MASTER
APPLIED ECONOMETRICS AND FORECASTING

MASTER´S FINAL WORK
DISSERTATION

DOCUMENT SPECIFICALLY ELABORATED TO OBTAIN THE MASTER'S DEGREE

UNCOVERING PROFESSIONAL FOOTBALL TRANSFER FEE DRIVERS, USING HEDONIC REGRESSION MODELS:

EVIDENCE FROM THE MAJOR EUROPEAN LEAGUES

DUARTE DOS SANTOS SOARES

SUPERVISION:
ESMERALDA ARRANHADO

2023

This page was intentionally left blank

Glossary

FA – Football association

FIFA - Fédération Internationale de Football Association

OLS - Ordinary least squares

PQML – Poisson quasi-maximum-likelihood

PQMLE – Poisson quasi-maximum-likelihood estimator

QML – Quasi-maximum-likelihood

RESET - Regression specification error test

SPL - Saudi Pro League

SW – Summer window

UCL – UEFA champions league

UEFA - Union of European football associations

VIF – Variance inflation factor

WW – Winter window

ABSTRACT, KEYWORDS AND JEL CODES

ABSTRACT:

Football transfer market news are each year a major topic of interest, as football clubs invest substantial sums of money to acquire new players for their teams. This study utilizes a distinctive cross-sectional database comprising information from 503 player transfers, regarding the 2022/23 football season, considering the six major European leagues. The primary goal of this study is to uncover the determinants of the transfer fee agreed by two clubs using hedonic price models, taking into consideration the set of characteristics included in the database.

Many previous studies on this topic have typically used a straightforward approach to the problem by employing log-linear models, which, although not necessarily incorrect, can be a restrictive approach. This dissertation takes an alternative approach by utilizing a non-linear, the Poisson, estimated through quasi-maximum-likelihood. The aim is to uncover the transfer fees drivers, while also comparing the obtained results to the classical approach based on linear models.

Furthermore, the most suitable regression model (Poisson) among the available options will be utilized for prediction exercise, using data from the 2023-24 season – using the six leagues and additionally a Saudi Arabian one. This will allow for a thorough evaluation of player valuation discrepancies, between the market value and the model's predictions, thereby illustrating the potential of the proposed model and suggesting whether the transferred player is potentially undervalued or overvalued. This approach can be a great tool for not only researchers, but also provides valuable information for team-management and investors.

KEYWORDS: Football; Transfer fee; Hedonic-model; Cross-sectional analysis; Poisson Quasi-maximum-likelihood estimation; Prediction

JEL CODES: B23; C01; C21; C53; Z23

TABLE OF CONTENTS

Glossary i

Abstract, keywords and JEL codes ii

Figure Index iv

Table Index v

Acknowledgments vi

1. Introduction 1

2. Football – an overview 3

  2.1. Origins and evolution 3

  2.2. Football dimensions 4

  2.3. Football Economics 5

  2.4. Players' transfers 7

3. Player characteristics, hedonic models and transfer fees 9

  3.1. Measuring the determinants of transfer fees through the use of hedonic price model 9

  3.2. Methodological framework 11

4. Data and descriptive statistics 15

  4.1. Dependent variable – transfer fee 17

  4.2. Explanatory variables 18

5. Empirical application 21

5.1. Regression results and model selection 21

5.2. Selected model interpretation 23

6. Player valuation discrepancies: market valuation *vs.* model predictions 26

7. Conclusion 30

References 33

Websites 37

Appendices 38

FIGURE INDEX

FIGURE 1 - European football market size 2011/12 - 2020/2021 (€ billions) 5
FIGURE 2 - Big-Five European league clubs' revenue 20/21 (€ millions) 6
FIGURE 3 - Brand value comparison - 2018/20 (€ billions) 7
FIGURE 4 - Player’s contract decision tree 9
FIGURE 5 - Transfer windows and football season moments 17
FIGURE 6 - Transfer Fee plot and histogram 18

TABLE INDEX

TABLE I - Transfer fee's descriptive statistics (10 thousand euros) 18

TABLE II - Regression results: log-linear and Poisson models 25

TABLE III - Predicted transfer fee values and outcome, based on the PQML estimated model.. 29

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to all those who have stood beside me throughout this journey, which ultimately led to the successful completion of this master´s final work.

To all the professors of the Master in Applied Econometrics and Forecasting for all their teachings, which were the initial foundation for making of this dissertation.

A special, and sincere note of gratitude to my professor and supervisor Esmeralda Arranhado, for the guidance and wavering support along every step of the process, and for always pushing me forward, to achieve the best outcome.

To my esteemed professional team leader Patrícia Gomes, for all the support provided when most needed, and all the advice given that were indispensable.

To my family, including my mom and dad, as well as Luísa, for always be there during times of utmost necessity, and for fostering an environment that promoted the successful completion of this dissertation.

Finally, my deepest gratitude goes to my beautiful girlfriend Mariana for all the encouragement and constant motivation. Your love, comprehension, and patience have been my source of strength in this journey. I'm truly grateful for your presence in my life, as well as your sacrifices and selflessness for my academic and professional aspirations.

1. INTRODUCTION

Football presents itself as one of the most iconic sports, with billions of followers all around the world. Apart from a social perspective, football can be seen as an industry, that can be compared to the biggest enterprises in a very diversified industry portfolio. Football generates its profit through different segments: broadcasting rights, sponsorships, or ticket and merchandising sales. Nonetheless, these profit generators are always linked to the sporting performance of each club. The connectivity between these links is established by a specific agent: the football players, upon whom a team depends to attain favourable outcomes. Good results lead to better sponsorship deals, more ticket and merchandising revenue, fostering interest from possible investors. Ultimately, more economic power, gives the club available cashflow to invest into the best players on the market, re-initiating the cycle.

For the period of interest (2022-2023) of this dissertation, the major five European leagues, in terms of sport and economic performance, are: Premier League (England), Bundesliga (Germany), LaLiga (Spain), Serie A (Italy) and Ligue 1 (France). The competitiveness level of these leagues and their gigantic economic power, makes it possible to attract and retain the more talented players worldwide, elevating the sporting level to its peak.

Being the players the centre key for all football sporting and economic activities, it is of great interest to analyse what drives the transfer fee agreed by two clubs, to acquire a specific player, based on its unique set of characteristics. A parallel type of analysis is commonly performed using hedonic methodology in the housing market, where the asset in that case are properties. The aim is measuring the effect of the housing characteristics, on their corresponding prices. Applying a similar approach, hedonic models can be used to measure the impact of the players' characteristics on their transfer fee.

Different studies were already done on this topic. An early example is given by Carmichael and Thomas (1993) that examined transfers from the 1990-91 English football season, applying two-person bargaining theory, to investigate the determination of transfer fees. Later, Dobson and Gerard (1999) analysed the effect on the transfer fee in the English professional football, based not only the player characteristics, but also on the buying/selling club characteristics. Also, they considered time effects and the different market segmentations, using then the estimated model to investigate the rate of inflation in transfer fees. Another example can be found in Lucifora and Simmons (2003), that using a similar approach, tried to analyse the relationship between a player's individual productivity and salary, which can lead to a superstar effect. A recent illustration can be found on Ezzeddine (2020).

Using hedonic regression models, this dissertation aims to analyse data collected from 503 player transfers regarding the season 2022-23, uncovering the transfer fee drivers, based on the collected sample. Moreover, given the fact that the majority of the existent literature focus exclusively on using simple linear models estimated by the ordinary least squares (OLS), there exists a gap on using more complex models. This dissertation tries to fulfil this gap, using a nonlinear model, the Poisson, estimated by quasi-maximum-likelihood (QML).

The construction of a model that can correctly and with precision evaluate the drivers of a player's transfer fee, is not only of interest from a theoretical and academic point of view, but can also be a tool of great use for decision making. After a first diagnosis analysis on the collected data, the regression results of the different set of models considered are presented, with one of them being selected for further inference. The estimated regression model, will then be used to analyse if the real transfer fee paid is over/undervalued compared to the predicted value of that player, supporting buying decisions with statistical information. The period of interest for this prediction exercise will be the season 2023-24, including an additional league from Saudi Arabia.

This dissertation is structured as follows: section 2 provides a comprehensive examination of football as a social and economic phenomenon, encompassing its origins and evolution. Section 3 explores the utilization of hedonic models to determine whether there is indeed a causal relationship between the transfer fee and the characteristics of the players and clubs. Furthermore, the section encompasses the methodological framework and literature support that will be utilized for modelling purposes. In section 4, the collected database and its primary descriptive statistics are presented, offering a thorough understanding of the analysed data. Section 5 shows the regression outcomes derived from both linear and non-linear model, as elucidated in the preceding section. Thereby shedding light on the estimated factors that have influence on the determinants of the transfer fee. Subsequently, the selection of the most appropriate model among the alternatives is undertaken. In Section 6, using the selected model from the previous section, an examination is conducted through a comparative analysis exploring the disparities in player valuation, between model-based predictions and actual market valuation. The final section encompasses the presentation and evaluation of the principal conclusions drawn from this dissertation.

2. Football – an overview

2.1. Origins and evolution

The roots of football, take us back in time for more than 2000 years, according to Fédération Internationale de Football Association (FIFA) Museum. The most famous ball games were played in Greece (ephebike or epikoinos), Rome (harpastum), Meso-America (ball-game), China (Cuju), Japan (Kemari), using the feet, hands and also the hips in the Meso-American game. Throughout history various forms of village football arise, played sometimes with teams composed by hundreds of players with the simple objective of forcing the ball by running with it, kicking it or by any other means previously agreed to pass it through the goals; Murray and Murray (1994).

Before the industrial revolution, football was largely seen as plebeian entertainment, without any set of defined rules, but nonetheless captured a lot of attention and enthusiasm. As a result, between 1850 and 1870 numerous football clubs would emerge, such as the historic English club Liverpool. On 26 October of 1863, the Football Association (FA) was founded in England with the purpose of promoting the adoption of a general code of rules for football. Football then began to spread all over the world, capturing the attention of all society classes, and many new clubs and federations were created; Collins (2018). In May 1904, the representatives of the seven European football nations met in Paris to establish FIFA as the organism responsible for overseeing football associations in the different continents, imposing football rules worldwide, as they are known nowadays. FIFA recognizes 6 football confederations that oversee continental football: Asian Football Confederation (AFC), Confederation of African Football (CAF), Confederation of North, Central American and Caribbean Association Football (CONCAF), Confederación Sudamericana de Fútbol (CONMEBOL), Oceania Football Confederation (OCF), and Union of European Football Associations (UEFA) – being the last one the focus in this dissertation.

From simple games played in the ancient times, to organized games with a defined set of rules, using modern technology such as the video assistant referee (VAR), football evolution was astonishing throughout history.

2.2. Football dimensions

Football embodies a universal, cross-cultural mode of communication that is comprehensible to every individual across the globe. The core fundamentals of this sport are readily grasped and involve the act of propelling a ball towards a designated goal. The universality of football is undeniable, and it knows no age bounds. According to FIFA, a staggering number of 5 billion individuals are devoted followers of this sport across the globe, with Latin America, the Middle East, and Africa being the regions harbouring the most substantial core of enthusiasts. Considering the earth's population - approximately 8 billion individuals (according to Worldometers), it can be inferred that around 62.5% of individuals worldwide are football fans. Sports are widely viewed as advocates of social inclusion, and football is no exception. Community-based initiatives have been implemented to leverage sports as a catalyst for achieving various societal goals, including but not limited to: promoting a healthy lifestyle, involving adolescents in both formal and informal learning, preventing criminal and anti-social behaviours among younger generations; Tacon (2007). Throughout the years, football leagues, associations, clubs, and FIFA itself have undertaken numerous initiatives aimed at advancing social inclusion and equality.[1]

When analysed from a cultural perspective, the influence of football on a worldwide level is of considerable magnitude. Fans' unwavering support for their teams is universally recognized and spans generations, leading to long-standing rivalries between rival fans that can last over a century. Examples thereof include the hostility between Sporting and Benfica, or the rivalry between Manchester City and Manchester United fans. The impact of football culture on the behaviour of individuals is apparent not only within the stadium, but also extends beyond it. Sports bars are animated with zealous supporters cheering for their beloved teams every time their club plays, in a fashion that bears resemblance to the devout attendance of Christians at church every week. Despite the diverse club affiliations, individuals worldwide come together to

[1] FIFA campaigns: Education for All; UEFA campaigns: HatTrick development programme; Premier league: No room to racism.

support their country's national team during any form of football competition, showing their fervent loyalty by vociferously singing the national anthem with a level of emotional investment that almost fuses their own identity with that of the players.

