

Please cite the Published Version

Ezzeddine, Moussa, Pradier, Pierre-Charles and Scelles, Nicolas  (2025) Pricing football transfers using video gaming data. Journal of Sports Analytics, 11. ISSN 2215-020X

DOI: <https://doi.org/10.1177/22150218251390384>

Publisher: SAGE Publications

Version: Published Version

Downloaded from: <https://e-space.mmu.ac.uk/642238/>

Usage rights:  [Creative Commons: Attribution-Noncommercial 4.0](https://creativecommons.org/licenses/by-nc/4.0/)

Additional Information: This is an open access article published in Journal of Sports Analytics, by SAGE Publications.

Enquiries:

If you have questions about this document, contact openresearch@mmu.ac.uk. Please include the URL of the record in e-space. If you believe that your, or a third party's rights have been compromised through this document please see our Take Down policy (available from <https://www.mmu.ac.uk/library/using-the-library/policies-and-guidelines>)

Pricing football transfers using video gaming data

Journal of Sports Analytics
Vol. 11(0): 1–19
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/22150218251390384
journals.sagepub.com/home/san

Moussa Ezzeddine¹ , Pierre-Charles Pradier¹  and Nicolas Scelles² 

Abstract

Most studies exploring the global market of association football (soccer) transfers have relied on the hedonic pricing approach, where the characteristics of the player are the key features. While this has largely been useful, there are some potential challenges since it does not represent the global market and is susceptible to selection bias. In this exploratory study, we aimed to address these two key issues by proposing a different modelling approach that relies on an original solution (i.e., video gaming data). First, we built a random global sample by starting from an existing player universe, and appended video gaming data, transfer and salary data. Second, we developed new measures of transfer prices (without selectivity). This design is superior as it addresses the challenges of non-representativeness and selection bias. Indeed, the results are homoscedastic across different segments (regions and player positions) and provide results that are robust and representative of the global market than previous studies.

Keywords

asset pricing, association football (soccer), football transfers, hedonic method, selection bias

Received: 13 June 2025; accepted: 6 October 2025

Introduction

Every summer and winter, a multitude of association football (soccer) fans, analysts, media, and researchers eagerly await the rumours and indicators of possible new signings that could surpass the preceding transfer window announcements. The Mercato has become an amusing festival for the stakeholders because it engages big names and significant amounts of money, boosting trade wars for talent acquisition among clubs. Such wars have fuelled speculation and affected the valuation of transfer fees and salaries of top football players among top football clubs.

Media coverage, including social media, has also played a crucial role in reshaping the financial football model, which has been evolving over the years, influencing top management's decisions and fanbases' reactions to certain transfers. In this context, the French giant Paris Saint-Germain spent €220 m to lure the services of FC Barcelona star Neymar da Silva Santos Júnior in the summer of 2017, the highest ever amount paid for gaining the services of a football player. An all-time high, historical record, sale forced by meeting the player's release clause has led to a panic-buying reaction by the Spanish giant FC Barcelona, triggering a spending spree of €145 m and €144 m for the purchases of Ousmane Dembele and Philippe Coutinho.

Figure 1 reveals the growing expenditure in transfer fees by the top five European men's football leagues (England, France, Germany, Italy and Spain), the so-called 'Big Five', until 2019/2020, i.e., before Covid-19. The buying club usually pays an estimate of the transferred player's market worth. Thus, the issue is to identify key elements affecting football player pricing. Football clubs are struggling financially due to various reasons, including negative demand and productivity shocks, relegations and more recently the COVID-19 pandemic, with the potential consequence being insolvencies (Scelles et al., 2018; Szymanski, 2017; Szymanski & Weimar, 2019).

Understanding the player pricing function's determinants is crucial to the financial management of football clubs. In this respect, video gaming data¹ can be very helpful. First, several advancements in sports technology

¹Sorbonne School of Economics, Université Paris I Pantheon-Sorbonne, Paris, Île-de-France, FR 75013

²Department of People and Performance, Manchester Metropolitan University, UK

Corresponding author:

Moussa Ezzeddine, Sorbonne School of Economics, Université Paris I Pantheon-Sorbonne, Paris, Île-de-France, FR 75013.

Email: moussa_ezz@hotmail.com



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and Open Access page (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

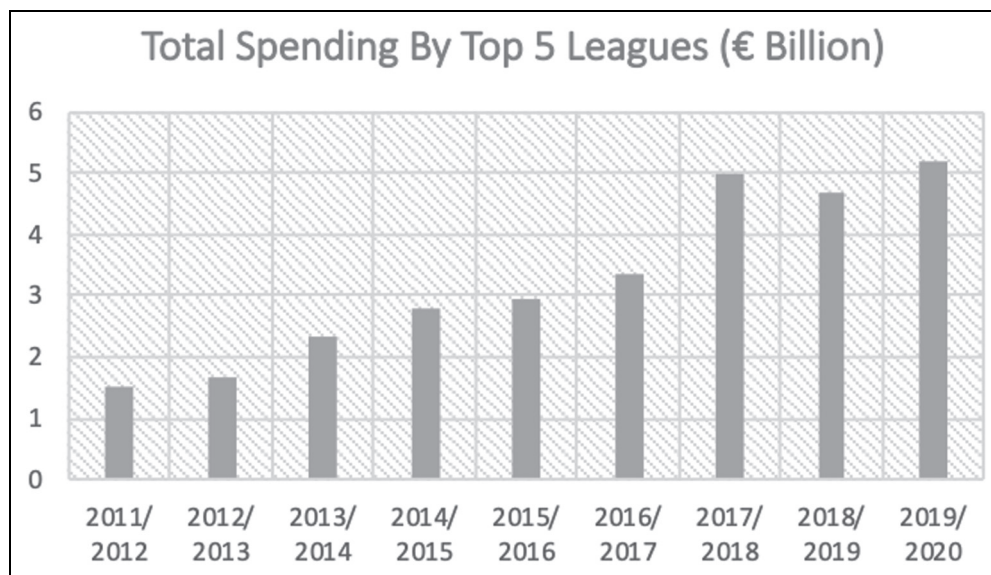


Figure 1. Growth of expenditure in transfer fees by the top five European men's football leagues. This FIFA data is used for a broader context on transfer fee growth. Source: FIFA (2019).

have made video game data more reflective of real-life player performance, thus providing a novel avenue for estimating player values more accurately than traditional methods, which often rely solely on historical data or subjective assessments. Second, with the recent trends where clubs face significant financial pressures to make informed decisions in player acquisitions, determining the extent to which transfer values can be predicted by combining video game characteristics with real-life statistics is particularly useful. Indeed, many models using the hedonic pricing of football transfers were influenced by several such studies, see, for instance, Payyappalli and Zhuang (2019). Third, pricing football transfers (while most associated with football) play a significant role in various sports and having a broader view of its predictors can thus be useful in further developing the broader economic landscape for professional athletes. Indeed, Franceschi et al. (2024) found that transfer fees were affected by the football player's age, height, goals, assists, and appearance, the timing of the transfer, and the selling and buying clubs' size and prestige. They also found that these parameters vary significantly among market segments (i.e., significant heterogeneity occurs) by geography (regions or nations), sports level (leagues), or in-game positions.

However, these studies and their approaches present two main issues. First, they relied on small samples and could not be deemed representative of the global market. Second, they were susceptible to selection bias because they valued the transfer prices from a set of transferring players. These econometric issues highlight the need for research based on larger samples representative of the global market and addressing the selectivity problem.

Most studies used independent variables like the players' characteristics, their competitive record and fame, and some

characteristics of the contract between them and the club. One may argue that these "real" performance variables are available not only for players who are transferred but also for those who are not transferred. However, their availability is not straightforward when it comes to leagues of lower standard. Therefore, using them may not allow researchers to build a sample representative of the global market. To address this issue, this paper relied on video gaming data which represents players' skills given by experts. The reason for using such type of data is that every player, transferred or not, is evaluated by experts globally and therefore we were able to include a more comprehensive subset of players in this study. The rationale behind this strategy was to contribute a solution to the lack of representativeness and selection bias problem, which has not been addressed properly in previous studies. The results show that valuation models can be consistent across time or space, while the difficulty of tackling selection bias and heteroscedasticity in a global model using the transfer fee alone was solved by aggregating various elements of players' cost in one overall package. Such aggregation generated promising results and findings. The rest of the paper is organized as follows. Section 2 reviews the literature. Section 3 presents the data, some econometric issues and the estimation strategy. This is followed by an analysis of the results (Section 5) before Section 6 provides some concluding remarks.

Literature review

The literature on football player transfer pricing is both historically deep and methodologically diverse, reflecting the growing complexity and commercial importance of player mobility in the sport. Early work, most notably by Carmichael and Thomas (1993), laid a theoretical

foundation for the subject by leveraging Nash bargaining theory to analyze how bargaining power, player attributes (such as age, appearances, and goal records), and positional factors affect transfer fees. Their research underscored that transfer pricing is fundamentally shaped by the negotiation context as well as by quantifiable features of both players and clubs.

Subsequently, Dobson and Gerrard (1999) advanced the field by introducing more nuanced market segmentation to account for geography and playing positions, enriching analytical frameworks with variables like international caps and differentiating between buyer and seller club motives. This segmentation made it clear that transfer fees arise from a combination of individual talent, strategic club behavior, and prevailing market conditions.

