1976	0,2592	0,0608	0,2165	0,2450	0,2594	Decreasing	Decreasing	Decreasing	1,3488	1,4872	1,2428
2015	Vitória Setúbal	0,3621	0,5190	0,4943	0,6425	0,7261	0,9810	Increasing	Increasing	Increasing	0,6977	0,7692	0,7402
2015	Benfica	0,1174	0,7633	0,0300	0,5857	0,1390	0,7633	Decreasing	Decreasing	Decreasing	2,0465	2,2564	1,843
2015	União	0,5744	0,8518	0,1073	0,1443	0,6318	0,9393	Increasing	Increasing	Increasing	0,6744	0,7436	0,638
2015	Marítimo	0,3748	0,4604	0,0911	0,1016	0,4391	0,5419	Increasing	Increasing	Increasing	0,814	0,8974	0,7383
2015	Vitória	0,2751	0,2957	0,1276	0,1442	0,3718	0,3893	Increasing	Decreasing	Increasing	0,9302	1,0256	0,9126
2015	Paços	0,5133	0,5891	0,2002	0,4572	0,6712	0,6873	Decreasing	Decreasing	Decreasing	1,1395	1,2564	1,0912
2015	Rio Ave	0,3917	0,4578	0,1334	0,3268	0,4975	0,5091	Decreasing	Decreasing	Decreasing	1,1628	1,2821	1,0901
2016	Sporting CP	0,0754	0,1901	0,0131	0,1020	0,0814	0,1901	Decreasing	Decreasing	Decreasing	1,6279	1,7949	1,5581
2016	Porto	0,1032	0,3121	0,0197	0,1944	0,1141	0,3121	Decreasing	Decreasing	Decreasing	1,7674	1,9487	1,6661
2016	Braga	0,1863	0,2321	0,0521	0,1588	0,2255	0,2354	Decreasing	Decreasing	Decreasing	1,2558	1,3846	1,1373
2016	Vitória Setúbal	0,4354	0,4927	0,5615	0,5762	0,8328	0,8893	Increasing	Increasing	Increasing	0,8837	0,9744	0,9365
2016	Benfica	0,0917	0,3703	0,0290	0,3780	0,1144	0,3780	Decreasing	Decreasing	Decreasing	1,907	2,1026	1,765
2016	Marítimo	0,5825	0,6808	0,1680	0,4113	0,7106	0,7213	Decreasing	Decreasing	Decreasing	1,1628	1,2821	1,0589