2.3. Football Economics

Football has transitioned from a sport-centric activity to an expansive and noteworthy commercial sector on a global scale; Hamil and Chadwick (2009). Clubs' commercial earnings are generated through different sources: merchandising, tickets, broadcasting agreements and sponsorship; Garcia-del-Barrio and Pujol (2004). Also, the reliance on the stock market is an important funding source to support football development, dating back to 1983 when Tottenham Hotspurs listed on the United Kingdom's stock market; Scholtens and Peenstra (2009). Moreover, the sport performance of a listed club, was demonstrated to have positive or negative impact depending on the club's results, on the associated stock price; Palomino *et al.* (2005).

As one can observe from Figure 1 the European football market exhibited a valuation of 27.6 billion euros in the 2020/21 season, presenting a growth rate of 42.27% in contrast to the 2011/12 season. During the 2018/19 season, the market value attained its apex, resulting in a total sum of 28.9 billion euros. Notwithstanding, the influence of the pandemic in the subsequent year resulted in a reduction of 12.8%, causing a reversion to the preceding values observed in the 2016/17 season. Despite the challenges it faced, the football industry's resilience and competitiveness enabled it to bounce back very effectively in 2020/21, resulting in a significant boost to its market value of approximately 9.52%.

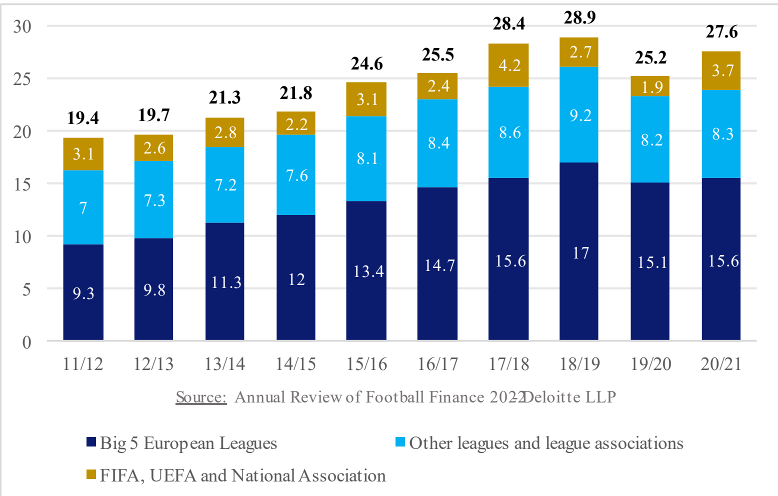


FIGURE 1 - European football market size 2011/12 - 2020/2021 (€ billions)

Analysing the information presented in Figure 2, on the 2020/21 season the Big-Five[2] present as the highest contribution for the total revenue Broadcasting agreements, with a weight of more than 50% on each league. The cumulated revenue generated by the Big-Five surpassed the mark of 15.5 billion euros. This figure accounts for over 50% of the European market's worth during the same period.

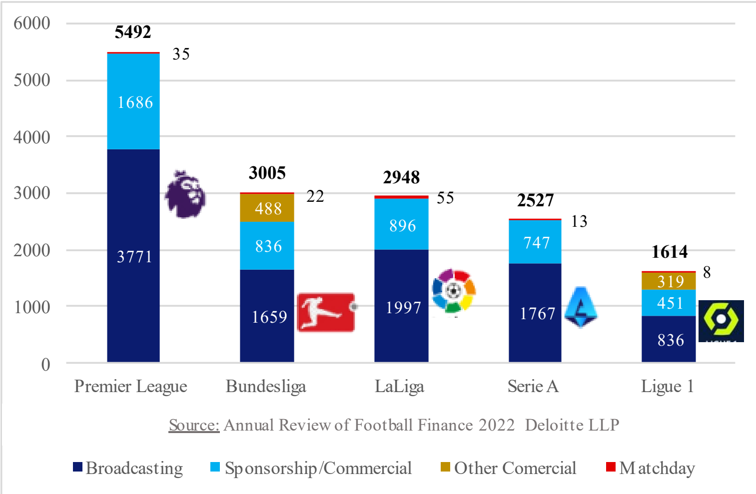


FIGURE 2 - Big-Five European league clubs' revenue 20/21 (€ millions)

At club level, among the five different leagues regarding season 2021/22, attributes the total revenue podium to: Manchester City (Premier League), Real Madrid (LaLiga) and Liverpool (Premier League) – with a total revenue respectively of 731, 714 and 701 million euros; for full details see appendix – Table A.1.

For a broader view of the huge dimension of the European football market, I compare it with some of the world biggest brands, well known by the public, namely: Pepsi (food and beverage industry), LG (electronics industry) and Vodafone (telecommunications industry), between the years of 2018 and 2021. As one can simply conclude by observing Figure 3, in all 4 years the Big-Five had a brand value higher than the other three brands considered. A more attentive analysis and upon further investigation, there is a small detail that makes a huge difference when making a comparison. Pepsi, LG, and Vodafone manage their business across, respectively 200,

[2] Big-Five European leagues: a) Premier League (England), b) Bundesliga (Germany), c) LaLiga (Spain), d) Serie A (Italy) and e) Ligue 1 (France) – accordingly to the UEFA coefficient, for the most powerful European leagues.

128 and 21 territories (and this is where it gets interesting) the Big-Five are present only in their respective home country, and each league that compose the Big-Five are one single entity. So, when comparing it to these companies one can conclude that the Big-Five have a nominal value close to them, but analysing its real brand value, it surpasses the intrinsic value of the compared entities, exhibiting the massive economic and financial power of the European football leagues.

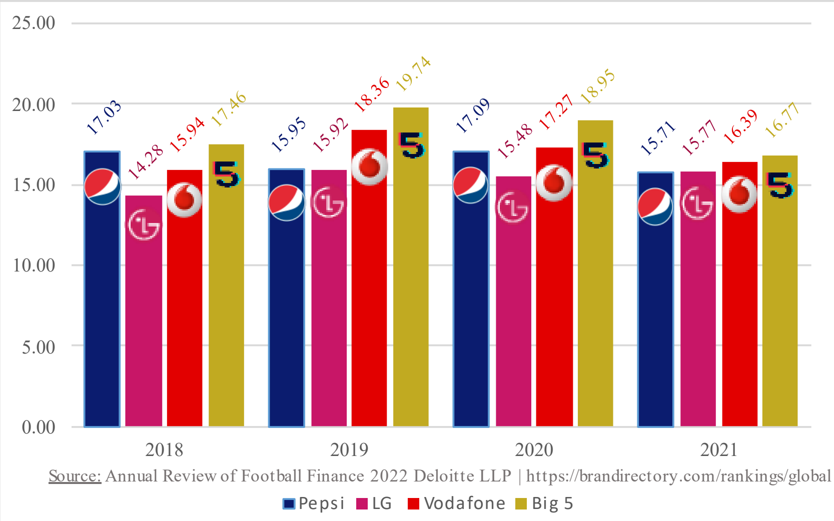


FIGURE 3 - Brand value comparison – 2018 - 21 (€billions)

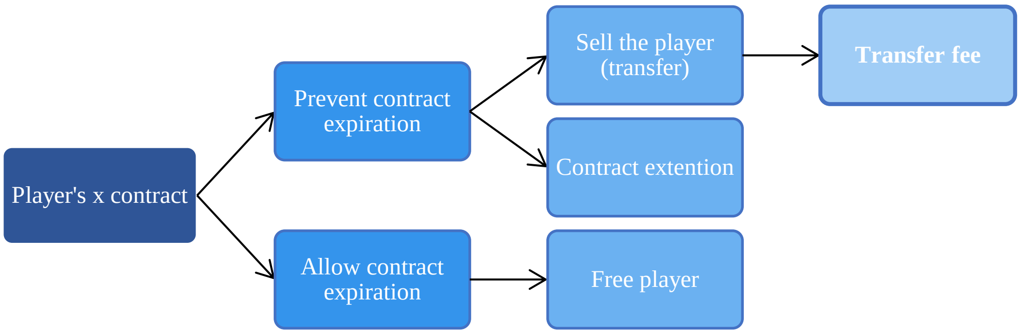
2.4. Players' transfers

The players' transfers and the ensuing rumors that circulate prior to their realization, holds significant importance within the football industry, being his associated value (fee) the main variable in analysis. To analyze it properly, it must be understood from its origins and evolution, until the present. The football transfers significance, extends beyond the desires of fans to recruit the most talented players to their respective clubs, and witness their contribution towards securing championships, and triumphs in major competitions. Also, the players aspire to compete for top-tier clubs in prominent leagues and tournaments, as well as earn larger incomes through their careers. Additionally, there is a significant advantage for clubs to possess the most skilled players, given the higher likelihood of winning championship titles and attaining international recognition. Finally, from an economic perspective revenue will in principle be higher, assuming that the squad is composed of as many football superstars as possible.

The FA implemented the first transfer system in 1885. Clubs were required to register their players annually, and the renewal was based on the discretion of the latter, possessing the authority to hold a player even if the contract had not been renewed; Dobson and Goddard (2001). The players were only permitted to transfer to a new club if the current one chose to sell them or terminate their contract - “retain-and transfer system”; Sloane (1969). In 1963, this system was abolished and was introduced an "option-and-transfer system" allowing a player to be contracted for a pre-determined period with the club having the option to extend it. If not exercised, the player became a free agent. In 1977, "freedom of contract" was established, and the players became free agents if their current club failed to offer terms that were at least as good as the ones of the previous contract final year. Even though this was an evolution towards the present transfer system, it did not remove the selling club’s property rights over a player since it still allowed for a fee at the moment of the transfer when the player was already at the end of the contract; Gerrard and Dobson (2000).

In 1995, the Bosman ruling constituted a significant legal landmark that established the modern automatic free transfer system, following the consolidation of the legal cases related to Jean-Marc Bosman. He expressed a desire to be transferred upon the expiry of his current contract, but its club requested a transfer fee that exceeded the buying club’s willingness to pay. Unable to move, Bosman took legal action against the club, the Belgian football association and even UEFA. The court declared that the previous regulation impeded the unrestricted mobility of laborers, which is guaranteed by Article 48 of the Treaty of the European Union. Deciding in favor of Bosman and originating the famous “Bosman Ruling”; Radoman (2017). This decision serves as a pivotal moment in football history. While its impact yielded increased autonomy for football players, it inadvertently catalyzed the sport's present state, wherein clubs with the higher financial power have an even greater chance of achieving the best sporting results over smaller clubs.

This changes in the football transfer system, have given rise to the football market as we know it nowadays, simple to understand and fair, both for the players and the club – Figure 4 illustrates in a simple format the essence of the football transfer market, where in the upper part one can observe what kind of decision give rise to the variable of interest of this dissertation (transfer fee)



Source: the author

FIGURE 4 - Player's contract decision tree

3. PLAYER CHARACTERISTICS, HEDONIC MODELS AND TRANSFER FEES

The connection between a player's characteristics and the transfer fee, is based on the idea that the fair transfer worth of a player, is reliant on his characteristics and intimately tied to his success on the football pitch. A player with exceptional (poor) performance, is typically associated with above-average (below-average) traits, which in principle is translated to a larger (smaller) transfer fee. Furthermore, if the buying and/or selling club have a track record of success (either economic or in sporting terms), the transfer fee agreed upon by the two clubs can be expected to be higher; Frick (2007).