A key trend in the modern literature has been the incorporation of new factors reflecting the sport's commercialization. Recent studies consider media presence, player popularity, and community-driven valuations, exemplified by the use of crowdsourced data from Transfermarkt or Google search counts, as influential determinants alongside traditional performance metrics. Herm et al. (2014) is notable for analyzing the market's "branding" dimension, finding that both media visibility and online community valuations can meaningfully predict transfer prices. Barbuscak (2018) used multivariate regression models to demonstrate that variables such as contract duration, popularity, and market value, as determined by highly engaged fans, frequently hold as much relevance as purely athletic factors. These insights introduce a layer of subjectivity into the market, suggesting that the perception and media representation of a player can be nearly as valuable as their on-field performance.

The role of bargaining and negotiation has been identified as central. Speight and Thomas (1997), for example, studied how negotiated transfer fees compare with arbitration settlements, identifying systematic differences in bargaining outcomes where selling clubs often received less via arbitration, a pattern reinforcing the strategic value of negotiations in determining transfer fees. Meanwhile, Dobson et al. (2000) further validated these approaches across non-league contexts, affirming the general applicability of early theoretical models.

In terms of methodology, initial studies often relied on ordinary least squares (OLS) regression, but more recent and sophisticated research has increasingly adopted the Heckman two-step approach to correct for selection bias, a recognition that not all professional players have an equal likelihood of being transferred. Although early adoption was evident in Carmichael et al. (1999), this methodology is now more widely used, evidenced by further contributions from Depken and Globan (2021), and others. These studies first estimate the probability of a player being transferred, then use this insight to adjust the

transfer fee equation, mitigating the bias introduced when analyses focus solely on transferred players.

The importance of meticulously accounting for the characteristics of both buying and selling clubs has become a consensus in the literature, with research noting the elevated bargaining power of successful selling clubs (Carmichael and Thomas, 1993) and the positive price premium for contracts with longer duration or expiring later in time. At the same time, there is a persistent need for richer models that incorporate direct and position-specific measures of player quality. While goals and appearances are commonly included, more granular or nuanced statistics (such as those measuring defensive or creative contributions) are increasingly recognized as important for accurately capturing player value, given positional biases in traditional models.

Large-scale systematic reviews, such as Franceschi et al. (2024), have addressed data limitations and sample selection issues by reviewing dozens of studies and thousands of player transactions, revealing that the literature's empirical scope is much broader than previously characterized. Criticisms regarding limited coverage or sample size are thus being addressed as transfer market data becomes more robustly available, especially for top European leagues. Similarly, the influence of "superstar effects" and market segmentation remains an active research area, with newer studies applying advanced econometric and machine learning techniques to parse out the unique attributes of high-value transfers and to predict transfer outcomes across diverse segments of the player population.

Recent methodological advances reflect a trend towards increasingly data-driven and precise analysis. The International Centre for Sport Studies (CIES) and scholars like McHale and Holmes (2023) now leverage massive datasets from major football leagues, integrating variables ranging from granular performance data to contract specifics, team performance, and digital followership. These efforts yield models that not only predict transfer fees with greater accuracy but also unearth novel predictors absent in earlier frameworks.

Despite the progress, unresolved issues persist, especially concerning the direct quantification of non-performance factors (such as marketability, media exposure, and fan following) and how these interact with established determinants like player productivity and contractual complexity. The literature continues to grapple with documenting the full extent of transfer market segmentation and with refining theoretical explanations for observed price disparities among otherwise comparable players.

In sum, research on football player transfer pricing demonstrates a mature, evolving field, characterized by theoretical innovation, methodological rigor, and increasing empirical breadth. Contemporary studies not only confirm the importance of performance and contractual factors but also recognize the growing influence of brand value,

media presence, and digital engagement. Ongoing debates about superstar effects, segmentation, and fairness speak to the dynamic interplay of economic, social, and cultural forces shaping the modern football transfer market.

Methodology

Data

In this paper, we address the challenges associated with the non-representativeness and selection bias of previous studies that used the hedonic pricing approach to estimate football prices. To solve these puzzles, we need not start with a set of transferring players but with a somewhat *representative* set of players, to which we will later append further data such as transfer prices (if any) and salaries. The FIFA games series by EA Sports (the series having been now replaced by EA Sports FC) or Pro Evolution Soccer by Konami on PlayStation and other platforms provide a large set of players. Player information is compiled in a so-called “Futhead” database, which is available on fan pages around the internet: it contains tens of thousands of players rated by experts for their skills. They provide data for all players, *not just players who transfer*. While we acknowledge the limitations of the “Futhead” database such as expertise of raters, consistency in evaluations, update frequency, and potential biases, we selected it due to the absence of any substantive alternative. Such data have already been used in recent research, see e.g., Yan (2020) on the evaluation of performance and market value of football players in the top five European men’s football leagues, Behravan and Razavi (2021) on the estimation of football players’ value in the transfer market through machine learning, Richau et al. (2021) on the impact of investors and ownership structures on transfer fees in the English Premier League, as well as Coates and Parshakov (2022) on the determinants of the actual fees paid in a transfer of a player. The use of these data is consistent with and strengthens the more general growing relationship between sports and video games or esports, showing the mutual benefits between both (García and Murillo, 2020; Pizzo et al., 2022; Scelles et al., 2021). The data was then supplemented by transfer data between 2007/2008 and 2018/2019 and salaries between 2012/2013 and 2018/2019, all harvested on the internet, notably from *Transfermarkt* (similar to Muller et al. (2017) mentioned above or, more recently, Felipe et al. (2020) on their study of the team variables and player positions that most influence the market value of professional male footballers in the top five European men’s football leagues) and *Sofifa* (similar to other authors such as Payyappalli and Zhuang (2019) in their data-driven integer programming model for football clubs’ decision making on player transfers)².

From these transfer data, it appears that the transfer fees are not normally distributed. Even when zero transfer fees

and loans are taken out of the sample, the distribution does not appear to be normal, nor are the logarithms or any simple transformation of the transfer fees. This has an important implication since it means the Heckman correction, which rests on a normal dependent variable, could not be applied to correct for possible selection.

Most variables from the database are usual (see the complete list in Table 1), except the number of days remaining in contracts, which we could not obtain for every player. However, it has not been extensively tested so far to value transfer contracts – exceptions being articles published recently by Coates and Parshakov (2022), Garcia-del-Barrio and Pujol (2020), Richau et al. (2021) and Rubio Martin et al. (2022) – and was thus worth trying even if it reduced the number of observations greatly. Moreover, age is measured on both sides of the age of peak transfer price: *lagem* and *lagem2* measure the years below 24 years while *lagep* and *lagep2* count the years above to test for possible asymmetry of an age effect. Eventually, six skills are singled out, two for each group of positions (forward players = pace + shooting, midfielders = dribbling + passing, defenders = defending + physicality).

Then, the measurement of club transfer activity must be explained. While several global rankings of clubs do exist, they do not contain all the clubs we have in the database. We thus had to develop an endogenous measure. The idea was to count clubs especially active on the transfer market by counting the number of transfer contracts in the database: *co_cuclu* and *co_preclu* provide such a count for the current (*i.e.*, buying) and previous (*i.e.*, selling) clubs. The magnitude of transactions is recorded by *tot_cuclu* and *tot_preclu* which sum up the total value of transfers (in the database) for the current (*i.e.*, buying) and previous (*i.e.*, selling) clubs. Since the database only spans seven seasons, the hierarchy of clubs does not move much; it would be appropriate to consider those measures in a moving time window.

Eventually, we could not append the full set of information to every player in the Futhead database since only 14,051 observations of salary (Table 2) and 13,500 observations of contract duration were gathered, the intersection being around 8000. Nevertheless, the database may be considered to retain the same properties as the starting universe (*i.e.*, the Futhead database). Our sample is larger³ than all previous research and internationally diversified, so we can inquire if transfer charge pricing is worldwide or segmented. It is worth mentioning this global approach to the subject has never been tried in the literature except by Coates and Parshakov (2022). By contrast to these authors, we not only analyse our data worldwide with all player positions aggregated, but also per region as well as per position and region to identify any potential differences across these segments.

Accordingly, we built a random global sample by starting from an existing player universe and appending data gathered over the internet. Second, we developed new measures of transfer prices which are not susceptible to being censored (*i.e.*, without selectivity) and reflect the expectations of the

Table 1. Variables categories and groups.