2016	Chaves	0,4814	0,5447	0,1108	0,1137	0,5571	0,6331	Increasing	Increasing	Increasing	0,8837	0,9744	0,8084
2016	Paços	0,4894	0,5846	0,1645	0,1782	0,6198	0,7332	Increasing	Increasing	Increasing	0,8372	0,9231	0,7831
2016	Rio Ave	0,4264	0,4894	0,1360	0,3107	0,5334	0,5413	Decreasing	Decreasing	Decreasing	1,1395	1,2564	1,0566
2016	Feirense	0,8387	0,9441	0,2218	0,4687	1,0000	1,0000	Decreasing	Decreasing	Constant	1,1163	1,2308	1
2017	Sporting CP	0,0844	0,2679	0,0133	0,1398	0,0885	0,2679	Decreasing	Decreasing	Decreasing	1,814	2	1,7663
2017	Porto	0,0892	0,5800	0,0215	0,4206	0,1044	0,5800	Decreasing	Decreasing	Decreasing	2,0465	2,2564	1,8587
2017	Braga	0,2218	0,6534	0,0573	0,5470	0,2632	0,6534	Decreasing	Decreasing	Decreasing	1,7442	1,9231	1,5679
2017	Benfica	0,0991	0,3561	0,0185	0,2200	0,1090	0,3561	Decreasing	Decreasing	Decreasing	1,8837	2,0769	1,7819
2017	Marítimo	0,4616	0,5090	0,1100	0,2128	0,5384	0,5390	Decreasing	Decreasing	Increasing	1,093	1,2051	0,9945
2017	Portimonense	0,2978	0,3370	0,7148	0,7336	0,9288	0,9768	Increasing	Increasing	Increasing	0,8837	0,9744	0,9508
2017	Chaves	0,5549	0,6118	0,1037	0,2005	0,6103	0,6203	Decreasing	Decreasing	Decreasing	1,093	1,2051	1,0339
2017	Paços	0,5315	0,7618	0,0918	0,1193	0,5722	0,8218	Increasing	Increasing	Increasing	0,6977	0,7692	0,6687
2017	Rio Ave	0,4353	0,5177	0,1153	0,3006	0,5192	0,5333	Decreasing	Decreasing	Decreasing	1,186	1,3077	1,0628
2017	Feirense	0,6764	0,9382	0,1098	0,1381	0,7154	0,9938	Increasing	Increasing	Increasing	0,7209	0,7949	0,6984
2018	Sporting CP	0,0996	0,2855	0,0120	0,1103	0,0996	0,2855	Decreasing	Decreasing	Decreasing	1,7209	1,8974	1,7209
2018	Porto	0,0696	0,3698	0,0231	0,3800	0,0879	0,3800	Decreasing	Decreasing	Decreasing	1,9767	2,1795	1,8458
2018	Braga	0,1924	0,4303	0,0451	0,2965	0,2236	0,4303	Decreasing	Decreasing	Decreasing	1,5581	1,7179	1,4211
2018	Benfica	0,0646	0,3948	0,0177	0,3278	0,0777	0,3948	Decreasing	Decreasing	Decreasing	2,0233	2,2308	1,8249
2018	Marítimo	0,6273	0,6917	0,0948	0,0948	0,6494	0,7167	Increasing	Constant	Increasing	0,907	1	0,8897
2018	Portimonense	0,1967	0,2169	1,0000	1,0000	1,0000	1,0000	Increasing	Constant	Constant	0,907	1	1
2018	Vitória	0,2854	0,3450	0,3596	0,9921	0,5354	0,9921	Decreasing	Decreasing	Decreasing	1,2093	1,3333	1,2808
2018	Chaves	0,3863	0,5191	0,0640	0,0780	0,4110	0,5533	Increasing	Increasing	Increasing	0,7442	0,8205	0,7184
2018	Rio Ave	0,3706	0,3904	0,0778	0,1318	0,4198	0,4225	Decreasing	Decreasing	Increasing	1,0465	1,1538	0,9711
2018	Feirense	0,4604	0,9899	0,0590	0,1150	0,4604	0,9899	Increasing	Increasing	Increasing	0,4651	0,5128	0,4651
2019	Sporting CP	0,0972	0,1419	0,0112	0,0423	0,0972	0,1419	Decreasing	Decreasing	Decreasing	1,3953	1,5385	1,3953
2019	Porto	0,1039	0,4194	0,0184	0,2408	0,1127	0,4194	Decreasing	Decreasing	Decreasing	1,907	2,1026	1,8193
2019	Braga	0,1584	0,2313	0,0379	0,1439	0,1849	0,2313	Decreasing	Decreasing	Decreasing	1,3953	1,5385	1,2686
2019	Benfica	0,0700	0,2170	0,0145	0,1481	0,0791	0,2170	Decreasing	Decreasing	Decreasing	1,7907	1,9744	1,665
2019	Marítimo	0,6823	0,7523	0,0980	0,0980	0,6952	0,7669	Increasing	Constant	Increasing	0,907	1	0,8977
2019	Portimonense	0,4079	0,5315	0,0737	0,0871	0,4446	0,5807	Increasing	Increasing	Increasing	0,7674	0,8462	0,73
2019	Gil Vicente FC	1,0000	1,0000	0,1356	0,2012	1,0000	1,0000	Constant	Decreasing	Constant	1	1,1026	1