3.1. Measuring the determinants of transfer fees through the use of hedonic price model

The application of the hedonic price model has become a prevalent practice within the real estate industry, as a mean of measuring the effect of specific property attributes on their corresponding price. The first evidence of an early similar approach to the hedonic framework was performed by Waugh (1929). He regressed prices of different types of asparagus on their color, diameter, and homogeneity, to assist farmers in meeting market demands. Ten years later, Court (1939) published an article mandated by General Motors, to defend the company against Congress' accusations of monopolistic price pushing. It is often viewed as the first article on hedonic models, although the study actually developed a hedonic price index for automobiles.

The hedonic price formal theory was developed and well defined, by the work of Rosen (1974), where he defined hedonic prices has: "*the implicit prices of attributes and are revealed to economic agents from observed prices of differentiated products and the specific amounts of characteristics associated with them.*". Therefore, the hedonic-pricing methodology employs statistical analysis to deconstruct the price of a composite asset into a sequence of implicit (hedonic) prices for each characteristic that comprises the distinctive asset; Gerrard (2001). More recently Malpezzi (2003) argues that the hedonic model arises from the market's inherent heterogeneity and the diversity of consumer preferences - the unique attributes of each asset are not only distinct, but also hold varying preferences for consumers. Another excellent alternative of hedonic pricing models review is given by Sirmans *et. al.*, (2005).

As one may easily conclude from the above stated, the hedonic approach can be applied to different assets, and not just limited to the real estate market, granted that they possess quantifiable attributes that can be deduced from the composite asset. As per the International Accounting Standards 18 description of asset: "*A resource controlled by the entity as a result of past events and from which future economic benefits are expected to flow to the entity"* - a football player fits perfectly into this description, as well as a house. Therefore, it is reasonable to use a hedonic-type framework to uncover the determinants of a house price, just as it can be used to ascertain the transfer fee determinants of a professional football player.

For the specific case of football, numerous studies have focused on the valuation of transfer fees, with three worth mention works by Carmichael and Thomas (1993), Dobson and Gerrard (1999), and Gerrard (2001). These studies employ hedonic-pricing methods, revealing a causality between the transfer fee and the distinctive characteristics of players. Age, appearances, goals scored and also the characteristics of the buying and selling clubs are examples of covariates included in these studies. More recent illustrations are given by Karnik (2010) for cricketers[3], and by Ezzeddine (2020) for the specific topic under discussion.

[3] Cricketer: an athlete who plays cricket

By estimating a hedonic price function composed by the different characteristics associated to the heterogenous players, it is then possible to statistically assess the effect (positive or negative), the magnitude, and the implicit price for changes in each attribute, that determine the value of the variable of interest. Empirically the function of the transfer fee can be written as:

$$TF_i = f(P_i, C_i), \tag{1}$$

where TF represents the transfer fee, $i = 1,2, \dots n$ denotes the player, n is the sample size, P are the k_1 attributes of the target player, and C are the k_2 attributes of the buying club.

As stated by Evangelista *et al.*, (2019) the functional relationship between prices and characteristics, is a fundamental and central element within the hedonic price model, and can be expressed as:

$$TF_i = \beta_0 + \sum_{k_1=1}^{k_1} \beta_{k_1} . P_{k_1,i} + \sum_{k_2=1}^{k_2} \theta_{k_2} . C_{k_2,i} + u_i, \tag{2}$$

where β and θ are the parameters to be estimated; and u is a term that represents the additional random factors that are not captured by the $k_1 + k_2$ variables included in the system. Which is assumed to satisfy the exogeneity assumption:

$$E(u_i | P_i, C_i) = 0 \tag{3}$$

In the football context, commonly used covariates for P are: age, appearances in national team (Lucifora and Simmons, 2003), player's height (Bryson *et al.*, 2013), number of goals, and assists and passes (Poli *et. Al*., 2021). Also, numerous studies use dummy variables for the specific position of the player: attacker, midfielder, defender, or goalkeeper (Ruijg and Van Ophem, 2015). Regarding C, the league position of the buying/selling club (Dobson and Gerard, 1999) is one of the most used covariates. In contrast, there are some explanatory variables that have received relatively less attention in the existing literature. Some are included in the dataset considered, such as: rate of successful passes, and a dummy variable accounting for the location of the transfer between European clubs, or not.

3.2. Methodological framework

As mentioned in section 2.1., the applicability of the hedonic framework is not limited to the real estate market but can actually be applied to different areas. Several studies have been conducted in distinct areas, including: wine selection (Panzone and Simões, 2009), food products (Giombi *et. al*., 2018), energy efficiency (Fesselmeyer, 2018), impact of traffic noise (Nelson, 1982), air pollution (Fernández *et. Al*., 2012). and also the niche market of art (Arvin and Scigliano, 2004).

Given the strong economic dimension of the football industry and the general interest worldwide, it is a topic of great interest in economic research, with some examples given by Malcolm (2000), Roberts *et. al*. (2016) and a more recent study is given by Bernardo *et. al*. (2021). The hedonic framework and football economics, when correctly combined can generate meaningful conclusions and insights, with relevant information not only for the general market overview, but also for team management and stakeholders. That said, a hedonic price model will be used in the next section, in order to determine the impact and magnitude of a certain set of characteristics, on the players' transfer fees.

Since the variable of interest (transfer fee) is a non-negative continuous variable, the use of a log-linear model is an adequate candidate to explain its relationship with the explanatory factors to be included in the regression. The use of the natural logarithm of the dependent variable will result in an improved symmetry of its distribution, and so less prone to the presence of outliers. This typology of models is commonly used in the literature, by different authors associated with hedonic framework. Following Lucifora and Simmons (2003), and using equation 2 as base model, and $x_i = (P_i, C_i)$ the empirical specification will be:

$$l\,n(TF_i) = \beta_0 + x_i'\beta + u_i, \tag{4}$$

where $l\,n(TF_i)$ represents the natural logarithm of the transfer fee of the i^{th} player. Considering the heteroskedastic nature of the data, it's expected that a standard OLS estimation will produce invalid standard errors and invalid inference. To account for this problem, the White (1980) robust covariance-matrix estimator, informally called the sandwich estimator, will be used.

Log-linear models are standard in the hedonic literature, but they may create problems, namely when the aim is predicting in the original scale. Therefore, more general alternatives such as non-linear models are considered. Given the continuous and nonnegative nature of the variable of interest, the Poisson model appears to be an appealing candidate to model the data.

Following Wooldridge (2001), the Poisson distribution, commonly used for counts, implies in this specific case, the following probability mass function conditional on x_i:

$$P(TF_i|x_i) = \frac{exp\{-\mu_i\}\, {\mu_i}^{TF_i}}{TF_i!}, \tag{5}$$

where $TF_i!$ is TF_i factorial (with $0! = 1$) and $\mu_i = E(TF_i|x_i)$ represents the conditional mean. The most common mean function assumes an exponential form: $\mu_i = E(TF_i|x_i) = exp(x_i'\beta)$, $i = 1,2, \dots n$.

The equality between the first two moments (equidispersion), is one of the most relevant distributional assumptions of the Poisson distribution, that is:

$$Var(TF_i|x_i) = E(TF_i|x_i) \tag{6}$$

Given (5) and the mean specification, and assuming that the observations $(TF_i|x_i)$ are independent, as in Cameron and Trivedi (2005), the maximum likelihood (ML) emerges as a suitable estimator. Being the log-likelihood function:

$$Log\ L(\beta) = \sum_{i=1}^{n} \{TF_i - x_i'\beta - exp(x_i'\beta) - lnTF_i!\}, \tag{7}$$

the Poisson maximum-likelihood estimator, denoted as $\hat{\beta}_p$, is the solution to K nonlinear equations, that corresponds to the first-order condition:

$$\sum_{i=1}^{n} \{TF_i - exp(x_i'\beta)\}x_i = 0 \tag{8}$$

The continuous nature of the variable of interest under analysis, requires that the Poisson quasi-maximum-likelihood estimator (PQMLE) is employed. This estimator is equivalent to (8), having identical first-order conditions to that of ML. The summation on the left-hand side has expectation zero if $E(TF_i|x_i) = e^{x_i'\beta}$, thereby ensuring the consistency of the Poisson quasi-maximum-likelihood (PQML) under the weak assumption of correct mean specification. Therefore, the data does not need to have a Poisson distribution, and TF_i does not have to be an integer for the estimator to be consistent; Gourieroux et. al. (1984).

The difference between ML and quasi-maximum-likelihood (QML) for the Poisson distribution, lies in the different variances among the estimators; Cameron and Trivedi (2005); That is:

$$Var_{QML}(\hat{\beta}_p) = \left(\sum_{i=1}^{n} \mu_i \, x_i x_i'\right)^{-1} \left(\sum_{i=1}^{n} \omega_i \, x_i x_i'\right) \left(\sum_{i=1}^{n} \mu_i \, x_i x_i'\right)^{-1} \quad (9)$$

$$Var_{ML}(\hat{\beta}_p) = \left(\sum_{i=1}^{n} \mu_i \, x_i x_i'\right)^{-1},$$

where $\omega_i = Var(TF_i|x_i)$.

Although PQMLE is commonly used as a count data estimator, it is also suitable for regressions with continuous data; Santos Silva and Tenreyro (2006). Being the nature of the target variable, non-negative and continuous, PQMLE presents itself as a viable option for modelling the data.

Nonetheless, in several cases the data rejects the Poisson assumption that the variance is equal to the mean. Relaxing this assumption, but maintaining the mean specification as $exp(x_i'\beta)$, one can denote the conditional variance of TF_i, as:

$$\omega_{NB_i} = Var(TF_i|x_i) \quad (10)$$

According to Cameron and Trivedi (1998), one can continue to model the variance as a function of the mean, for some function of $\omega(.)$, that is: $\omega_{NB_i} = \omega(\mu_i, \alpha)$ where α is a scalar parameter. Using the general variance function:

$$\omega_{NB_i} = \mu_i + \alpha \, \mu_i^p, \quad (11)$$

where p is a specified constant. Except for the Poisson case, encompassed when $\alpha = 0$ (equidispersion), the analysis is limited to two cases, the negative binomial 1 and 2, that set $p = 1$ and $p = 2$ respectively. The negative binomial 2 (NB2) with density :

$$f(TF|\, \mu, \alpha) = \frac{\Gamma(TF + \alpha^{-1})}{\Gamma(TF + 1)\Gamma(\alpha^{-1})} \left(\frac{\alpha^{-1}}{\alpha^{-1} + \mu}\right)^{\alpha^{-1}} \left(\frac{\mu}{\mu + \alpha^{-1}}\right)^{TF}, \quad (12)$$

where the function $\Gamma(.)$is the gamma function, can be estimated by QML, as long as the scalar parameter α is known. Given that in this application, and in the generality of the practical situations, α is not known, the NB2 is not an option for modelling the nonnegative continuous variable of interest.

The aforementioned models and estimation methods, however, can suffer from a misspecified conditional mean functional form, leading to the inconsistency of the estimators. To account for this issue, and test for correct form specification the test proposed by Ramsey (1969) will be employed. Ramsey proposed a regression specification error test (RESET), which has proven to be a valuable tool for identifying potential functional form misspecification.

Considering the linear model presented in Equation (2) and let $\widehat{TF}_i^p$ (with $p = 2{,}3$) denote the OLS fitted values from the estimation, the test implementation is done with two steps. Firstly, an augmented equation will be estimated, adding powers of the fitted values (squares and cubes) which will be just non-linear functions of the covariates:

$$TF_i^* = \alpha_0 + x_i'\alpha + \rho_1\widehat{TF}_i^2 + \rho_2\widehat{TF}_i^3 + error, \qquad (13)$$

Secondly, using Equation (13) the RESET will simply be applied as the F-*statistic* for testing the null hypothesis (H_0) against the alternative (H_A):

$$H_0: \rho_1 = \rho_2 = 0 \quad vs. \quad H_A: \rho_1 \neq 0 \vee \rho_2 \neq 0 \qquad (14)$$

Under H_0, the model has a correct functional form specification as there is no evidence of any other non-linear functions being statistically significant. Consequently, the model can be used for subsequent inference. The F-*statistic* distribution is approximately $F_{2,n-k-3}$ (where n and k represent respectively, the number of observations and the number of covariates); Wooldridge (2013). Under H_A the functional form is incorrectly specified. The model potentially suffers from omitted variables, and/or the relation between the dependent variable and one or more covariates is not linear, but rather quadratic or cubic. For nonlinear models, the RESET test is performed as a Wald test in the framework of QML estimation. Specifically, the significance of the two considered powers of the fitted liner index is tested.