Goals (ZGOALS1)	Goals - Previous 2 (ZGOALS2)
Goals - Previous	Goals - Previous - 2
CL Goals - Previous	CL Goals - Previous - 2
International Goals - Previous	International Goals - Previous - 2
CL Penalty Goals - Previous	CL Penalty Goals - Previous - 2
CL Qualifications Goals - Previous	CL Qualifications Goals - Previous - 2
CL Qualifications Penalty Goals - Previous	CL Qualifications Penalty Goals - Previous - 2
Local Competition Goals - Previous	Local Competition Goals - Previous - 2
Local Competition Penalty Goals - Previous	Local Competition Penalty Goals - Previous - 2
International Goals - Previous	International Goals (Previous - 2)
Assists (ZASSISTS1)	Assists - Previous 2 (ZASSISTS2)
Assists - Previous	Assists (Previous - 2)
CL Assists - Previous	CL Assists - Previous - 2
CL Qualifications Assists - Previous	CL Qualifications Assists - Previous - 2
Local Competition Assists - Previous	Local Competition Assists - Previous - 2
International Assists - Previous	International Assists (Previous - 2)
Negative Characteristics - Previous (ZBAD1)	Negative Characteristics -Previous 2 (ZBAD2)
CL Own Goals - Previous	CL Own Goals - Previous - 2
CL Red Cards - Previous	CL Red Cards - Previous - 2
CL Yellow cards - Previous	CL Yellow cards - Previous - 2
CL Yellow/Red cards - Previous	CL Yellow/Red cards - Previous - 2
CL Qualifications Own Goals - Previous	CL Qualifications Own Goals - Previous - 2
CL Qualifications Red Cards - Previous	CL Qualifications Red Cards - Previous - 2
CL Qualifications Yellow Cards - Previous	CL Qualifications Yellow Cards - Previous - 2
CL Qualifications Yellow/Red Cards -Previous	CL Qualifications Yellow/Red Cards - Previous - 2
Local Competition Own Goals - Previous	Local Competition Own Goals - Previous - 2
Local Competition Red Cards - Previous	Local Competition Red Cards - Previous - 2
Local Competition Yellow Cards - Previous	Local Competition Yellow Cards - Previous - 2
Local Competition Yellow/Red Cards - Previous	Local Competition Yellow/Red Cards - Previous - 2
Experience Previous (ZAPPI)	Experience Previous 2 (ZAPP2)
CL App (Starting 11) - Previous	CL App (Starting 11) - Previous - 2
CL App (Substituted On) - Previous	CL App (Substituted on) - Previous - 2
CL App (Substituted off) - Previous	CL App (Substituted off) - Previous - 2
CL Qualifications App (Starting 11) - Previous	CL Qualifications App (Starting 11) - Previous - 2
CL Qualifications App (Substituted on) - Previous	CL Qualifications App (Substituted on) - Previous - 2
CL Qualifications App (Substituted off) -Previous	CL Qualifications App (Substituted off) - Previous - 2
CL (minutes played) - Previous	CL (minutes played) - Previous - 2
Local Competition App. (Starting 11) -Previous	Local Competition App. (Starting 11) - Previous - 2
Local Competition App. (Substituted on) - Previous	Local Competition App. (Substituted on) - Previous - 2
Local Competition App. (Substituted off) - Previous	Local Competition App. (Substituted off) - Previous - 2
International App - Previous	International App - Previous - 2
International (minutes played) - Previous	International (minutes played) - Previous - 2

Notes: CL: Champions League, App.: Appearance. The “-2” column refers to lagged (previous two seasons) values of each skill or performance metric. The two for each group of positions” means two core skill metrics are selected for forwards, midfielders, and defenders respectively.

stakeholders through the inclusion of the current and expected cost of the player in addition to the transfer fee. The overall design is demanding in terms of information quality and quantity, but the modelling of the overall transfer cost is better than with the transfer fee alone. Thus, such a global model, including 22 countries and 25 leagues, can satisfactorily explain the expenditure by buying clubs from 2007 to 2018 on players’ transfers and wages (see Table 3 for the balance of trade in these 25 leagues). We achieved homoskedasticity on segments of the global market (regions and player positions), proven to be consistent by a series of Chow tests. This

might illustrate how the global market for transfers might be adequately studied using this methodology on an even larger scale, as more data become available over time, complementing regional markets. Appendix 2 presents our descriptive statistics.

Getting around selection issues

While most studies focus on transfer fees and, more recently, transfer market value (Franceschi et al., 2024), which is outside our scope, it might be worth refining the

Table 2. Unique salary observations in our database per country 2007–2019.

Country	# obs.	Country	# obs.
Argentina	1010	Italy	2819
Brazil	155	Mexico	1344
Chile	334	Spain	948
China	124	USA	1202
England	2212	Uruguay	165
France	1355	Other	1133
Germany	1250	Total	14051

The salary observations in Table 2 are direct estimates of annual base salary awarded to each player via club contract, gross of tax and excluding bonuses or external income. This data enables our modeling of the relationship between individual and team performance metrics and actual player compensation, providing an accurate reflection of the wage-setting processes in elite football. Transfer fees and market values, often speculative and unrelated to individual pay, are not used in this analysis.

analysis before we decide on a dependent variable to be explained by the hedonic analysis. A transfer is a bargain between three sides: a buying club paying a selling club a transfer fee and to a player some future (certain) salaries and (uncertain) bonuses (one may add the agents, but we do not have figures about their earnings and assume they perceive a percentage of the other payments). From the player's perspective, he may want to extract the maximum out of the various clubs he is going to play

with; hence, his program is to maximise the sum of his incoming cash flows:

$$\text{Max} \sum_{t=1}^{\infty} \frac{CF_t}{(1+r)^t} \quad (1)$$

where CF_t represents the incoming cash flows per period, comprising fixed salary and contingent payments (*i.e.*, various bonuses as well as sponsorship revenues), and r is the interest rate. Contingent payments are an expectation since bonuses are contingent on objectives, and future salaries beyond the contract are not known; hence, a more developed expression of this quantity for a contract lasting n period should be:

$$\text{Max} \sum_{t=1}^n \frac{w_t}{(1+r)^t} + EU \left(\sum_{t=1}^{\infty} \frac{\tilde{CF}_t}{(1+r)^t} \right) \quad (2)$$

where w_t represents the wages per period and the tilde ' $\tilde{CF}$ ' denotes a random quantity, whose expected utility (EU) might depend on the player's risk preferences.

From an objective point of view, the club is willing to minimise the cost incurred when hiring the player, and this cost breaks down into a transfer fee plus an agreed-on salary for the duration of the contract, and some additional *contingent* costs such as bonuses, which are not known on the day the contract is signed (but the list of events

Table 3. Balance of trade in the competitions studied (2007–2020).

Competition	Country	Arrivals	Expenditure (€)	Departures	Income (€)	Balance (€)
Premier League	England	5588	15.39bn	5990	8.11bn	−7276.90m
Championship	England	8247	2.03bn	9113	2.69bn	660.18m
Serie A	Italy	10952	9.67bn	11136	8.15bn	−1523.50m
Serie B	Italy	8288	472.51m	8644	1.00bn	529.37m
LaLiga	Spain	4058	7.90bn	4373	6.74bn	−1165.87m
Bundesliga	Germany	3002	5.47bn	3315	4.38bn	−1089.90m
Bundesliga 2	Germany	3059	364.13m	3360	744.64m	380.51m
Ligue 1	France	3512	5.01bn	4075	5.31bn	301.87m
Liga NOS	Portugal	5532	1.32bn	5854	3.10bn	1.78bn
Eredivisie	Netherlands	3111	937.92m	3692	2.08bn	1.15bn
Jupiler Pro League	Belgium	3824	824.77m	4157	1.32bn	499.32m
Super League	Switzerland	1943	220.96m	2143	579.39m	358.43m
Scottish Premiership	Scotland	2500	244.45m	2898	324.15m	79.70m
Premier Liga	Russia	3231	2.19bn	3324	1.48bn	−712.72m
Süper Lig	Turkey	5370	1.28bn	5567	912.50m	−363.89m
Camp. Brasileiro Série A	Brazil	9748	1.07bn	10897	2.59bn	1.52bn
Superliga	Argentina	2792	424.47m	3154	957.06m	532.60m
Chinese Super League	Chinese	2232	1.97bn	2169	558.36m	−1409.38m
Major League Soccer	USA	4147	388.28m	4286	167.55m	−220.73m
Saudi Prof. League	KSA	2045	469.44m	2242	93.07m	−376.37m
Qatar Stars League	Qatar	1391	268.32m	1535	58.50m	−209.83m
Arabian Gulf League	UAE	1383	285.98m	1484	112.79m	−173.19m
Premier Liga	Ukraine	3077	710.01m	3422	760.37m	50.37m
Super League 1	Greece	4555	363.81m	5137	405.31m	41.50m
Liga 1	Romania	4385	212.03m	4582	294.08m	82.05m

Source: Transfermarkt.

triggering bonuses may be in the contract):

$$\text{Min } TF + \sum_{t=1}^n \frac{w_t}{(1+r)^t} + \sum_{t=1}^n \frac{\widetilde{CF}_t}{(1+r)^t} \quad (3)$$

where TF represents the transfer fee.

These quantities may be approximated by (in order of greater complexity):

a. The transfer fee (TF): this has been done by the previous studies.

b. An “objective package” (OP): this is the sum of the transfer fee and annual salary during the duration of the current/new contract. This can be objectively measured it as long as we have the player’s salary and duration of the contract, as well as the transfer fee when applicable, we thus compute:

$$OP = TF + \sum_{t=1}^n \frac{w_t}{(1+r)^t} \quad (4)$$

as an approximation for the whole

$$TF + \sum_{t=1}^n \frac{w_t}{(1+r)^t} + \sum_{t=1}^n \frac{\widetilde{CF}_t}{(1+r)^t} \quad (5)$$

The advantage over the raw transfer fee is not just adding some marginal information for transferring players: taking salaries into consideration guarantees that the dependent variable is not censored. For better adequacy, the remaining contract duration should be considered. To ensure the model is meaningful, we also included in the database players on loan and players whose contracts have ended, who should thus transfer for free. Those latter players help test the consistency of estimations provided by the model, since strictly speaking, a player with six months remaining in the contract should not have a package very different from a player with a contract that just ended, albeit the distribution between the club and the player may differ significantly. Our model is not suitable to analyse this effect, though.

c. A “Subjective Complete Package” (SCP): this package is *complete* since it features all elements of costs, it is *subjective* as well since there is no objective assessment of it all:

$$\text{Total cost} = TF$$

$$+ PV \text{ of expected future income in } t$$

$$= TF + \sum_{i=1}^{+\infty} \frac{E(\tilde{w}_{t+i}|w_t)}{(1+r)^i} \quad (6)$$

where PV stands for potential value.