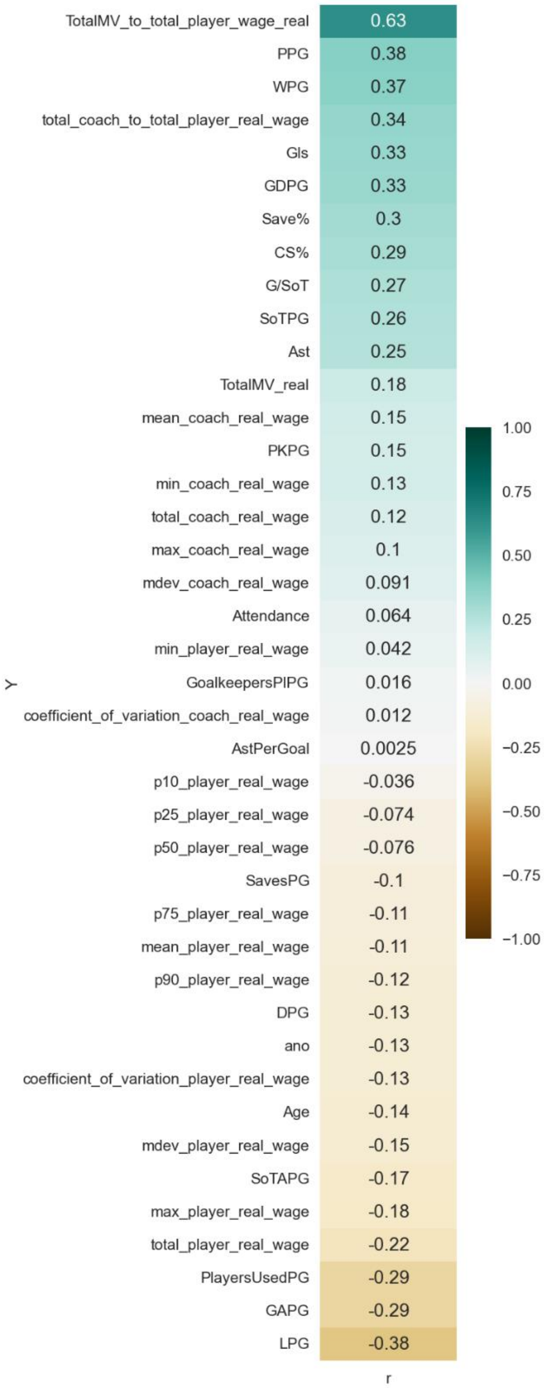
2019	Paços	0,7727	0,8519	0,1250	0,1250	0,8164	0,9015	Increasing	Constant	Increasing	0,907	1	0,8791
2019	Rio Ave	0,7518	0,9496	0,0948	0,3016	0,7518	0,9496	Decreasing	Decreasing	Decreasing	1,2791	1,4103	1,2791
2019	Famalicão	0,3942	0,4910	0,1548	0,4714	0,5161	0,5507	Decreasing	Decreasing	Decreasing	1,2558	1,3846	1,2037
2020	Sporting CP	0,0942	0,5004	0,0151	0,2482	0,0993	0,5004	Decreasing	Decreasing	Decreasing	1,9767	2,1795	1,9187
2020	Porto	0,0696	0,2310	0,0237	0,2630	0,0884	0,2630	Decreasing	Decreasing	Decreasing	1,8605	2,0513	1,7435
2020	Braga	0,1277	0,2457	0,0351	0,1848	0,1539	0,2457	Decreasing	Decreasing	Decreasing	1,4884	1,641	1,3436
2020	Benfica	0,0574	0,1737	0,0117	0,1152	0,0645	0,1737	Decreasing	Decreasing	Decreasing	1,7674	1,9487	1,6491
2020	Marítimo	0,4256	0,5229	0,0712	0,0794	0,4542	0,5590	Increasing	Increasing	Increasing	0,814	0,8974	0,7844
2020	Portimonense	0,4000	0,4914	0,1090	0,1215	0,4808	0,5926	Increasing	Increasing	Increasing	0,814	0,8974	0,7334
2020	Gil Vicente FC	0,6620	0,7299	0,0876	0,0876	0,6620	0,7299	Increasing	Constant	Increasing	0,907	1	0,907
2020	Paços	0,5516	0,6771	0,1514	0,4399	0,6643	0,6899	Decreasing	Decreasing	Decreasing	1,2326	1,359	1,1121
2020	Rio Ave	0,2949	0,3730	0,1511	0,1733	0,4054	0,4971	Increasing	Increasing	Increasing	0,7907	0,8718	0,7854
2020	Famalicão	0,1757	0,1889	0,0768	0,0868	0,2349	0,2466	Increasing	Decreasing	Increasing	0,9302	1,0256	0,9053