4. Data and descriptive statistics

The dataset exploited in this dissertation includes the dependent variable, transfer fee, retrieved from the digital football platform Transfermarkt. This platform contains a large amount of information regarding football - including but not limited to: results, statistics, transfer news and player values. Nonetheless, the information provided by Transfermarkt on the players' and clubs' characteristics (covariates) was deemed relatively superficial from an econometric standpoint. Consequently, it did not offer any noteworthy enhancements compared to previous studies on the topic. After a strong and well-structured research, it was possible to assemble a complete and rich dataset of explanatory factors, resorting to FootyStats - a football analysis website covering club, league and player stats covered in (almost) full detail.

In the 2022-23 season a staggering 700 transfers[4] took place among the Big-Five. Nonetheless, to enhance the dataset and provide a more comprehensive analysis, the decision to incorporate the Liga Portugal was made – bringing the total number of transfers up to more than 900 observations. Considering the varying levels of attention given to minor leagues or youth football divisions, and the lack of technological development, the data collection on certain players was limited. Consequently, only players with the full set of characteristics available were included in the analysis. After combining the two separate databases, a comprehensive and diverse dataset was created, encompassing information on 503 transfers from the 2022-23 season, excluding loan agreements or free transfers.

Before analyzing the data further, it is important to understand the complexities of the football calendar and its transfer windows. The football season can be broken down into three key periods: 1 - Summer window (SW); 2 - Football season; 3 - Winter window (WW). The SW marks the first opportunity for athletes to be transferred, spanning a three-month timeframe from July to September. This period aligns with the start of the football season, which spans from August to June next year. Roughly halfway through the football season the WW is available for new transfers with a shorter period of only two months, between January and February.

[4] The term transfer refers to the trade of a certain player in exchange for a pre-determined price. It is considered a transfer whenever a club from the six leagues under analysis sell or buy a player that can be from an inferior or superior league.

For a comprehensive understanding of these main moments, Figure 5 illustrates in a simple way its different durations and dates.

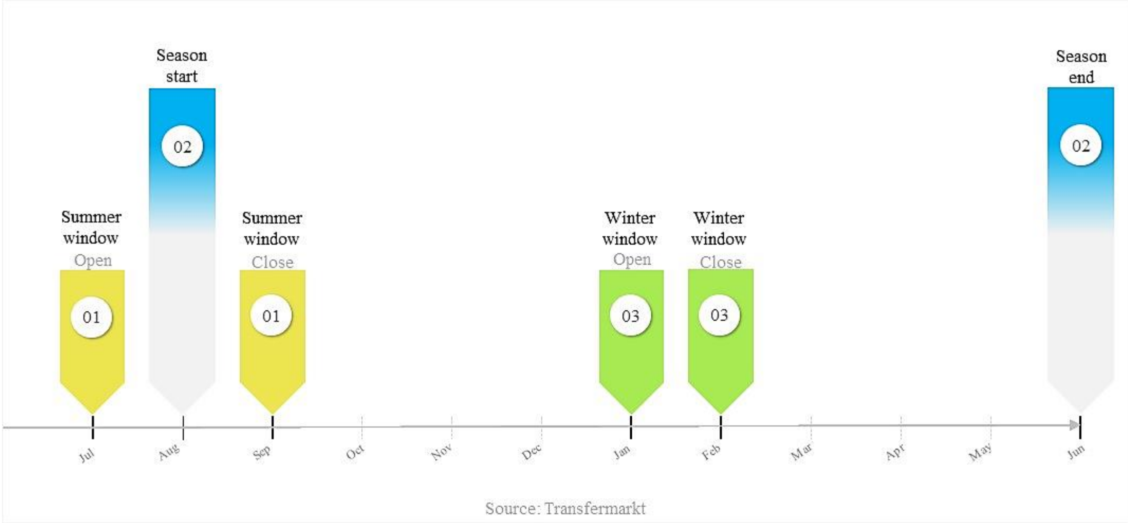


FIGURE 5 - Transfer windows and football season moments

The dates presented pertains to the 2022-23 football season and may not be applicable to all preceding and forthcoming seasons. Also, among the six leagues there is a small difference in the start and closing dates of each of the moments – a full description of these dates is presented in the appendix (Table A.I.)

4.1. Dependent variable – transfer fee

The transfer fee (TF) represents, as stated in the previous section, the amount agreed by two different clubs (the seller and the buyer) for the transfer of a specific player. Moreover, the transfer fee can have future added values to the initial agreed value, depending on the future performance of the player. This type of transfers is not considered since the interest lies on the past performance until the time of the transfer.

The transfer fees are in the context of professional football a variable that presents very high values, reaching in this dataset a minimum of 50 thousand euros, and a maximum of 121 million euros (see Table I), resulting into a wide degree of amplitude between observations, confirmed by the standard deviation value of 14,807,670. To enhance interpretability, all the 503 observations will be measured in ten thousand (10,000) euros. $TF^{\#}$ will then represent the transformed dependent variable that will be used for model estimation purposes. The variable does not exhibit a normal distribution behavior, with the values of the Kurtosis and Skewness lying far from the standard values.

Presenting a leptokurtic and left-skewed behavior – which can also be confirmed by the difference between the mean and median values, presented in Table I.

TABLE I

TRANSFER FEE'S DESCRIPTIVE STATISTICS (10 THOUSAND EUROS)

Transfer fee ($TF^{\#}$)						
Mean	*Median*	*Std. Deviation*	*Min*	*Max*	*Skewness*	*Kurtosis*
1024.24	450	1480.77	5	12100	3.11	15.56

As anticipated and elucidated in the preceding sections, through visual inspection of the plot and the histogram presented in Figure 6, one can observe some sort of heterogeneity, and a concentration on lower values, confirming the previous statistical values.

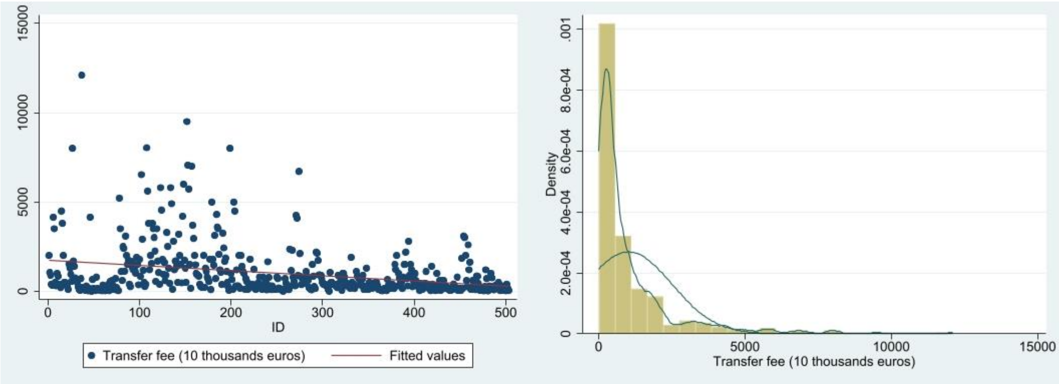


FIGURE 6 - Transfer fee plot and histogram

4.2. Explanatory variables

The covariates extraction process posed a significant challenge to overcome, as there was a lack of available databases that could directly yield the target data with the specific predefined attributes in detail. Given that, all the covariates were extracted player by player to ensure that the best set of differentiating characteristics were collected. This yields the dataset uniqueness, compared to previous studies.

It was duly acknowledged that the variables ought not to be solely gathered based on their quantitative value, but rather in terms of their qualitative relevance. The reason for this choice is related to different factors, such as: the difference in the number of total games of the six leagues considered is not equal, so different players have more (less) opportunities to increase (decrease) their stats. Additionally, the presence in major

European competitions, such as the UEFA Champions League (UCL) or even domestic cups, can influence the choices of each coach to play with a certain player, planning on recovery, for competitions perceived as more important. Even the coach or management decisions to play different styles of football (more defensive or offensive) can influence some set of players to play more or less, throughout the season. That said, the majority of the covariates were extracted in terms of success rate for a certain action, or as ratio of the overall standard play time by game (90 minutes). The previous literature shades little light on this type of variables, but they appear to be very powerful and informative, in a hedonic framework.

As stated previously, there are two different periods where the players can be transferred amongst two clubs, the SW and the WW, so the data collected had to be adjusted accordingly. That is, for the SW of the 2022-23 season was considered all the information available from the previous season (2021-22) until the 30^{th} of June 2022. In respect to the WW, it was considered the information of the previous season, plus the information until 31^{st} of December 2022. The reason for that, is related with the idea that a football club will be interested in a specific player given its past performance with the main objective to achieve equal or better future performance influencing positively the club results.

Regarding the typology of players, they can be divided into 4 different positional categories: attackers, midfielders, defenders, and goalkeeper. All these positions have a different set of intrinsic characteristics, e.g., an attacker is expected to have more goals than a defender or a midfielder more passes than an attacker. Therefore, the covariates selection was made taking into consideration the different set of intrinsic characteristics of each positional category. Having that way balanced analysis, and also a broader view and understanding of the impact of each category characteristics on the transfer fee. Nevertheless, the goalkeepers will not be considered in the analysis since this type of position needs a completely different set of characteristics to be possible to analyze – a goalkeeper is not expected to have any goals or assists, and the number of passes and dribbles will be close to zero. Since the main function of a goalkeeper is to maintain a clean sheet, the analysis should be done separately from the rest of the remaining positions, contrary to some studies that included them in the analysis and model estimation.

With a total of more than 30 covariates extracted, a pre-selection is needed to select the ones that can indeed help to explain the drivers of the football transfer fee. Resorting to a preliminary visual inspection, some of them were immediately dropped from the analysis given the weak or non-existent relationship, between them and the dependent variable. Additionally in a posterior analysis and relying on a correlation matrix analysis, covariates with a high correlation (>80%) amongst them were also dropped from the analysis. On the other hand, given the mediatic dimension of football players, they are usually associated to marketing campaigns via social media, not only for the club represented, but also for brands in the most diverse industry sectors (e.g., sports, cosmetics, fashion). So, a variable that can capture the added value for a transfer fee taking it into consideration the mediatic attention of a player, is of great interest. Nonetheless the actual social media platforms like Instagram, Facebook or Twitter do not provide a tool that allow to verify the mediatic exposure of an individual or collective person through the number of followers in a certain point in time, but only in real time. The interest for this type of variable, will be the number of followers of a specific football player in the moment before the transfer take place, to analyze whether a transfer value is affected by the recognition of a player through the social media. Given the impossibility of extract this type of variable in different points in time, it will not be included in the dataset since the inclusion considering the number of followers in the present, will be in principle affected by the transfer itself introducing bias to future estimates. Following the aforementioned preliminary examination, the outcome yields a dataset comprising 19 covariates, featuring a diverse array of players with distinct characteristics – a full description of each variable is presented in the appendix (Table A.II).

The dataset players' ages span from 18 to 34 years, with the average player being approximately 24 years old and playing 67.5 minutes per game. Amongst the different variables considered in this phase, it's noteworthy that seven of them were deliberately included to accurately capture the main characteristics[5] of each one of the 3 positional categories, namely: <u>Attackers</u>: dribbles, goals; <u>Midfielders</u>: aerial duels, and passes; <u>Defenders</u>: interceptions and shots blocked. The dominant positional category are the attackers, with a presence of 62.03% in the sample, with the remaining 191 players falling

[5] The main characteristics associated with each one of the positional categories, are not mutually exclusive, and some player may present strong statistics in characteristics that are not usually associated with its main positional category.

into the defenders and midfielders category. In respect to the geographical distribution of transfers, over 90% of them occurred between European clubs, predominantly during the SW transfer window. Which can be attributed to the extended absence of club games during this period, resulting in more time and focus for the management team and directors to plan the next season. Deciding whether there is a need to reinforce the team, or if they need additional funds, being more prone to sell players. In terms of the players that have participated at least in one game for their respective National Serie A team over the past 2 years, the sample is rather balanced, with 55.27% who did not represent their national team. For a comprehensive overview of all variables' main statistics, please refer to the detailed information provided in the appendix (Table A.III).