This package features undisclosed elements (such as the contingent payment scheme) and a double uncertainty, both on the realization of the contingent events and on what will happen beyond the horizon of the contract. We can think of all those elements to be conditional to the current salary, and it is not unreasonable to think that future salaries can be expected to vary according to the cross-sectional variation of salaries in the base. That is to say, when a player’s age grows by one unit, his salary is adjusted according to the average variation for players of his age and the probability that he remains a professional player. Is given by the average probability of players of his age. Those salaries and probabilities certainly do not evolve uniformly across the spectrum of all players (Figure 2), but this coarse approximation of evolution patterns is a starting point to compute this subjective package as:

$$SCP \approx TF + \sum_{i=\text{age}}^{42} \frac{w_{age}}{E(w_{age})} E(w_{i+1}) \times \frac{\text{number of players of age } i+1}{\text{number of players of age } i} \quad (7)$$

Where $E(w_{age})$ and *number of paid players of age i* are taken from the whole database.

Eventually, SCP can be written as:

$$SCP \approx TF + w \times \text{multiplier}(\text{age}) \quad (8)$$

where the multiplier has to be estimated from the wage distribution in the sample (Figure 2). It should be mentioned here that the income multiplier was commonly used by UEFA in the 1990s to determine the transfer prices of players between European football clubs. This price had to be at least equal to the gross salary of the player multiplied by a coefficient depending on the age of the player (art. 3, UEFA, 1992). Accountants such as Morrow (1999) and Scarpello and Theeke (1989) criticised the inconsistency of the method with standard economic theory. The main difference between our approach and UEFA-1990’s own is that our multiplier is estimated from the data.

Eventually, the player may be interested in the transfer fee as a signal of the willingness of the club to pay, but he is likely concerned only by what he will take from his club; hence we can define:

d. A Subjective Salary Package (SSP): this package assumes clubs keep the player during his whole career. If markets and information were perfect, this should match the income generated by the player, hence:

$$\sum_{t=1}^n \frac{w_t}{(1+r)^t} + \sum_{t=1}^{\infty} \frac{\widetilde{CF}_t}{(1+r)^t} \quad (9)$$

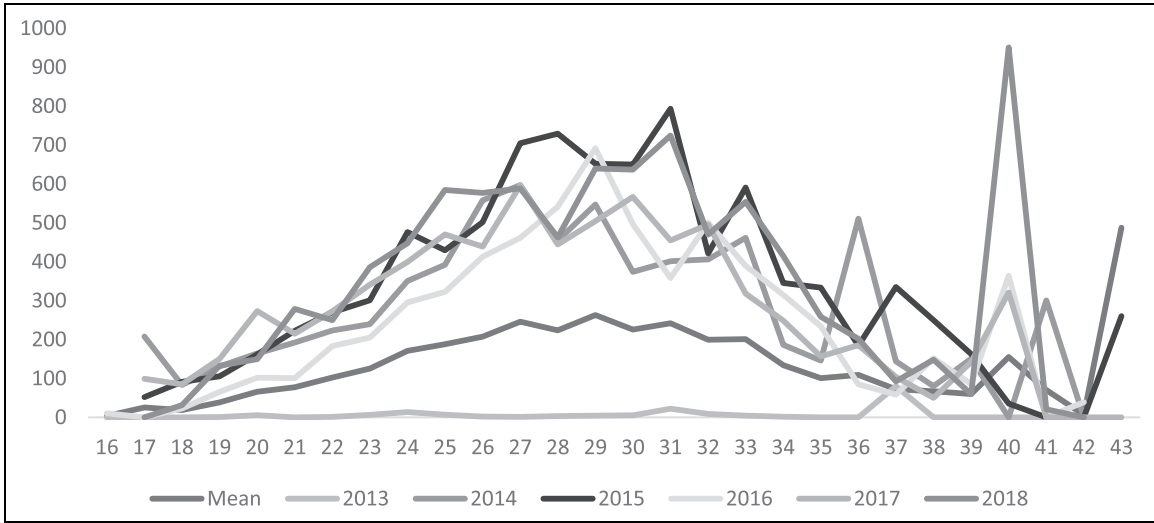


Figure 2. Average salaries per age per season.

We assumed as with the SCP that the future salaries are dependent on the current relative salary and the average evolution in the database hence:

$$SSP \approx w \times multiplier(age) \quad (10)$$

It seems pretty obvious that the complete specification of the subjective package is well beyond our current knowledge of the stakeholders: we do not have data on the risk preferences of football players, nor on the interest rate players and clubs consider discounting future opportunities. We had to make a series of assumptions to discuss the general idea that we want to capture a non-linear relationship between transfer fee, age and current salary. We assumed players to be somewhat risk-neutral and adopt a zero-discount rate.

It is quite obvious that the aforementioned packages do not provide a *direct* estimate for the transfer fee. The transfer fee can be computed very simply, though, from the packages predicted by the models, since:

$$\widehat{transfer\ fee} = \widehat{package} - salaries \quad (11)$$

Eventually, we have four dependent variables to try to value, and three of them are *not* censored. We can thus use these packages in the selection equation of a Heckman-inspired regression. Since the OP and SCP depend on the transfer fee, it might be better to look only at the SSP, as done in the next section.

Estimation strategy

Our standard hedonic model uses a log-linear equation to determine the dependent variable's value based on the

player's skills, personal traits, and control factors, *i.e.*,

$$\ln depvar = \sum_i^m \alpha_i \ln X_i + \sum_j^n \beta_j \ln Y_j + \sum_k^l \delta_k \ln Z_k + u_i \quad (12)$$

where *depvar* (dependent variable) can be either the transfer fee or any of the packages explained in the previous section and X_i is the players' skills vector, Y_j is personal characteristics vector, and Z_k is the control variables vector (country, position and time). We applied full standardization to all continuous features. This approach facilitates a meaningful comparison/interpretation of coefficients and assessment of use of video games data. We started estimating the dependent variables on the whole sample, which includes all countries and positions aggregated, then we tried disaggregated regressions per continent, per position and performed a series of Chow tests to choose between the aggregate and the multi-level model⁴. To achieve the proper segmentation, many estimations have been generated per country, per continent, per year and per position. While the whole process might look like pointless data mining, it appears *a posteriori* that the results of regressing the four dependent variables are mostly convergent; the main difference is the ability of a given specification to reduce heteroscedasticity in and across market segments. While providing all the Chow tests would be very fastidious, we only give an intuition of how segmentation works by providing some decisive examples in the results section. It is worth noting that the USA and China were grouped, despite not being in the same geographical region. This is because they share some similar characteristics, *e.g.*, they attract players who used to play in some of the best leagues but are not able to perform at the same level as they used to do as they age.

4 Results

Regressing the transfer fee (as previous studies have done) brings many significant results (Table 4): the transfer price is an increasing function of the duration of the contract, internet visibility or google hits (although not significant at the world level), player skills, buying club transfer activity, and is negatively affected by the end of the contract ('free transfer') or the transfer being a loan. There are some consistency problems when the regression is broken down by continent or position: for instance, yearly dummies tend to be significant and negative at the world level but not necessarily at the continent level (except for Latin America), while the Chow test shows that the disaggregated model is better. The Breusch-Pagan statistics ('Chi2') indicate heteroscedasticity, which disaggregation cannot reduce. While this does not make the model irrelevant, it means that the regression does not well render the granularity of the data. We find that the r-square is largest in the case of Europe relative to that of the global sample. For Latin America and USA/China it is relatively lower. There is a better fit for the case of European data.

Figure 3 is the scatterplot between actual transfer values and predicted values. The fit is quite good here, maybe not as good as Poli et al. (2017), who claimed that their models gave R^2 above 80%. This is to say, their model provides an account for more than 80% of the observed differences in the observed transfer fees (or, more precisely, the squares of the logarithms of those fees). This is a very good result for a statistical estimation procedure.

Using the packages to look at the breakdown by continent gives the same kind of results as with the transfer fee, while the Breusch-Pagan statistic is significantly lower to a point where heteroscedasticity can disappear in some instances (Table 5). The yearly dummies are consistent between the global market and the European market but not with other continents: although these dummies may be interpreted as a price index due to inflation over time (see Figure 1 for the Big Five), it seems that the price of transfers is not evolving consistently across continents; hence, breaking down the regression is required, as the Chow test shows.

Breaking down by positions is performed only with the subjective complete package (SCP): while the result is consistent with other packages, SCP almost removes heteroscedasticity. This indicates not only that the estimated coefficients are unbiased, but also that the segmentation may be relevant to the resolution of the data. Table 6 is devoted to the subset of forward players: while there is heteroscedasticity, the Chow test indicates the aggregate model is better than a breakdown by continent. Unsurprisingly, the value of forward players' transfer is linked to their specific skills such as "*shooting*", "*dribbling*" and "*passing*" rather than other skills, which are more related to other positions. In addition to those skills variables, some other factors were

significant, like some age variables. The involvement of the buying and selling clubs in the transfer business was also noticeable through appropriate variables. We again notice that the remaining duration of the contract before a transfer deal is reached is positively significant. While a Chow test does not favour disaggregation by continent, a separation between England and the rest of the world makes sense. Interestingly, England favours more "defending" and "physicality", in contrast with the world level. This is consistent with the perception that England is a highly competitive and intense league (Bradley et al., 2016).

The same feature is true for the defenders and defensive midfielders (Table 7): the best segmentation (according to Chow tests) is England vs the rest of the world. Unsurprisingly, the defending and physicality skills are valued. The English market is more visible since Google Trends has a positive impact on player value. The other variables have the same impact as the forward players. Defensive midfielders have been added to defenders as the result of another Chow test.