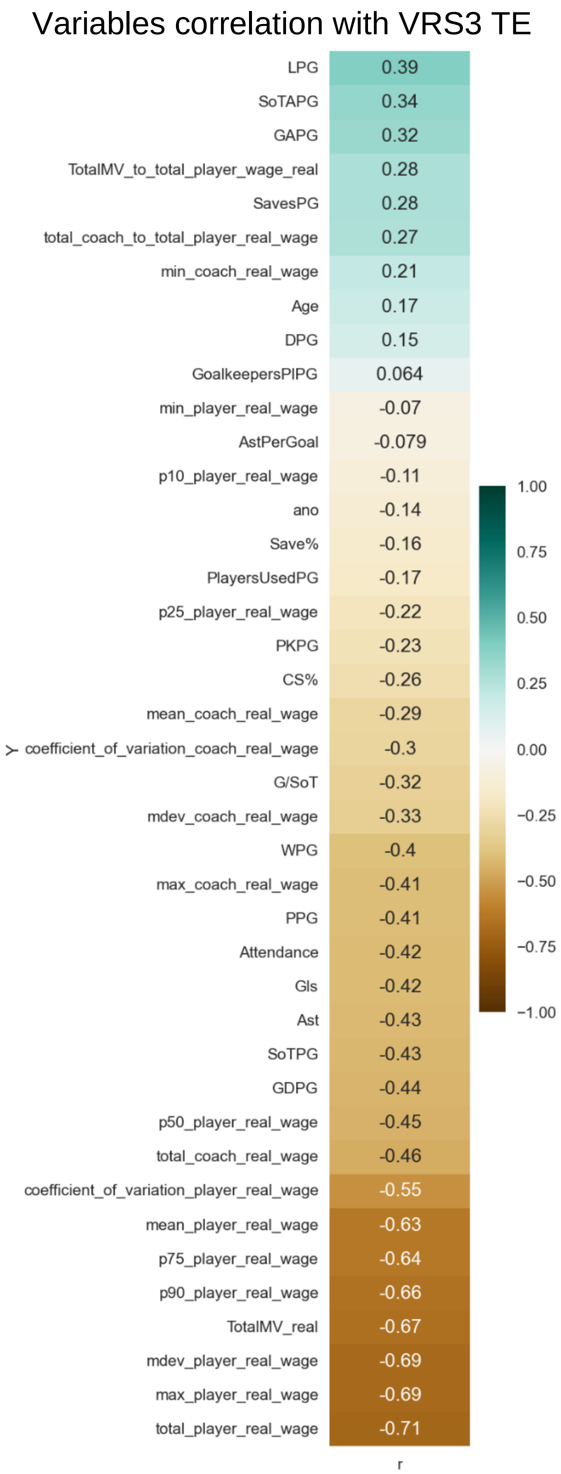

Correlation Heatmap
year
max_player_real_wage
max_coach_real_wage
mean_player_real_wage
coefficient_of_variation_player_real_wage
mean_coach_real_wage
mdev_player_real_wage
mdev_coach_real_wage
coefficient_of_variation_coach_real_wage
min_player_real_wage
min_coach_real_wage
p10_player_real_wage
p25_player_real_wage
p50_player_real_wage
p75_player_real_wage
p90_player_real_wage
total_player_real_wage
total_coach_to_total_player_real_wage
total_coach_real_wage
TotalMV_to_total_player_wage_real
TotalMV_real
WPG
DPG
LPG
GDPG
PPG
Attendance
Gls
GAPG
PlayersUsedPG
Age
PKPG
Ast
AstPerGoal
GoalkeepersPlPG
SoTAPG
SavesPG
Save%
CS%
SoTPG
G/SoT
deacrs1
deavrs1
deacrs2
deavrs2
deacrs3
deavrs3
lambsum1
lambsum2
lambsum3
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75