Although these variables seem at first glance good candidates to proceed to the next phase of model estimation and were carefully chosen, multicollinearity problem can be present in the data. Although it will not affect the model consistency, it does affect the estimates accuracy. Therefore, the Variance inflation factor (VIF) is calculated, using $VIF_i = \frac{1}{1-R_i^2}$. The VIF results (presented in detail in appendix – table A.IV) are in all cases smaller than 10, suggesting that no multicollinearity was detected.

5. Empirical application

5.1. Regression results and model selection

Following the existent literature on the topic being discussed, the estimation starting point was the log-linear model by OLS with robust standard errors, accounting for the possible presence of heteroskedasticity. The specification of the final model has been determined by a general-to-specific approach, where the non-significant variables, considering a maximum confidence interval of 10%, were eliminated consecutively until the final specification. These results are presented on Table II.

The Ramsey RESET specification test results, presents empirical evidence that both the unrestricted and restricted models are well specified, with the non-rejection of the null hypothesis at 5% confidence interval. Suggesting that the log-linear model can be a suitable candidate for this type of data, and it presents itself as a useful solution. The final specification of the model aligns with previous studies on this topic, as presented in the appendix (table A.V).

Resorting now to the use of the non-linear model, the Poisson, the same general-to-specific approach is used. The Ramsey RESET specification test results, in both unrestricted and restricted forms, shows empirical evidence that both models are well specified, since the null hypothesis of a corrected specified model is not rejected at 5% confidence level.

Despite the fact that RESET tests for both unrestricted and restricted version of the two models, present evidence of a correct functional form, the variables that were eliminated from the analysis are jointly non-significant – see appendix (table A.VI). Thus, both restricted versions of the log-linear and Poisson models are preferred, over the unrestricted ones.

When compared to the results of the linear model final specification, the restricted Poisson model present as expected different results – as shown in table II. This occurs because both the model specification and the estimation method is different from that of the simple OLS estimation. The domestic league classification that was only significant at 10% confidence level in the log-linear model, turned out to have a stronger significance, at 1% confidence level in Poisson. Moreover, some of the covariates turned out to be non-significant in explaining the dependent variable, such as the number of assists and the number of domestic league titles in the Poisson model.

Although the log-linear model estimated by OLS has been frequently utilized in prior investigations, inference on the original scale requires additional transformations, which can be both computationally expensive and time-consuming. Since the interest lies in selecting a model for prediction in the original scale, the use of the Poisson model is preferred over the log-linear, in this specific case.

So, the selected model is the restricted PQML, which will be used for a prediction exercise with real data, presented in the next section. The transfer fee equation will subsequently adopt a form of exponential conditional mean regression, in accordance with the final specification of the selected model as presented in table II, such as:

$$\widehat{TF}_i^{\#} = e^{x_i'\hat{\beta}} \tag{15}$$

$$\begin{aligned} x_i'\hat{\beta} = 5.281 - 0.115\, Age - 0.024\, Class\_L + 0.775\, Goals + 3.163\, Pass + 1.578\, Min \\ + 0.597\, Min\_UCL - 0.362\, Pos\_MD + 0.529\, Nat\_Team \\ + 0.238\, Titles\_UCL + 0.599\, Titles\_Nat + 0.415\, Europe \end{aligned}$$

5.2. SELECTED MODEL INTERPRETATION

Before utilizing the model in the previously mentioned prediction exercise, it is crucial to firstly derive relevant conclusions and interpret the regression results. In terms of individual significance, the variables considered in the selected model, are significant at 1% confidence level, apart from *Europe and Pos_MD*, which are significant at only 5%, as presented in Table II. Joint significance is also observed at the 1% significance level.

The duration of a football player's career tends to be relatively brief, as most players do not continue playing beyond the age of 40. As players age, their probability of retirement escalates as a result of a reduction in both their physical and cognitive capacities, required to perform at the utmost level. Consequently, as players' age increase, the value of their transfer fees tends to decrease, which aligns with negative impact of the *Age* variable.

Historically, when it comes to positional categories, clubs have consistently allocated the highest sums of money towards attackers. This trend is evident when we analyse notable examples from the past, including Neymar, Kylian Mbappé, João Félix, and Cristiano Ronaldo, where substantial investments have been made on attackers. The aforementioned group of players are those who exhibit a high proficiency in goal-scoring, thereby augmenting the likelihood of triumph in matches. This puts midfielders and defenders at a disadvantage, when it comes to their associated transfer fees as well illustrate by the estimated results for the dummy positional variable (*Pos_MD*).

Regarding the negative impact of the buying club classification (*Class_L*), it is linked to the relationship between the club's sporting performance and associated revenue. When football clubs attain higher classifications (with the highest being 1st place), they tend to draw more attention from supporters, investors, sponsors, and the media. Consequently, this results in a rise in the available cashflow, which can be utilized to acquire players of higher value and uphold a competitive and successful team. Conversely, as indicated by the model, when clubs descend further down the rankings, it is expected to have a negative effect on the transfer fee.

Additional goals scored by a player per 90 minutes (*Goals)*, as anticipated, give rise to a positive transfer fee impact in comparison to the average player. Clubs highly value a player who possesses the skill to score numerous goals, as the higher the number of goals a team scores, the greater the likelihood of achieving victory.

The rate of successful passes (*Pass*) has a positive impact as expected in the transfer fee. This association can be attributed to a player's ability to minimize the margin of error in their actions. In other words, a higher rate of successful passes reduces the probability of a player losing the ball possession to the opposing team.

The percentage of minutes played both in the domestic league (*Min*) and in the UCL (*Min_UCL*), have a positive impact on the transfer fee. This result was already anticipated, as the more minutes a player participates in a game, lead to enhanced visibility within the transfer market domain and gradual acquisition of expertise. Consequently, this leads to an increase in their transfer fee.

The impact on the transfer fee is also positive when it comes to the number of titles attained in the UCL (*Titles_UCL*) and on the national senior A team (*Titles_NAT*). The most prominent stages of football undoubtedly encompass the UCL (in the context of club-level competition), as well as the national senior A team tournaments (such as the World Cup, European Championship, and Nations League). Players who have managed to win such competitions are regarded as top-performance players with immense potential, as they were able demonstrated exceptional performance compared to all participants, in a competition where only the most skilled players are present. Finally, if the transfer take place between European clubs, the effect on the transfer fee is positive. Considering the fact that the top-tier leagues are all situated within Europe, it was only expected this result. Furthermore, this outcome is also consistently supported by historical evidence, as all of the most significant transfers in history have consistently taken place between European clubs.

TABLE II

REGRESSION RESULTS: OLS LOG-LINEAR, NEGATIVE BINOMIAL 2 AND POISSON

	Log-linear				***Poisson***			
	Unrestricted		*Restricted*		*Unrestricted*		*Restricted*	
Explanatory variables	*Param. estimate*	*Robust t-stat*	*Param. estimate*	*Robust t-stat*	*Param. estimate*	*Robust t-stat*	*Param. estimate*	*Robust t-stat*
Age	*-0.118******	*-8.12*	*-0.117******	*-8.27*	*-0.118******	*-7.70*	*-0.115******	*-8.18*
Class_L	*-0.015**	*-1.68*	*-0.015**	*-1.72*	*-0.027******	*-2.82*	*-0.024******	*-2.73*
Goals	*0.719******	*2.85*	*0.866******	*3.59*	*0.666******	*2.78*	*0.775******	*3.56*
Dribble	*-0.209*	*-0.68*	-	-	*0.246*	*0.72*	-	-
Assist	*0.790****	*1.97*	*0.848******	*2.18*	*0.285*	*0.64*	-	-
Pass	*3.937******	*4.87*	*3.256******	*4.54*	*3.276******	*3.68*	*3.163******	*4.20*
Min	*1.422******	*5.32*	*1.337******	*5.19*	*1.551******	*5.02*	*1.578******	*5.53*
Min_UCL	*0.630******	*3.75*	*0.604******	*3.69*	*0.596******	*4.48*	*0.597******	*4.67*
Cards_R	*-0.120*	*-1.39*	-	-	*-0.085*	*-1.03*	-	-
Duels_A	*0.064*	*1.37*	-	-	*0.029*	*0.59*	-	-
Shots_BL	*-0.200*	*-0.83*	-	-	*0.004*	*0.02*	-	-
Pos_MD	*-0.267**	*-1.65*	*-0.387******	*-2.74*	*-0.305**	*-1.76*	*-0.362****	*-2.47*
Summer	*0.008*	*0.05*	-	-	*-0.222*	*-1.49*	-	-
Nat_Team	*0.660******	*6.29*	*0.675******	*6.47*	*0.527******	*4.59*	*0.529******	*4.67*
Titles_DOM	*0.133**	*1.74*	*0.15****	*1.98*	*-0.019*	*-0.28*	-	-
Titles_UCL	*0.247******	*2.72*	*0.257******	*2.87*	*0.242******	*4.76*	*0.238******	*5.77*
Titles_NAT	*0.740****	*5.37*	*0.733******	*5.55*	*0.597******	*5.79*	*0.599******	*5.62*
Europe	*0.395****	*2.28*	*0.396******	*2.32*	*0.484******	*2.71*	*0.415****	*2.36*
Interc	*-0.149*	*-1.30*	-	-	*-0.13*	*-1.06*	-	-
Constant term	*4.307******	*5.50*	*4.658******	*6.57*	*5.356******	*6.80*	*5.281******	*8.25*
Number of obs.	*503*		*503*		*503*		*503*	
Regressions' R^2	*0.149*		*0.129*		-		-	
Regressions' pseudo-R^2	-		-		*0.483*		*0.473*	
RESET type test								
Test statistic	*1.19*		*1.73*		*0.03*		*0.01*	
p-value	*0.3066*		*0.1791*		*0.9855*		*0.9969*	

6. PLAYER VALUATION DISCREPANCIES: MARKET VALUATION *VS.* MODEL PREDICTIONS

The topics covered in the preceding sections were primarily based on the general econometric theory and previous studies conducted using a similar methodology, apart from alternative modelling techniques, as the presented PQML model, which has never been used in this area. In order to fully exploit the theoretical findings, I shall employ the regression results (equation 15), to new data pertaining to the ongoing football season of 2023-24.

The main objective of this prediction exercise is through the use of the selected model, predict the estimated transfer fee, for a specific set of players. A subsequent examination will then be conducted, to compare the predicted results with the actual transfer fee, analyzing whether the players are undervalued or overvalued. For that purpose, the target leagues will be the ones previously considered in the initial sample. The data collection process was performed in the exact same way as the one used for gathering the initial database. In total, considering the six leagues, were collected 23 random observations regarding the transfer market SW of the 2023-24 football season.

The predicted results are presented in Table III, show evidence that, from the 23 transfers 10 of them were overpriced, taking into consideration the characteristics of the specific player being transferred. Liga Portugal presents one of the most significant overvaluation, with a disparity of 5.4 million euros between the estimated and actual transfer fee. Similarly, the Premier League follows suit, with a difference of approximately 5.5 million euros. Upon further examination of these two transfers, one can deduce that these outcomes are far from random. Firstly, the players involved, Viktor Gyökeres (Sporting) and Chermiti (Everton), both belong to the attacking sector, and possess a relatively youthful age. Secondly, upon analysing the results of the 2022-23 season, Sporting missed the chance to participate in the UCL. Furthermore, Everton, finished in the seventeenth position, only managed to maintain a solitary position above relegation to a lower tier league. These factors, may have influenced the clubs to actively reinforce their teams with players that allow to boost the clubs' results, thereby ensuring a more favourable classification outcome. Selling clubs are aware of these factors and will negotiate the players at the highest possible value, given the current situation of the buying clubs.