For (non-defensive) midfielders (Table 8), there is a higher heteroscedasticity than for other positions. Not only the English but also the Italian market as well are singled out: there seem to be some singularities both in the appreciation of the players (height is preferred, youth is an asset) and their situation (loans seem to be priced like regular transfers of shorter duration). England displays a significant negative impact of the loan variable with a coefficient higher in absolute value than the one at the world level. This result can be due to the intention of clubs to lend their players, even at little or no cost, to the clubs, to put them on temporary display to sell them more easily. Clubs outside the Premier League may also favour a loan of their players in the latter rather than in their domestic league so that they do not take the risk of strengthening their domestic competitors (Feuillet et al., 2021). The number of observations may be too small, and the interest of the regression is more in showing the difference with the rest of the world than in specifying a very precise model.

Discussion and conclusion

Main findings

The analysis of football transfer fees in this article reveals a complex interplay of factors influencing their value, highlighting the need for a nuanced understanding that goes beyond simple aggregate models. While contract duration, internet visibility, player skills, and buying club activity consistently demonstrate a positive correlation with transfer fees, the significance and relative importance of these factors exhibit considerable geographic variation. The superior fit of the model for European data compared to other regions underscores the need for regionally specific

Table 4. Regression with the transfer fee being the dependent variable.

	World	Europe	Latin America	USA and China
Dummy for players moving after finishing previous contract	−0.856***	−1.421***	−0.134***	−0.232***
Dummy for players moving on loan	−0.679***	−1.306***	−0.114***	−0.105**
Players age (standardized)	−0.131***	−0.116***	−0.025***	0.012
Player's age squared (standardized)	−0.050***	−0.058***	−0.015***	−0.072***
Duration of contract (standardized)	0.254***	0.252***	0.084***	0.072***
Player's height (standardized)	0.010	0.057**	−0.005	0.045*
Pace (standardized)	0.010	0.048*	0.009	0.046**
Dribbling (standardized)	0.035	−0.001	0.000	0.087*
Shooting (standardized)	0.126***	0.149***	0.035***	0.113***
Defending (standardized)	0.047*	0.048	0.005	0.016
Passing (standardized)	0.058***	0.069**	0.001	−0.090*
Physicality (standardized)	0.134***	0.117***	0.032***	0.107***
Player's search frequency as a google trend (standardized)	0.015	0.055***	−0.009*	0.052***
Player's search frequency on google (standardized)	0.036	0.054	−0.010	−0.003
Dummy for players playing in both feet	0.339**	0.518**	−0.041	0.169
Dummy of players playing in left foot	−0.170***	−0.189*	−0.022	−0.136
Dummy of the players playing in right foot	−0.156***	−0.204*	0.016	−0.101
Dummy of forward players (Strikers)	0.092	0.256	0.049**	−0.312**
Dummy of midfield players (Midfielders)	−0.060	0.127	−0.012	−0.287**
Dummy of defensive players (Defenders)	0.014	0.170	−0.013	−0.294**
Number of transfers done by the previous/selling club 2007/2008–2018/2019 (standardized)	−0.123***	−0.330***	0.013*	−0.007
Volume of transfers (£) done by the previous/selling club 07/08–18/19 (standardized)	−0.120***	−0.253***	−0.011*	−0.127**
Volume of transfers (£) done by the current/buying club 07/08–18/19 (standardized)	0.206***	0.261***	0.054***	0.278***
Number of transfers done by the current/buying club 2007/2008–2018/2019 (standardized)	0.328***	0.339***	0.187***	0.450***
Dummy of players of Asian origin	−0.127*	−0.094	−0.173***	−0.159*
Dummy of players of African origin	−0.047	−0.008	0.031	−0.024
Dummy of players of Australian origin	−0.181**	−0.067		−0.196*
Dummy of players of European origin	−0.154***	−0.034	−0.018	−0.083
Dummy of players of South American origin	0.082**	0.322***	−0.044**	0.242***
Yearly Dummy for transfers of the 2008/2009 Two transfers windows	−0.164	−0.145	−0.330**	
Yearly Dummy for transfers of the 2009/2010 Two transfers windows	−0.437***	−0.130	−0.428**	
Yearly Dummy for transfers of the 2010/2011 Two transfers windows	−0.191*	0.079	−0.397**	
Yearly Dummy for transfers of the 2011/2012 Two transfers windows	−0.317***	−0.184	−0.406**	0.186
Yearly Dummy for transfers of the 2012/2013 Two transfers windows	−0.258***	−0.020	−0.399**	0.187
Yearly Dummy for transfers of the 2013/2014 Two transfers windows	−0.195**	−0.054	−0.387**	0.251**
Yearly Dummy for transfers of the 2014/2015 Two transfers windows	−0.213**	−0.068	−0.371**	0.218*
Yearly Dummy for transfers of the 2015/2016 Two transfers windows	−0.194**	−0.024	−0.375**	0.185
Yearly Dummy for transfers of the 2016/2017 Two transfers windows	−0.132	0.052	−0.331**	0.409***
Yearly Dummy for transfers of the 2017/2018 Two transfers windows	−0.004	0.213	−0.312*	0.390**
Constant	−0.622***	−0.803**	−0.052	−0.094
Chi2	1753.93	361.57	1521.37	848.45
Prob>Chi2	0.000	0.000	0.000	0.000
Observations	13,977	8514	3007	1326
R-squared	0.282	0.325	0.272	0.266
Adj. R-Squared	0.280	0.311	0.243	0.222

analyses, acknowledging that the global market for football players is not monolithic.

Heteroscedasticity, a persistent challenge in the regression analysis, underscores the limitations of a purely aggregate approach. While disaggregating the data by continent partially addresses this issue, the unequal variance of

errors suggests that unobserved heterogeneity, possibly related to specific market dynamics or institutional factors, remains a significant influence on transfer fee determination. Further research should delve deeper into these contextual factors to refine the model and improve predictive accuracy.

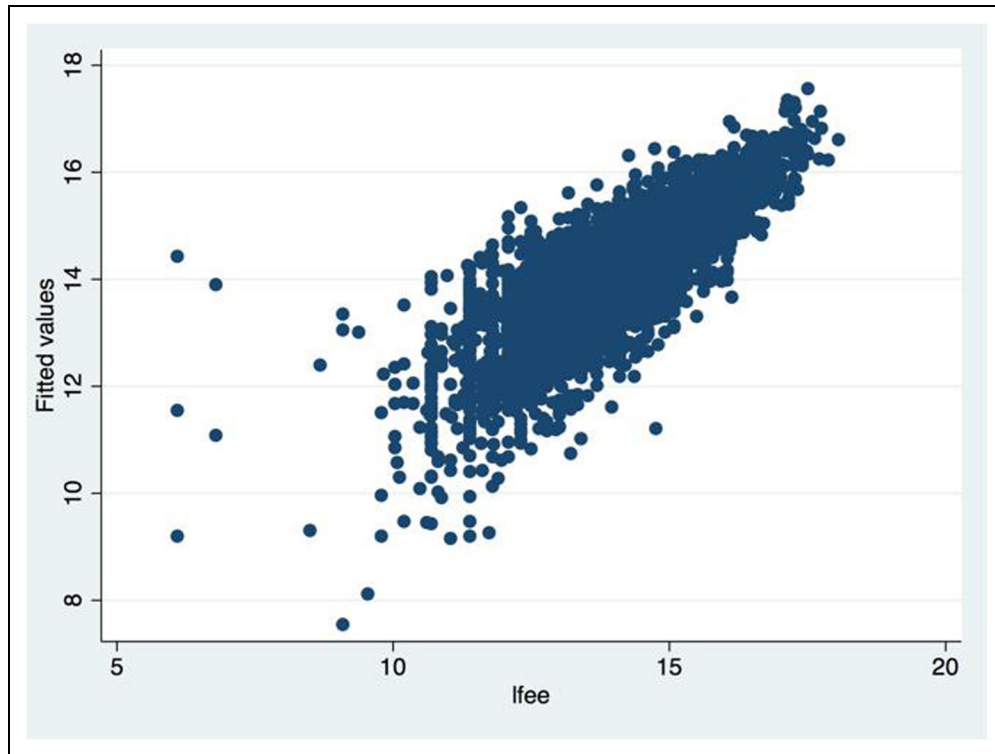


Figure 3. Scatterplot between actual transfer values and predicted values.

The incorporation of yearly dummy variables acknowledges the temporal dimension of transfer values, suggesting that inflation or other macroeconomic factors are at play. However, the inconsistent temporal patterns across continents indicate that these influences are not uniformly distributed globally. This highlights the need to account for regional economic conditions and their specific impacts on the market for football players. Analysing transfer fees by player position reveals crucial insights into the relative importance of specific skills. For forward players, shooting and dribbling abilities are particularly significant, while defensive players' values are more strongly influenced by their defensive skills. Even within positions, regional variation is evident, with England often exhibiting distinct characteristics compared to the rest of the world. This underscores the limitations of a one-size-fits-all model and the importance of regional context.

The unique characteristics of the English market in loan deals, showing a negative correlation for non-defensive midfielders, highlight potential strategic market behaviour, consistent with Feuillet et al. (2021). The practice of loaning players at little or no cost, primarily within the Premier League, suggests a strategy of showcasing players for potential future sales. This behaviour contrasts with markets outside the Premier League, implying that strategic considerations and differing risk tolerances are crucial elements of the transfer market. Future research should explore these strategic aspects of player transfers

more deeply, for example by building from Feuillet et al. (2021) and our own study.