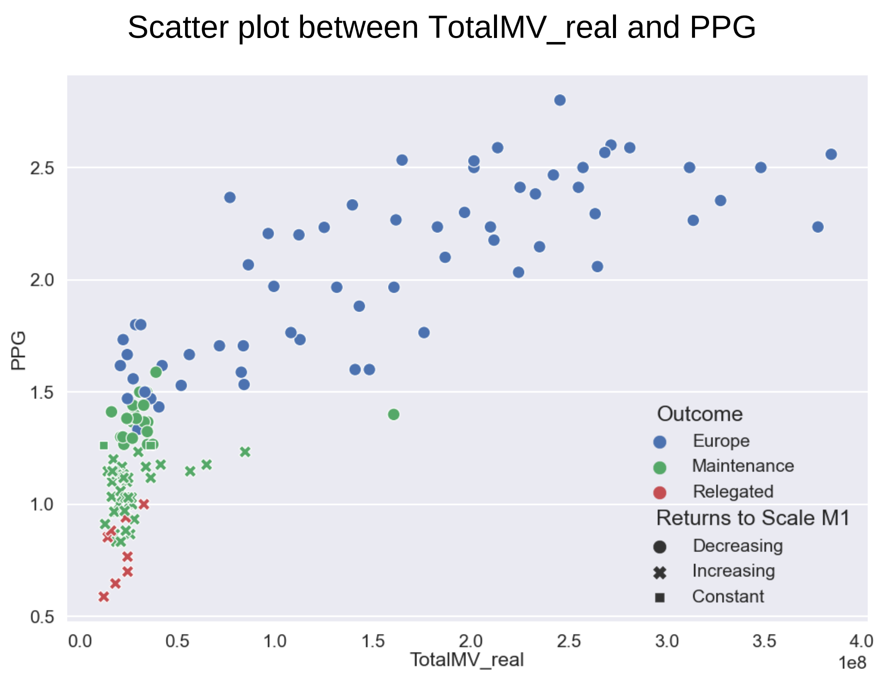
Variables correlation with VRS1 TE

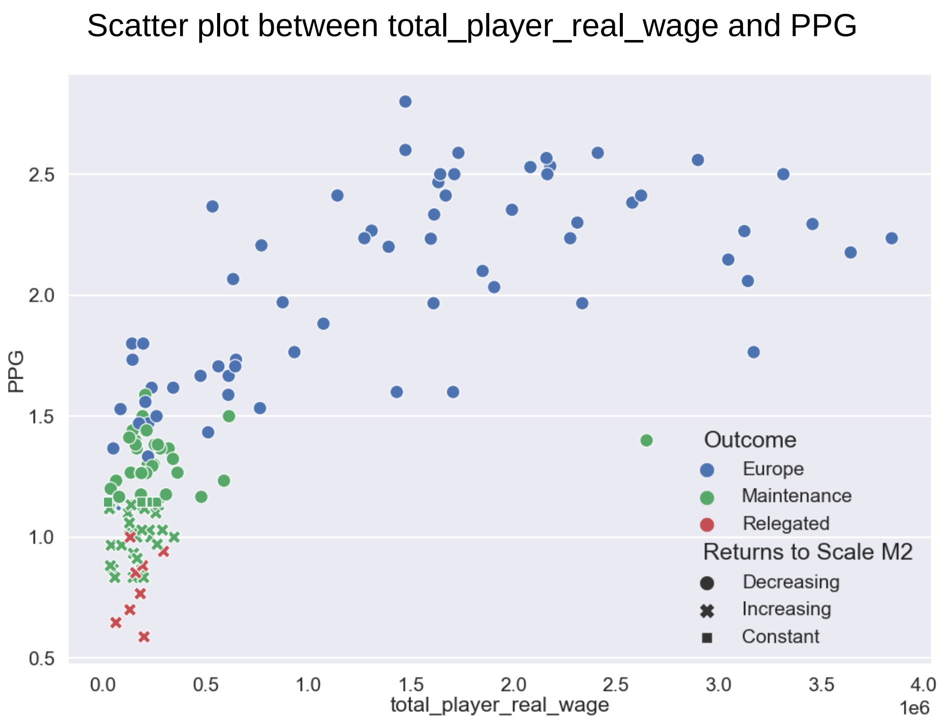


Variables correlation with VRS2 TE




Variables correlation with VRS3 TE
LPG 0.39
SoTAPG 0.34
GAPG 0.32
TotalMV_to_total_player_wage_real 0.28
SavesPG 0.28
total_coach_to_total_player_real_wage 0.27
min_coach_real_wage 0.21
Age 0.17
DPG 0.15
GoalkeepersPIPG 0.064
min_player_real_wage -0.07
AstPerGoal -0.079
p10_player_real_wage -0.11
ano -0.14
Save% -0.16
PlayersUsedPG -0.17
p25_player_real_wage -0.22
PKPG -0.23
CS% -0.26
mean_coach_real_wage -0.29
coefficient_of_variation_coach_real_wage -0.3
G/SoT -0.32
mdev_coach_real_wage -0.33
WPG -0.4
max_coach_real_wage -0.41
PPG -0.41
Attendance -0.42
Gls -0.42
Ast -0.43
SoTPG -0.43
GDPG -0.44
p50_player_real_wage -0.45
total_coach_real_wage -0.46
coefficient_of_variation_player_real_wage -0.55
mean_player_real_wage -0.63
p75_player_real_wage -0.64
p90_player_real_wage -0.66
TotalMV_real -0.67
mdev_player_real_wage -0.69
max_player_real_wage -0.69
total_player_real_wage -0.71
r
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
−0.75
−1.00


Scatter plot between TotalMV_real and PPG
PPG
2.5
2.0
1.5
1.0
0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
TotalMV_real
1e8
Outcome
Europe
Maintenance
Relegated
Returns to Scale M1
Decreasing
Increasing
Constant


Scatter plot between total_player_real_wage and PPG
PPG
2.5
2.0
1.5
1.0
0.5
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
total_player_real_wage
1e6
Outcome
Europe
Maintenance
Relegated
Returns to Scale M2
Decreasing
Increasing
Constant