In respect to the remaining 13 observation, they all turned out to be undervalued. Being able to pay less for a specific player than its estimated value, can be translated into a strong negotiation capacity of the buyer. It can also be perceived, as an effective utilization of advantageous market conditions. LaLiga seems to have achieved significant success in its negotiations, showcasing the highest level of transfer fee undervaluation at 7.5 million euros. This transfer involves the player Djibril Snow, and the move from Frankfurt (Bundesliga) to Sevilla (LaLiga). This particular player was an ideal prospect for Sevilla, considering the deterioration in the player's relation with the former club. Snow had merely one year on his contract before expiration, and Frankfurt did not desire to renew it. So, Frankfurt in its willingness to let the player go, sought to generate profit prior to the expiration of the contractual agreement. Sevilla took advantage of the situation for a beneficial transaction and effectively obtained the player at a price lower than the estimated value.

Another interesting outcome arose from the transfer fee paid by Marseille (Ligue 1) to Sheffield (Premier League) for the acquisition of Iliman Ndiaye. The disparity amounted to 7.1 million euros of undervaluation, and it is easily to understand how Marseille successfully disbursed only 17 million euros. In the 2022-23 season, Sheffield successfully ascended to the Premier League subsequent to securing the runner-up position in the subordinate England league. The elevation to a more fiercely contested and demanding league, demands a certain level of financial commitment from football clubs, thereby rendering the availability of cashflow indispensable. One possible strategy is to engage in player sales in order to enhance cash-flow, as exemplified by the case of Sheffield. However, the drawback associated with such actions in these particular circumstances is that clubs like Sheffield have limited bargaining power, thereby enabling larger clubs like Marseille to secure lower transfer fees through negotiations.

In addition to the six leagues previously referenced and examined, I have made the decision to incorporate a supplementary one, namely the Saudi Pro League (SPL). Considering the absence of this particular league from the initial sample, the results presented herein, despite interesting, should be regarded carefully. The SPL has, in recent years, showcased exorbitant sums of money being exchanged for player transfers, particularly after the move of Cristiano Ronaldo to the Saudi Club Al-Nassr. The underlying objective behind this phenomenon appears to be the amplification of global focus on the league, where the significance of players' athletic performance seems to be overshadowed by their international recognition. All the six transfers considered for the SPL, turned out, as expected, overvalued, with a maximum difference between the estimated and actual transfer fee of 26.3 million euros, in respect to transfer of Neymar from PSG (Ligue 1) to Al-Hilal (Saudi Pro League). The difference between these outcomes and those from the remaining six leagues is substantial, suggesting that the SPL is trying at all costs to become future football reference globally in the future, by recruiting as much as possible the best European players.

TABLE III

PREDICTED TRANSFER FEE VALUES AND OUTCOME, BASED ON THE PQML ESTIMATED MODEL

League	***Player's name***	***Selling club***	***Buying club***	***Real fee***	***Estimated fee***	***Difference***	***Outcome***
Bundesliga	*Jessic Ngankam*	*Hertha BSC*	*E. Frankfurt*	*4 000 000*	*2 726 436*	*1 273 564*	*Overvalued*
Bundesliga	*Lucas Tousart*	*Hertha BSC*	*Union Berlin*	*2 800 000*	*3 629 750*	*-829 750*	*Undervalued*
LaLiga	*Alexander Sørloth*	*RB Leipzig*	*Villarreal*	*10 000 000*	*8 866 728*	*1 133 272*	*Overvalued*
LaLiga	*Arsen Zakharyan*	*Dyn. Moscow*	*Real Sociedad*	*13 000 000*	*17 700 000*	*-4 700 000*	*Undervalued*
LaLiga	*Carles Pérez*	*AS Roma*	*Celta de Vigo*	*5 200 000*	*6 039 244*	*-839 244*	*Undervalued*
LaLiga	*Djibril Sow*	*E. Frankfurt*	*Sevilla*	*10 000 000*	*17 500 000*	*-7 500 000*	*Undervalued*
LaLiga	*Oriol Romeu*	*Girona*	*Barcelona*	*3 400 000*	*5 747 627*	*-2 347 627*	*Undervalued*
LaLiga	*Raúl García*	*Real Betis*	*Osasuna*	*6 500 000*	*9 919 829*	*-3 419 829*	*Undervalued*
Liga Portugal	*Fran Navarro*	*Gil Vicente*	*FC Porto*	*7 000 000*	*11 900 000*	*-4 900 000*	*Undervalued*
Liga Portugal	*Francisco Moura*	*SC Braga*	*Famalicão*	*1 000 000*	*5 462 990*	*-4 462 990*	*Undervalued*
Liga Portugal	*Ricardo Mangas*	*Boavista*	*V. Guimarães*	*1 000 000*	*4 474 430*	*-3 474 430*	*Undervalued*
Liga Portugal	*Viktor Gyökeres*	*Coventry*	*Sporting*	*20 000 000*	*14 600 000*	*5 400 000*	*Overvalued*
Ligue 1	*Clinton Mata*	*Club Brugge*	*Lyon*	*5 000 000*	*3 612 888*	*1 387 112*	*Overvalued*
Ligue 1	*Iliman Ndiaye*	*Sheffield U.*	*Marseille*	*17 000 000*	*24 100 000*	*-7 100 000*	*Undervalued*
Ligue 1	*Terem Moffi*	*FC Lorient*	*Nice*	*22 500 000*	*19 700 000*	*2 800 000*	*Overvalued*
Ligue 1	*Wilfried Singo*	*Torino FC*	*Monaco*	*10 000 000*	*9 456 098*	*543 902*	*Overvalued*
Premier League	*Calvin Bassey*	*Ajax*	*Fulham*	*22 500 000*	*21 900 000*	*600 000*	*Overvalued*
Premier League	*Chermiti*	*Sporting*	*Everton*	*12 500 000*	*6 968 019*	*5 531 981*	*Overvalued*
Premier League	*Mateo Kovacic*	*Chelsea*	*Man. City*	*29 100 000*	*24 900 000*	*4 200 000*	*Overvalued*
Saudi Pro League	*Alex Telles*	*Man. United*	*Al-Nassr*	*4 600 000*	*3 536 409*	*1 063 591*	*Overvalued*
Saudi Pro League	*Habib Diallo*	*Strasbourg*	*Al-Shabab*	*18 000 000*	*12 300 000*	*5 700 000*	*Overvalued*
Saudi Pro League	*Jordan Henderson*	*Liverpool*	*Al-Ettifaq*	*14 000 000*	*3 662 697*	*10 337 303*	*Overvalued*
Saudi Pro League	*Jota*	*Celtic*	*Al-Ittihad*	*29 100 000*	*10 900 000*	*18 200 000*	*Overvalued*
Saudi Pro League	*Marcelo Brozovic*	*Inter Milan*	*Al-Nassr*	*18 000 000*	*6 024 308*	*11 975 692*	*Overvalued*
Saudi Pro League	*Neymar*	*PSG*	*Al-Hilal*	*90 000 000*	*63 700 000*	*26 300 000*	*Overvalued*
Serie A	*Arkadiusz Milik*	*Marseille*	*Juventus*	*6 300 000*	*10 600 000*	*-4 300 000*	*Undervalued*
Serie A	*Daniel Boloca*	*Frosinone*	*Sassuolo*	*10 000 000*	*7 477 023*	*2 522 977*	*Overvalued*
Serie A	*Gustav Isaksen*	*Midtjylland*	*Lazio*	*12 000 000*	*15 500 000*	*-3 500 000*	*Undervalued*
Serie A	*Yann Bisseck*	*Aarhus GF*	*Inter Milan*	*7 000 000*	*12 500 000*	*-5 500 000*	*Undervalued*

7. Conclusion

Several studies utilize linear models to uncover the drivers of professional football transfer fees. However, my findings demonstrate that a non-linear model, such as the Poisson, offers a viable alternative in a hedonic framework, when the goal is to make prediction in the original scale.

With the regression results of the selected model, some conclusions were possible to be drawn and be compared with the existent literature. The results are aligned with previous studies on the topic, and are the following:

i) As a midfielder or defender, a player's transfer fee is expected to be lower than that of an attacker. The same effect occurs when the player's age and the buying club's league classification increase. From the perspective of a buying club, if the planned budget to spent on a new player is conservative, the club should look for older players from clubs with higher classifications, and preferably not attackers;

ii) Players who score a greater number of goals per 90 minutes are expected to witness an increase in their transfer fee compared to the average player. As attackers typically score the most goals, this positive impact on the number of goals aligns with the negative impact on other positional categories;

iii) Additional minutes played (both in the domestic league and UCL), along with the additional titles won (UCL and in the National Senior A team), have also a positive impact in the transfer fee. Which can be associated with the valuation of experience by the clubs. The more experienced a player is both in time played and titles, the more accurate and better the players' actions will in principle be, with an higher mental capability in key moments;

iv) Players that are transferred between European clubs are expected to have a higher transfer fee compared to the ones who are not, which is explained by the higher concentration of the top tier leagues and clubs in Europe;

v) Representing the national senior A team confers upon the player a higher level of distinction among their peers, given the exclusive access to join a national team. Influencing positively the players’ transfer fee, given that only the best player are called upon for representing their home countries;

vi) Finally, an increased rate of successful passes also exerts a favorable impact on the transfer fee. Given that an enhanced accuracy in a player's actions, such as successful passes, contributes to the overall performance of the team, which increases the likelihood of achieving victories in the majority of the games.

. Football players are the biggest assets of each club, so it is important that whenever a club buys a player, that decision is made in a cautious and thoughtful manner, in order to maximize both the future profits from the players performance, and future valuation in market value. Additionally, the money spent on a player should not in principle be higher than its actual value. The quantitative comparison between the market valuation and model predictions, suggests the existence of some clubs that managed to make suberb deals. These clubs paid less than the estimated tranfer fee of the player, due to some almost perfect market conditions, and, possibly, good research in identifying the best oppurtunities. On the other hand, some clubs agreed to pay more for a player than its estimated value, which is some cases, can be due to the team future perspectives to increase their results considering the preceding season, or, in other cases, it can be linked to the club weak negotial capabilities.

Naturally, apart from the measurable characteristics of both the players and the clubs, factors like speculation about future performances, or club profit from a marketing perspective can also influence the amount that a club is willing to pay for acquiring a specific player. Nonetheless, some of these factors are not measurable, and others do not have at the present moment mechanisms that allowed to extract that type of information.

This dissertation provides a different methodology to deal with the application of hedonic regression models to football transfers, and valuable information for future studies. First, it shows that log-linear models can easily be substituted by non-linear ones. Second, the results obtained from the models, can be used to access different points of view regarding football: the determinants of transfers were uncovered, and the selected model was used to compare predicted and observed transfers.

A suggestion for future research on this topic, would be to try to collect variables connected with the mediatic exposure of the players, that are not included in this dissertation, given the reasons previously explained. Another interesting avenue for research, would be to obtain results for different leagues individually, and compare between each league what are the drivers of the transfer fee, and what each one values the most. Finally, with the rise of the SPL on the football world, it would be valuable for future research to use it to analyze the impact on the players valuation and migration from European leagues to Saudi Arabia.

References

Arvin, B. Mak, & Scigliano, M. (2004). Hedonic prices in art and returns to art investment: Evidence from the Group of Seven at auction. Économie Appliquée 57, 137–162

Bernardo, G., Ruberti, M. & Verona, R. (2021). Image is everything! Professional football players' visibility and wages: evidence from the Italian Serie A. Applied Economics, 54(5), 595–614

Bryson, A., Frick, B. & Simmons, R. (2013). The Returns to Scarce Talent. Journal of Sports Economics, 14(6), 606–628

Cameron, A. C. & Trivedi, P. K. (2005). Microeconometrics: Methods and Applications. Cambridge University Press.

Cameron, A. C. & Trivedi, P. K. (1998). Regression analysis of count data. Cambridge University Press.