In conclusion, this research offers valuable insights into the multifaceted determinants of football transfer fees. While broadly identifying key variables, the study emphasizes the critical role of regional context and player position, revealing inconsistencies that require region- and position-specific models to fully understand the dynamics at play. The presence of heteroscedasticity necessitates further investigation into unobserved factors and potentially more sophisticated modelling techniques to achieve a more comprehensive understanding of this complex market.

Practical (managerial/policy) implications

Evaluating players based on their performance statistics is not a new approach. Studies on the determinants of transfer fees followed a similar path while evaluating players based on subjective scores given by experts (used in video games like PlayStation games). Clubs also use this type of analysis when considering buying or selling a player, players and/or their agents could also use it to understand whether a selling club asks for a realistic transfer fee. Besides, football governing bodies (e.g., FIFA) could use it to assess whether a transfer fee is realistic or if further investigation is required if it is not realistic. It's noteworthy to mention that some missing variables (remaining duration of contracts) in previous studies happen to be influential in using video gaming

Table 5. Global model + breakdown by continent for objective package / subjective complete package / subjective salary package.

VARIABLES	World			Europe			Latin America			USA and China		
	OP	SCP	SSP	OP	SCP	SSP	OP	SCP	SSP	OP	SCP	SSP
Dummy for players moving after finishing previous contract	-0.731***	-0.605***	-0.371***	-1.103***	-0.691***	-0.313***	-0.160***	-0.253***	-0.225***	-0.210**	-0.148***	-0.015
Dummy for players moving on loan	-0.528***	-0.306***	-0.105*	-0.935***	-0.417***	-0.049	-0.120***	-0.220***	-0.200***	0.005	0.064	0.077
Players age (standardized)	-0.025	-0.995***	-1.076***	0.005	-1.114***	-1.215***	0.022*	-0.575***	-0.636***	-0.004	-0.280***	-0.325***
Player's age squared (standardized)	-0.091***	0.415***	0.484***	-0.097***	0.519***	0.601***	-0.041***	0.206***	0.237***	-0.087***	0.052*	0.091***
Duration of contract (standardized)	0.480***	0.215***	0.143***	0.477***	0.240***	0.179***	0.287***	0.007	-0.022	0.193***	0.079***	0.058**
Player's height (standardized)	-0.005	0.041	0.047*	0.030	0.092**	0.089**	-0.006	0.067**	0.081***	0.051	0.032	0.006
Pace (standardized)	-0.006	0.038*	0.041*	0.025	0.102***	0.100***	0.019	0.022	0.019	0.039*	0.052**	0.037
Dribbling (standardized)	0.054*	0.098***	0.101***	0.034	0.083	0.094*	0.032	0.089***	0.104***	0.115	-0.002	-0.051
Shooting (standardized)	0.155***	0.188***	0.166***	0.160***	0.192***	0.172***	0.054***	0.103**	0.104*	0.150***	0.068**	-0.004
Defending (standardized)	0.105***	0.116***	0.109***	0.129***	0.162***	0.155***	0.037*	0.048	0.057	0.021	0.061	0.026
Passing (standardized)	0.083***	0.255***	0.272***	0.094**	0.350***	0.369***	0.028	0.104**	0.113**	-0.072	0.040	0.133***
Physicality (standardized)	0.145***	0.169***	0.140***	0.138***	0.209***	0.188***	0.073***	0.086***	0.079***	0.033	0.029	0.007
Player's search frequency as a google trend (standardized)	0.000	-0.002	-0.008	0.037*	0.031	0.011	-0.015	-0.026	-0.024	0.039	0.048**	0.030
Player's search frequency on google (standardized)	0.035	-0.020	-0.042	0.051	-0.089	-0.134*	-0.035	-0.004	0.031	0.016	-0.008	-0.006
Dummy for players playing in both feet	0.514***	0.298	0.152	0.701***	0.299	0.096	0.240**	0.072	0.094	0.321*	0.092	-0.094
Dummy of players playing in left foot	-0.275***	-0.038	0.033	-0.342***	0.178	0.276	-0.048	-0.103	-0.117	-0.193	-0.035	0.069
Dummy of the players playing in right foot	-0.267***	-0.038	0.041	-0.381**	0.161	0.284*	0.012	-0.023	-0.041	-0.144	-0.016	0.069
Dummy of forward players (Strikers)	0.179	-0.342	-0.403	0.411	0.273	0.219	0.093	-0.543	-0.599	-0.543**	-0.041	0.259*
Dummy of midfield players (Midfielders)	-0.117	-0.740***	-0.770**	0.106	-0.302	-0.338	-0.093	-0.604	-0.643	-0.538**	-0.111	0.181
Dummy of defensive players (Defenders)	0.040	-0.348	-0.365	0.227	0.109	0.102	-0.035	-0.438	-0.464	-0.447**	-0.119	0.163
Number of transfers done by the previous/selling club 2007/2008-2018/2019 (standardized)	-0.160***	-0.117***	-0.072***	-0.340***	-0.322***	-0.232***	-0.035**	-0.003	0.004	0.022	0.037	0.007
Volume of transfers (£) done by the previous/selling club 07/08-18/19 (standardized)	-0.142***	-0.084***	-0.051*	-0.246***	-0.128**	-0.056	-0.038***	-0.065***	-0.068***	-0.095*	-0.063*	-0.004
Volume of transfers (£) done by the current/buying club 07/08-18/19 (standardized)	0.215***	0.235***	0.193***	0.251***	0.271***	0.215***	0.115***	0.100***	0.093**	0.409***	0.212***	0.129***
Number of transfers done by the current/buying club 2007/	0.350***	0.364***	0.291***	0.359***	0.373***	0.298***	0.417***	0.490***	0.459***	0.374**	0.276***	0.010

(continued)

Table 5. Continued.

VARIABLES	World			Europe			Latin America			USA and China		
	OP	SCP	SSP	OP	SCP	SSP	OP	SCP	SSP	OP	SCP	SSP
2008–2018/2019 (standardized)												
Dummy of players of Asian origin	0.005	0.044	0.080	0.031	0.085	0.118	-0.263***	-0.152	-0.102	-0.070	-0.018	0.032
Dummy of players of African origin	-0.058	0.031	0.040	-0.025	0.054	0.048	-0.076	-0.058	-0.061	-0.108	-0.078	-0.057
Dummy of players of Australian origin	-0.267**	-0.419***	-0.410***	-0.099	-0.373*	-0.429**				-0.473**	-0.187	0.047
Dummy of players of European origin	-0.145***	-0.171***	-0.125***	-0.075	-0.038	-0.020	0.020	-0.000	0.005	-0.048	0.004	0.083**
Dummy of players of South American origin	0.144***	0.213***	0.197***	0.296***	0.311***	0.248***	-0.048	0.064	0.098*	0.108	0.121*	0.023
Yearly Dummy for transfers of the 2013/2014 Two transfers windows	0.058	0.238**	0.277***	0.141	0.258**	0.269**	-0.101	0.342**	0.494***	0.365***	0.116	0.021
Yearly Dummy for transfers of the 2014/2015 Two transfers windows	-0.038	-0.076	-0.043	0.060	-0.039	-0.032	-0.221*	-0.055	0.052	0.312**	0.080	0.014
Yearly Dummy for transfers of the 2015/2016 Two transfers windows	0.016	-0.008	0.024	0.119	0.033	0.035	-0.146	0.056	0.171	0.330***	0.119	0.068
Yearly Dummy for transfers of the 2016/2017 Two transfers windows	-0.072	-0.422***	-0.468***	-0.003	-0.529***	-0.626***	-0.118	-0.000	0.090	0.372***	0.111	-0.063
Yearly Dummy for transfers of the 2017/2018 Two transfers windows	0.078	0.429***	0.472***	0.243**	0.643***	0.657***	-0.178	-0.001	0.088	0.237**	0.033	-0.075
Constant	-1.983***	-0.323	-0.093	-2.131***	-1.565***	-1.458***	-1.164***	0.966	1.085	-0.452	-0.252	-0.494
Chi2	113.01	148.47	97.18	131.57	118.66	73.87	5.65	10.58	8.73	15.66	6.83	3.9
Prob>Chi2	0.000	0.000	0.000	0.000	0.000	0.000	0.0175	0.0011	0.0031	0.0001	0.009	0.0482
Observations	7880	7880	7880	5237	5237	5237	1507	1507	1507	530	530	530
R-squared	0.353	0.368	0.353	0.353	0.353	0.353	0.353	0.353	0.353	0.353	0.439	0.353
Adjusted R-squared	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309	0.309

Robust standard errors in parentheses; ***p<0.01, **p<0.05, *p<0.1.

Table 6. Breakdown by position and continent for the subjective complete package – forward players (strikers).

VARIABLES	World	World without England	England
Dummy for players moving after finishing previous contract	−0.260***	−0.234***	−0.445
Dummy for players moving on loan	−0.130	−0.125	−0.162
Player's age is below 24 years old (standardized)	−0.107	−0.135	−0.325
Player's age is above 24 years old (standardized)	−0.000	0.001	0.005
Player's age is below 24 years old - squared (standardized)	0.308*	0.265	0.008
Player's age is above 24 years old - squared (standardized)	−0.018***	−0.017***	−0.009
Duration of contract (standardized)	0.094**	0.076*	0.044
Player's height (standardized)	0.064	0.082*	0.059
Pace (standardized)	−0.037	0.006	−0.213
Dribbling (standardized)	0.193**	0.093	0.688**
Shooting scores (standardized)	0.178*	0.279**	0.002
Defending (standardized)	0.109	0.099*	0.297
Passing (standardized)	0.132**	0.125**	0.133
Physicality (physical strength) (standardized)	0.040	0.005	0.071
Player's search frequency as a Google trend (standardized)	−0.007	−0.014	0.002
Player's search frequency on Google (standardized)	0.123	0.190	−0.004
Dummy for players playing in both feet	0.523	0.534	1.078
Dummy of players playing in left foot	−0.065	−0.085	−0.413
Dummy of the players playing in right foot	−0.058	−0.127	0.014
Remaining duration of previous contract at the time of current contract (standardized)	−0.032	−0.041	−0.493
Number of followers of players on sofifa website (standardized)	−0.000	0.006	−0.282
Number of transfers done by the current/buying club 2007/2008–2018/2019 (standardized)	0.145***	0.145***	0.172
Number of transfers done by the previous/selling club 2007/2008–2018/2019 (standardized)	0.199***	0.223***	0.240**
Volume of transfers (£) done by the current/buying club 07/08–18/19 (standardized)	0.047***	0.040***	0.105*
Remaining duration of previous contract at the time of current contract (standardized)	0.620***	0.568***	0.887***
Dummy of players of Asian origin	−0.083	0.357	−3.116**
Dummy of players of African origin	0.079	0.057	0.146
Dummy of players of Australian origin	−0.106	−0.113	0.078
Dummy of players of European origin	0.022	0.023	0.302
Dummy of players of South American origin	0.168**	0.216**	−0.312
Yearly Dummy for transfers of the 2013/2014 Two transfers windows	0.276*	0.221	0.441
Yearly Dummy for transfers of the 2014/2015 Two transfers windows	−0.008	−0.030	0.255
Yearly Dummy for transfers of the 2015/2016 Two transfers windows	0.137	0.044	0.518
Yearly Dummy for transfers of the 2016/2017 Two transfers windows	−0.611***	−0.586***	−0.577
Yearly Dummy for transfers of the 2017/2018 Two transfers windows	0.494***	0.357**	1.318**
Constant	2.596	3.031*	4.307
Observations	2921	2390	531
R-squared	0.449	0.460	0.469
Adjusted R-squared	0.431	0.431	0.431

Robust standard errors in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

data. Moreover, the loans and free transfers were present, through their dummy variables, for the first time in a study making the pricing function of this research unique and comprehensive. The controversy about the greed of agents by media houses and media, blaming them for the hike in prices of transfer fees and salaries, made it crucial to investigate the significance of the agent's dummy variables, which proved such allegations insignificant, while buying clubs' financial strength was proved significant.

Theoretical implications

The findings of this paper do have some theoretical implications. The use of video gaming data to price football transfers may allow for the incorporation of new factors that affect a player's value, such as in-game performance, popularity among fans, and overall impact on the team. This implies the reciprocal benefits of using both actual sports data and video gaming data. This can lead to a more

Table 7. Breakdown by position and continent for the subjective complete package - defenders and defensive midfielders.

	World	World without England	England
Dummy for players moving after finishing previous contract	−0.234***	−0.211***	−0.455**
Dummy for players moving on loan	−0.050	−0.064	−0.305
Player's whose age is below 24 years old (standardized)	−0.390***	−0.331***	−0.553**
Player's whose age is above 24 years old (standardized)	0.006***	0.005***	0.008
Player's whose age is below 24 years old - squared (squared)	0.062	0.077	0.140
Player's age is above 24 years old - squared (standardized)	−0.014***	−0.013***	−0.023**
Duration of contract (standardized)	0.116***	0.090***	0.094
Player's height (standardized)	0.027	0.035	0.008
Pace (standardized)	−0.056*	−0.047	−0.099
Dribbling (standardized)	0.071*	0.076*	0.132
Shooting scores (standardized)	0.125*	0.158***	0.039
Defending (standardized)	0.405***	0.315**	0.899**
Passing (standardized)	0.164***	0.103**	0.359*
Physicality (physical strength) (standardized)	0.177***	0.121**	0.449**
Player's search frequency as a Google trend (standardized)	0.011	0.001	0.038
Player's search frequency on Google (standardized)	−0.038	0.000	0.296*
Dummy for players playing in both feet	0.215	0.178	0.458
Dummy of players playing in left foot	−0.258*	−0.231	−0.569
Dummy of the players playing in right foot	−0.195	−0.168	−0.486
Dummy of midfield players (Midfielders)	−0.169**	−0.122	−0.530**
Number of transfers done by the current/buying club 2007/2008–2018/2019 (standardized)	−0.069**	−0.051	−0.848**
Number of transfers done by the previous/selling club 2007/2008–2018/2019 (standardized)	0.003	0.006	−0.283
Volume of transfers (£) done by the current/buying club 07/08–18/19 (standardized)	0.052*	0.062*	0.028
Volume of transfers (£) done by the previous/selling club 07/08–18/19 (standardized)	0.139***	0.131***	0.189**
Remaining duration of previous contract at the time of current contract (standardized)	−0.001	−0.006	0.033
Number of followers of players on sofifa website (standardized)	0.474***	0.499***	0.358**
Dummy of players of Asian origin	−0.313*	−0.298*	
Dummy of players of African origin	−0.005	−0.053	0.338
Dummy of players of Australian origin	−0.347*	−0.121	−0.641
Dummy of players of European origin	−0.151***	−0.073	−0.345
Dummy of players of South American origin	0.127**	0.110*	0.540
Yearly Dummy for transfers of the 2013/2014 Two transfers windows	0.510***	0.364**	1.195*
Yearly Dummy for transfers of the 2014/2015 Two transfers windows	0.031	0.046	0.108
Yearly Dummy for transfers of the 2015/2016 Two transfers windows	−0.021	−0.023	0.113
Yearly Dummy for transfers of the 2016/2017 Two transfers windows	−0.447***	−0.380***	−0.670
Yearly Dummy for transfers of the 2017/2018 Two transfers windows	0.122	0.032	0.768
Constant	5.964***	5.343***	7.847**
Observations	2926	2416	510
R-squared	0.415	0.404	0.467
Adjusted R-squared	0.428	0.428	0.428

Robust standard errors in parentheses.

***p<0.01, **p<0.05, *p<0.1.

comprehensive understanding of a player's value and potentially more accurate pricing. The implication is that newer techniques that incorporate both types of data are becoming relevant and may be effective in pricing football transfers. For instance, the use of machine learning techniques, such as nonlinear regression methods, may show promise in improving the accuracy of pricing football transfers. These techniques can help identify complex relationships between variables and potentially outperform old methods. It

is again possible to use a comprehensive global analysis based on video gaming data and distinguishing different segments compared to previous literature.

Limitations and directions for future research

While the data used in this paper is certainly rich, it has limitations. We acknowledge that some variables in our regressions, such as goals and Champions League appearances, may be

Table 8. Breakdown by position and continent for the subjective complete package - non-defensive midfielders.

VARIABLES	World	World without England and Italy	England	Italy
Dummy for players moving after finishing previous contract	-0.513***	-0.366***	-0.959***	-0.658***
Dummy for players moving on loan	-0.327***	-0.209*	-1.058***	-0.275
Player's age is below 24 years old (Standardized)	0.132	0.216	0.595	-0.272
Player's age is above 24 years old (Standardized)	0.231*	0.266	0.0255	-0.418
Player's age is below 24 years old – squared (Standardized)	0.503***	0.415***	0.221	0.860***
players age is above 24 years old – squared (Standardized)	-0.347***	-0.380***	-0.164	-0.00698
duration of contract (Standardized)	0.138***	0.146**	-0.0215	0.103
player's height (Standardized)	0.212	0.107	-2.351	4.634**
player's search frequency as a Google trend (Standardized)	-0.0101	0.0333	-0.00105	-0.0842
Player's search frequency on Google (Standardized)	0.00811	0.019	0.000377	0.00118
Dummy for players playing in both feet	-0.116	0.157	-0.00443	-0.312
Dummy of players playing in left foot	0.287	0.154	0.48	-0.00998
Dummy of the players playing in right foot	0.194	-0.00733	0.426	
Pace (Standardized)	-0.146	0.102	1.643**	-1.128*
Shooting scores (Standardized)	0.0739	-0.504	-0.199	1.261*
Dribbling (Standardized)	1.847***	1.890***	1.459	2.017**
Passing (Standardized)	2.731***	2.666***	4.684***	2.556*
Defending (Standardized)	0.00341	-0.201	0.133	0.246
Physicality (physical strength) (Standardized)	1.098***	1.648***	0.915	0.488
Remaining duration of previous contract at the time of current contract (Standardized)	0.0248	0.0401	-0.0332	0.0391
Number of followers of players on sofifa website (Standardized)	0.191***	0.238***	0.120**	0.0194
Number of transfers done by theprevious/selling club 2007/2008–2018/2019 (Standardized)	-0.0112	0.0206	-0.0593	0.132*
Number of transfers done by the current/buying club 2007/2008–2018/2019 (Standardized)	0.0557	0.0656	0.064	0.361***
Volume of transfers (£) done by the previous/selling club 07/08–18/19 (Standardized)	0.0203**	0.0113	0.0177	0.00152
Volume of transfers (£) done by the current/buying club 07/08–18/19 (Standardized)	0.0286***	0.0164	0.0536**	-0.00115
Dummy of players of Asian origin	-0.307	-0.229	-1.667*	-0.99
Dummy of players of African origin	0.112	0.146	0.125	-0.0744
Dummy of players of Australian origin	-0.703*	-0.446		-0.51
Dummy of players of European origin	-0.0726	0.0386	-0.35	-0.400**
Dummy of players of South American origin	0.0254	0.0494	-0.112	0.0372
Yearly Dummy for transfers of the 2013/2014 Two transfers windows				
Yearly Dummy for transfers of the 2014/2015 Two transfers windows	-0.360***	-0.397**	-0.149	-0.208
Yearly Dummy for transfers of the 2015/2016 Two transfers windows	-0.516***	-0.529***	-0.341	-0.442
Yearly Dummy for transfers of the 2016/2017 Two transfers windows	-1.268***	-1.405***	-0.883***	-1.016***
Yearly Dummy for transfers of the 2017/2018 Two transfers windows	-0.351***	-0.667***	0.457*	0.179
Yearly Dummy for transfers of the 2018/2019 Two transfers windows	-0.468***	-0.583***	-0.681*	-0.111
Constant	-10.57***	-10.87***	-20.35***	-12.16**
Chi2	26.41	6.67	1.75	8.18
Prob>Chi2	0	0.0098	0.1864	0.0042
Observations	1227	674	227	326
R-squared	0.642	0.679	0.736	0.675
Adj. R-Squared	0.6318	0.6616	0.6891	0.6373

jointly determined, raising concerns about multicollinearity and endogeneity. Our analysis focused on predictive accuracy rather than strict causal inference, and we interpreted coefficient estimates descriptively. We performed correlation and VIF checks before estimation and found no evidence of problematic multicollinearity.