Carmichael, F. J. & Thomas, D. (1993). Bargaining in the transfer market: theory and evidence. Applied Economics, 25(12), 1467–1476

Collins, T. (2018). How Football Began: A Global History of How the World's Football Codes Were Born. Routledge

Court, Andrew T. (1939), "Hedonic Price Indexes with Automotive Examples." The Dynamics of Automobile Demand, General Motors Corporation

Deloitte Football Money League (2023), Deloitte United Kingdom. https://www2.deloitte.com/uk/en/pages/sports-business-group/articles/deloitte-football-money-league.html

Dobson, S., & Gerrard, B. (1999). The Determination of Player Transfer Fees in English Professional Soccer. Journal of Sport Management, 13(4), 259–279

Dobson, S., Gerrard, B. & Howe, S. (2000). The determination of transfer fees in English nonleague football. Applied Economics, 32, 1145–52

Dobson, S. & Goddard, J. (2001). The Economics of Football

Evangelista, R., Ramalho, E. A. & Silva, J. A. E. (2020). On the use of hedonic regression models to measure the effect of energy efficiency on residential property transaction prices: Evidence for Portugal and selected data issues. Energy Economics, 86

Ezzeddine, M. (2020). Pricing football transfers: determinants, inflation, sustainability, and market impact: finance, economics, and machine learning approaches

Fernández-Avilés, G., Mínguez, R. & Montero, J. R. (2012). Geostatistical Air Pollution Indexes in Spatial Hedonic Models: The Case of Madrid, Spain. Journal of Real Estate Research, 34(2), 243–274

Fesselmeyer, E. (2018). The value of green certification in the Singapore housing market. Economic Letters, 163, 36-39

Frick, B. (2007). The Football players´ labor market: Empirical evidence from the major European Leagues. Scottish Journal of Political Economy, 54(3), 422–446

Garcia-del-Barrio, P. & Pujol, F. (2004). Pay and Performance in the Spanish Soccer League: Who Gets the Expected Monopsony Rents?

Gerrard, B. (2001). A new approach to measuring player and team quality in professional team sports. European Sport Management Quarterly, 1(3), 219–234

Gerrard, B. & Dobson, S. (2000). Testing for monopoly rents in the market for playing talent – Evidence from English professional football. Journal of Economic Studies, 27(3), 142–164

Giombi, K., Muth, M. K. & Levin, D. C. (2018). A comparative analysis of hedonic models of nutrition information and health claims on food products: An application to soup products. Journal of Food Products Marketing

Gourieroux, A. C., Monfort, A. & Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. Econometrica, 52, 681–700

Hamil, Sean & Chadwick, S. (2009). Managing football: an international perspective

Karnik, A. (2010) Valuing cricketers using hedonic price models. Journal of Sports Economics, 11(4), 456–469

Lucifora, C. & Simmons, R. (2003). Superstar Effects in Sport. Journal of Sports Economics, 4(1), 35–55

Malcolm, D. (2000). Football business and football communities in the twenty-first century. Soccer & Society, 1(3), 102–113

Malpezzi, S. (2003). Hedonic Pricing Models: A Selective and Applied Review, in Housing Economics and Public Policy: Essays in Honor of Duncan Maclennan, T. O. Sullivan and K. Gibbs (Eds.), Blackwell, 2003

Murray, B. & Murray, W. J. (1994). Football: A History of the World Game. Routledge

Nelson, J. P. (1982). Highway noise and property values: A survey of recent evidence. Journal of Transport Economics and Policy, 16(2), 117-138

Palomino, F., Renneboog, L. & Zhang, C. (2005) Stock price reactions to short-lived public information: the case of betting odds, ECGI, Working Paper No. 81

Panzone, L. A. & Simões, O. (2009). The Importance of Regional and Local Origin in the Choice of Wine: Hedonic Models of Portuguese Wines in Portugal. Journal of Wine Research, 20(1), 27–44

Poli, R., Besson, R. & Ravenel, L. (2021). Econometric Approach to Assessing the Transfer Fees and Values of Professional Football Players. Economies, 10(1), 4

Ramsey, J. B. (1969). Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. Journal of the Royal Statistical Society Series B-methodological, 31(2), 350–371

Radoman, M. (2017). Labor Market Implications of Institutional Changes in European Football: The Bosman Ruling and Its Effect on Productivity and Career Duration of Players. Journal of Sports Economics, 18(7), 651–672

Reilly, B. & Witt, R. (1995). English league transfer prices: is there a racial dimension? Applied Economics Letters, 2, 200–22

Roberts, A., Roche, N., Jones, C. & Munday, M. C. R. (2016). What is the value of a Premier League football club to a regional economy? European Sport Management Quarterly, 16(5), 575–591

Rosen, S. (1974). Hedonic prices and implicit markets: Product differentiation in pure competition. Journal of Political Economy, 82(1), 34-55

Ruijg, J. & Van Ophem, H. (2015). Determinants of football transfers. Applied Economics Letters, 22(1), 12–19

Santos Silva, J. & Tenreyro, S. (2006). “The Log of Gravity,” Review of Economics and Statistics, 2006, 88 (4), 641–658

Scholtens, B. & Peenstra, W. (2009) Scoring on the stock exchange? The effect of football matches on stock market returns: an event study. Applied Economics, 41(25), 3231–3237

Sloane, P.J. (1969), “The labour market in professional football", British Journal of Industrial Relations, Vol. 7, 181-99

Sirmans, S., Macpherson, D. & Zietz, E. (2005) The Composition of Hedonic Pricing Models, Journal of Real Estate Literature, 13:1, 1-44

Tacon, R. (2007). “Football and Social Inclusion: Evaluating Social Policy.” Managing Leisure 12 (1), 1–23

Waugh, F. V. (1929). Quality as a Determinant of Vegetable Prices: A Statistical Study of Quality Factors Influencing Vegetable Prices in the Boston Wholesale Market, Ph.D thesis, Columbia University 1929

White, H. (1980). A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity. Econometrica, 48(4), 817-838

Wooldridge, J. M. (2001). Econometric Analysis of Cross Section and Panel Data, Second edition, The MIT Press, Cambridge, Massachusetts

Wooldridge, J. M. (2013). Introductory Econometrics: A Modern Approach, Fifth edition. Cengage Learning

WEBSITES

https://brandirectory.com/rankings/global - accessed on 24th of April 2023

https://digitalhub.fifa.com/m/1cf301829f1cf996/original/ifab-laws-of-the-game-2020-21.pdf – accessed on 13th of April 2023

https://fifa.com/about-fifa/organisation - accessed on 13th of April 2023

https://fifa.com/about-fifa/associations - accessed on 13th of April 2023

https://fifamuseum.com/en/exhibitions-events/special-exhibitions/origins/ - accessed on 13th of April 2023

https://footystats.org - accessed from 10th of March to 30th of April

https://ifrs.org/content/dam/ifrs/publications/pdf-standards/english/2021/issued/part-a/ias-38-intangible-assets.pdf - accessed on 4th of May 2023

https://lg.com/global/business/about-lg-business - accessed on 25th of April 2023

https://pepsico.com/who-we-are/about-pepsico - accessed on 25th of April 2023

https://publications.fifa.com/en/vision-report-2021/the-football-landscape/ accessed on 13th of April 2023

https://www.transfermarkt.com/ - accessed from 10th of March to 16th of April

https://vodafone.com/about-vodafone/where-we-operate - accessed on 25th of April 2023

https://worldometers.info/world-population/ - accessed on 18th of April 2023

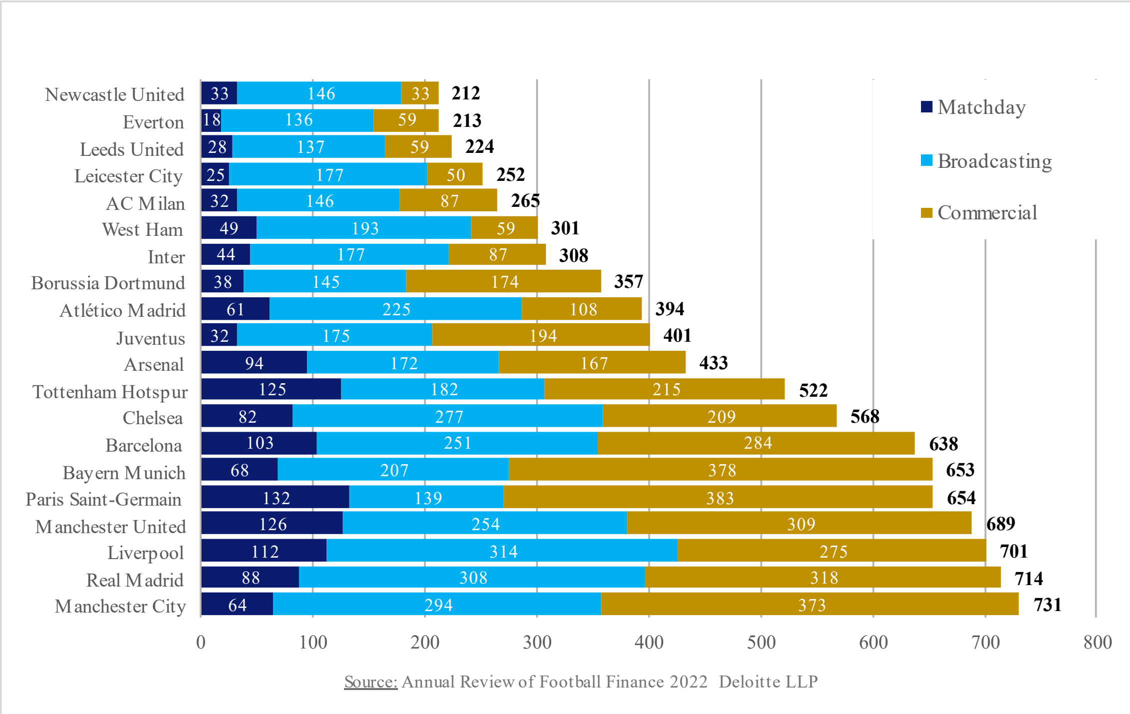


FIGURE A.1 - Football clubs with the highest revenue worldwide in 2021/22, by stream (€ millions)

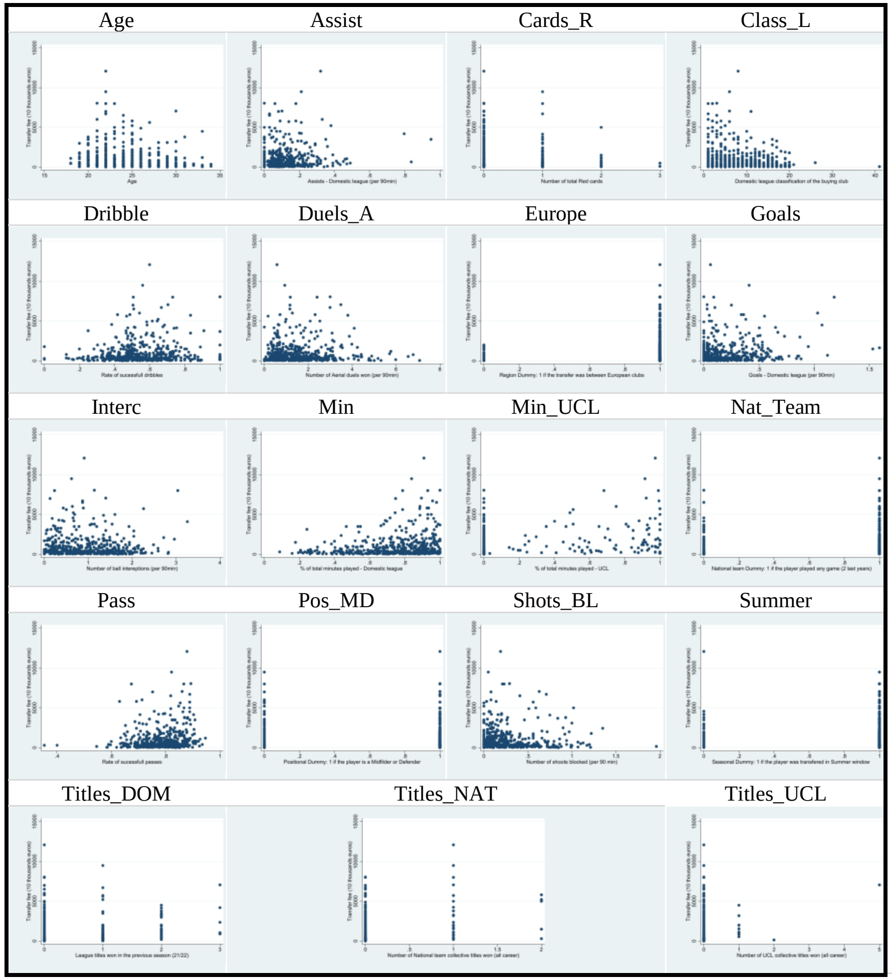


FIGURE A.2 - Explanatory variables vs dependent variable plots

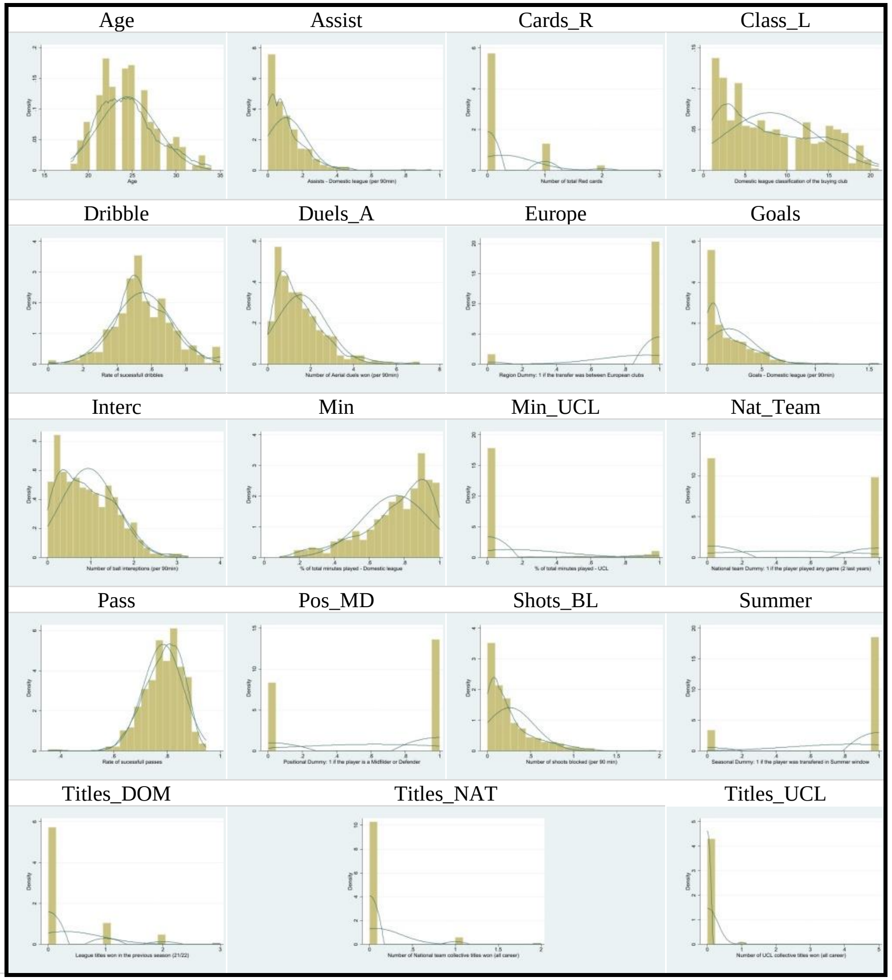


FIGURE A.3 - Explanatory variables histograms

TABLE A.I

2022-23 SEASON AND TRANSFER WINDOW DATES

	Transfer window				Football season	
	Summer		Winter			
Country	*Open*	*Close*	*Open*	*Close*	*Start*	*End*
Portugal	*Jul 01*	*Aug 31*	*Jan 03*	*Feb 02*	*Aug 05*	*May 27*
France	*Jun 10*	*Sep 01*	*Jan 01*	*Jan 31*	*Aug 05*	*May 07*
England	*Jun 10*	*Sep 01*	*Jan 01*	*Feb 01*	*Aug 05*	*May 28*
Germany	*Jul 01*	*Sep 01*	*Jan 01*	*Feb 01*	*Jul 07*	*May 27*
Spain	*Jul 01*	*Sep 01*	*Jan 02*	*Feb 01*	*Aug 12*	*Jun 01*
Italy	*Jul 01*	*Sep 01*	*Jan 03*	*Jan 31*	*Aug 14*	*Jun 04*

Source: Transfermarkt

TABLE A.II

EXPLANATORY VARIABLES DESCRIPTION

Explanatory variable	*Variable Description*
Age	Player's age at the time of the transfer
Assist	Number of assists for goals scored per 90 minutes
Cards_R	Number of red cards
Class_L	Buying club classification – 1 being the top classification
Dribble	Rate of successful dribbles, corresponding to whenever a player successfully maneuvers the ball avoiding any attempt of intercepting it
Duels_A	Number of aerial duels won per 90 minutes. An aerial duel is won whenever a player jumps with one or more opponent players to intercept an aerial ball, and is able to gain or maintaining it, into its team possession
Europe	Regional dummy variable = 1 when the transfer is made within European clubs
Goals	Number of goals scored per 90 minutes
Interc	Number of interceptions per 90 minutes. Every time a player takes possession of the ball after it has been passed or kicked by the opposing team
Min	Player's percentage of total minutes played, in the respective domestic league, taking into consideration the total number of matches
Min_UCL	Player's percentage of total minutes played, in the UCL, taking into consideration the total number of matches
Nat_Team	National team Dummy = 1 when the player represented its respective national senior A team, playing at least in one game in the past 2 years
Pass	Rate of successful passes, corresponding whenever a player successfully passes the ball to a teammate without being intercepted by any of the opponents
Pos_MD	Positional dummy = 1 when the player is a defender or a midfielder
Shots_BL	Number of shots blocked per 90 minutes
Summer	Seasonal dummy = 1 when the player is transferred in the summer window
Titles_DOM	Number of collective team titles won at club level which can include the domestic championship, other cups from each country, or European competitions (excluding UCL)
Titles_NAT	Number of national senior A team collective titles won by the player throughout its professional career
Titles_UCL	Number of UCL collective titles won by the player throughout its professional career

TABLE A.III

EXPLANATORY VARIABLES MAIN STATISTICS

Variable	***Mean***	***Median***	***Std. deviation***	***Min***	***Max***	***Skewness***	***Kurtosis***
Age	*24.44*	*24*	*3.32*	*18.00*	*34.00*	*0.55*	*2.89*
Assist	*0.11*	*0.80*	*0.12*	*0*	*0.95*	*2.42*	*13.67*
Cards_R	*0.26*	*0*	*0.53*	*0*	*3.00*	*2.12*	*7.40*
Class_L	*7.96*	*7.00*	*5.64*	*1.00*	*21.00*	*0.45*	*1.95*
Dribble	*0.55*	*0.53*	*0.17*	*0*	*1.00*	*0.16*	*3.61*
Duels_A	*1.54*	*1.28*	*1.18*	*0*	*7.08*	*1.50*	*5.87*
Europe	*0.92*	*1.00*	*0.27*	*0*	*1.00*	*-3.21*	*11.32*
Goals	*0.20*	*0.12*	*0.23*	*0*	*1.59*	*1.96*	*8.88*
Interc	*0.94*	*0.84*	*0.65*	*0*	*3.26*	*0.62*	*2.87*
Min	*0.75*	*0.79*	*0.20*	*0.09*	*1.00*	*-1.01*	*3.38*
Min_UCL	*0.14*	*0*	*0.31*	*0*	*1.00*	*1.96*	*5.18*
Nat_Team	*0.45*	*0*	*0.50*	*0*	*1.00*	*0.21*	*1.05*
Pass	*0.79*	*0.80*	*0.08*	*0.35*	*0.95*	*-0.98*	*5.80*
Pos_MD	*0.62*	*1.00*	*0.49*	*0*	*1.00*	*-0.50*	*1.25*
Shots_BL	*0.26*	*0.17*	*0.28*	*0*	*1.96*	*1.71*	*6.64*
Summer	*0.85*	*1.00*	*0.36*	*0*	*1.00*	*-1.91*	*4.63*
Titles_DOM	*0.30*	*0*	*0.64*	*0*	*3.00*	*2.12*	*6.84*
Titles_NAT	*0.07*	*0*	*0.30*	*0*	*2.00*	*4.34*	*22.77*
Titles_UCL	*0.03*	*0*	*0.27*	*0*	*5.00*	*13.52*	*226.23*

TABLE A.IV

VIF RESULTS

Variable	*VIF*	*1/VIF*
Age	*1.12*	*0.89*
Assist	*1.29*	*0.78*
Cards_R	*1.07*	*0.93*
Class_L	*1.20*	*0.83*
Dribble	*1.34*	*0.75*
Duels_A	*1.48*	*0.68*
Europe	*1.08*	*0.93*
Goals	*2.00*	*0.50*
Interc	*2.10*	*0.48*
Min	*1.27*	*0.79*
Min_UCL	*1.36*	*0.74*
Nat_Team	*1.18*	*0.85*
Pass	*1.96*	*0.51*
Pos_MD	*2.76*	*0.36*
Shots_BL	*1.77*	*0.56*
Summer	*1.16*	*0.86*
Titles_DOM	*1.42*	*0.70*
Titles_NAT	*1.17*	*0.86*
Titles_UCL	*1.16*	*0.86*
Mean VIF	*1.47*	

TABLE A.V

PRIOR STUDIES TOPICS AND MAIN FINDINGS

Author(s)	*Data*	*Estimation method*	*Dependent variable*	*Significant findings*	
				Positive	*Negative*
Carmichael and Thomas (1993)	214 transfers in the English football league - season 1990/91	OLS regression	Log of transfer fee	Average attendance of buying club (previous season), goal difference of buying club (previous season), buying club division, goal difference of selling club in previous season, selling club division, career games played, arbitrated fee (dummy)	League position of buying club in previous season squared, league position of selling club in previous season squared, player age squared
Reilly and Witt (1995)	202 transfers in the English football leagues - season 1991/92	OLS regression	Log of transfer fee	Appearances last season, goals scored in the current season, age, forward, international player, seller, and buyer clubs division	Number of previous clubs
Dobson, Gerrard, and Howe (2000)	114 transfers in semi-professional English football - 1988–1997	OLS regression	Log of transfer fee	Age, goals (previous season), average attendance of selling club (previous season), number of seats in buying club's stadium, average attendance of buying club (previous season)	Age squared, league position of selling club (previous season), goal difference of selling club (previous season), stadium capacity of buying club
Lucifora and Simmons (2003)	533 players appearing in Serie A or Serie B (Italy) - season 1995/96	OLS regression	Log of gross salary, net of bonuses and signing-on fees	Age, games in Serie A and/or Serie B (previous season), prior career games in Serie A and/or Serie B, goal rate, international caps, 'superstar status' (measured by deviation from goal rate)	-
Garcia-del-Barrio and Pujol (2005, 2006)	369 players appearing in the Primera Division (Spain) - season 2001/02	OLS regression	Log of market value	Google hits, aggregate performance index, international caps, European cup matches, European player, midfielder, forward	-
Poli, Raffaele, Roger Besson, and Loïc Ravenel. 2022	2045 transfers in the Big-Five leagues - between July 2012 and November 2021	Multiple linear regression	Transfer fee	Remaining contract duration, experience as: Goalkeeper, Center-back, Fullback, Midfielder, Offensive; goals, assists, passes, dribbles, National team participation since the start of the career, economic level of the buying club	Age

TABLE A.VI

JOINT SIGNIFICANCE TEST RESULTS

	Log-linear	***Poisson***
Variables in test	*Assist*	*Cards_R*
	Cards_R	*Dribble*
	Dribble	*Duels_A*
	Duels_A	*Interc*
	Interc	*Shots_BL*
	Shots_BL	*Summer*
	Summer	
	Titles_DOM	
Statistics	$F(6,483) = 1.02$	$\chi^2_{(8)} = 6.11$
	p-value = 0.4103	p-value = 0.6348