The completeness of the data used to price football transfers using video gaming data may be limited. The data may not capture all relevant factors that affect the pricing function of a football player, such as injuries, team dynamics, and player behaviour off the field. The use of video gaming data to price football transfers may

not be generalizable to all football players and teams. The data may only be representative of a specific subset of players and teams, and the results may not be applicable to other contexts. The use of video gaming data to price football transfers may lack transparency, as it may not be clear how the data is being used to determine the value of a player. This lack of transparency may lead to scepticism and mistrust among stakeholders in the football industry. To address these challenges, we suggest some new directions for future research. Future research could incorporate additional data sources, such as social media data, to improve the accuracy and completeness of the data used to price football transfers. New machine learning techniques could be developed to better analyse and interpret the data used to price football transfers. These techniques could help to identify new factors that affect the pricing function of a football player and improve the accuracy of transfer fee estimates. To address concerns about the lack of transparency in the use of video gaming data to price football transfers, future research could focus on developing more transparent and explainable models. This could help to build trust and confidence among stakeholders in the football industry. Future research could also consider the ethical implications of using video gaming data to price football transfers. This could involve examining issues such as data privacy, fairness, and bias, and developing guidelines and best practices for the responsible use of football data. Addressing post-pandemic effects is an important direction for future research, as the current paper covered the periods up to the 2019/2020 season.

In terms of future model refinement to enhance predictive accuracy and interpretability, the choice of model depends on the nature of the data, the complexity of relationships between variables, and the specific goals of the analysis. Machine learning models are particularly promising for improving predictive accuracy in transfer fee estimation, while advanced regression methods like GLM and fixed effects models offer robust alternatives to address statistical limitations of OLS. Finally, it is possible to explore threshold effects and ascertain the optimal age for footballers in terms of their transfer fees. We were unable to do this, probably because of the different packages. However, it is important to note that this is very feasible with actual transfer fees. Future research could build from our analysis and apply these directions for further model refinement.

ORCID iDs

Moussa Ezzeddine  <https://orcid.org/0000-0002-0021-9139>
 Pierre-Charles Pradier  <https://orcid.org/0000-0003-0808-7668>
 Nicolas Scelles  <https://orcid.org/0000-0002-6177-5307>

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Notes

1. The video gaming database refers to systematically collected data derived from interactive football video games played on platforms such as PlayStation, Xbox, PC, and mobile devices. In this context, football gaming primarily involves sports simulation titles (for example, EA Sports FC, formerly FIFA) and management or strategy-based games (for example, Football Manager or Championship Manager). It can also include fantasy football platforms, where participants manage virtual teams based on real-world athletes. The database captures diverse categories of information, including in-game performance metrics (such as match statistics, tactical choices, and progression levels), behavioural patterns of players (such as decision-making frequency, time spent in gameplay, and interaction styles), demographic data, and qualitative feedback. These data are generated directly from player interactions within the game environment, logged systematically by the gaming platforms and publishers. To ensure validity, the database draws on standardized and replicable metrics incorporated by the game providers themselves, such as skill rating systems and performance scores that are regularly calibrated across global user bases. Certain platforms further incorporate expert-based assessments, where professional coaches, analysts, or vetted gaming professionals contribute to evaluations of player skill and realism of performance modelling. While some elements of feedback can be community-driven, the primary structure and metrics of the database are designed by game developers to ensure consistency, comparability, and objectivity across players and regions.
2. As well as numerous known sites such as “www.google.com”, “www.bigsoccer.com”, “www.tifosobilanciato.it”, “www.footyrate.com”, “www.sportune.fr”, “www.calcioefinanza.com”, “www.totalsportek.com”, “www.sportlens.com”, “@swissramble”, www.deloitte.com (Deloitte Football Money League), Sky Sports, and Sports Intelligence Report.
3. See Appendix 1 for information about the sample of recent studies.
4. We have also tried estimating the models using random forest regressions. We preferred the results from the ordinary regression as it had a good fit, and the results are quite intuitive. It also quite clearly addresses the challenges of selection bias.

References

- Scarpello V and Theeke HA (1989) Human resource accounting: A measured critique. *Journal of Accounting Literature* 8: 265–280.
- Carmichael F and Thomas D (1993) Bargaining in the transfer market: Theory and evidence. *Applied Economics* 25(12): 1467–1476.
- Speight A and Thomas D (1997) Football league transfers: A comparison of negotiated fees with arbitration settlements. *Applied Economics Letters* 4(1): 41–44.

- Dobson S and Gerrard B (1999) The determination of player transfer fees in English professional soccer. *Journal of Sport Management* 13(4): 259–279.
- Carmichael F, Forrest D and Simmons R (1999) The labour market in association football: Who gets transferred and for how much? *Bulletin of Economic Research* 51(2): 125–150.
- Morrow S (1999) *The new business of football: accountability and finance in football*. Edinburgh: Macmillan Business.
- Dobson S, Gerrard B and Howe S (2000) The determination of transfer fees in English nonleague football. *Applied Economics* 32(9): 1145–1152.
- Herm S, Callsen-Bracker HM and Kreis H (2014) When the crowd evaluates soccer players' market values: Accuracy and evaluation attributes of an online community. *Sport Management Review* 17(4): 484–492.
- Bradley PS, Archer DT, Hogg B, et al. (2016) Tier-specific evolution of match performance characteristics in the English premier league: It's getting tougher at the top. *Journal of Sports Sciences* 34(10): 980–987.
- Müller O, Simons A and Weinmann M (2017) Beyond crowd judgments: Data-driven estimation of market value in association football. *European Journal of Operational Research* 263(2): 611–624.
- Szymanski S (2017) Entry into exit: Insolvency in English professional football. *Scottish Journal of Political Economy* 64(4): 419–444.
- Scelles N, Szymanski S and Dermit-Richard N (2018) Insolvency in French soccer: The case of payment failure. *Journal of Sports Economics* 19(5): 603–624.
- Payyappalli VM and Zhuang J (2019) A data-driven integer programming model for soccer clubs' decision making on player transfers. *Environment System and Decisions* 39(4): 466–481.
- Szymanski S and Weimar D (2019) Insolvencies in professional football: A German sonderweg? *International Journal of Sport Finance* 14(1): 54–68.
- Yan H (2020) How soccer players' box score statistics effect on their rating and market value. *The Frontiers of Society, Science and Technology* 2(15): 82–104.
- Felipe JL, Fernandez-Luna A, Burillo P, et al. (2020) Money talks: Team variables and player positions that most influence the market value of professional male footballers in Europe. *Sustainability* 12: 3709.
- Garcia-del-Barrio P and Pujol F (2020) Recruiting talent in a global sports market: Appraisals of soccer players' transfer fees. *Managerial Finance* 47(6): 789–811.
- García J and Murillo C (2020) Sports video games participation: What can we learn for esports? *Sport, Business and Management: An International Journal* 10(2): 169–185.
- Feuillet A, Terrien M, Scelles N, et al. (2021) Determinants of cooptation and contingency of strategic choices: The case of professional football clubs in France. *European Sport Management Quarterly* 21(5): 748–763.
- Scelles N, Peng Q and Valenti M (2021) Do the peculiar economics of professional team sports apply to esports? Sequential snowballing literature reviews and managerial implications. *Economies* 9(1): 31.
- Richau L, Follert F, Frenger M, et al. (2021) The impact of investors on transfer fees in the English premier league: A study of the ownership structures. *Corporate Ownership & Control* 18(3): 241–256.
- Behravan I and Razavi SM (2021) A novel machine learning method for estimating football players' value in the transfer market. *Soft Computing* 25(3): 2499–2511.
- Pizzo AD, Su Y, Scholz T, et al. (2022) Esports scholarship review: Synthesis, contributions, and future research. *Journal of Sport Management* 36(3): 228–239.
- Rubio Martin G, García CMM, Rodríguez-López Á, et al. (2022) Measuring football clubs' human capital: Analytical and dynamic models based on footballers' life cycles. *Journal of Intellectual Capital* 23(5): 1107–1137.
- Coates D and Parshakov P (2022) The wisdom of crowds and transfer market values. *European Journal of Operational Research* 301(2): 523–534.
- Franceschi M, Brocard JF, Follert F, et al. (2024) Determinants of football players' valuation: A systematic review. *Journal of Economic Surveys* 38(3): 577–600.
- Barbuscak L (2018) What makes a soccer player expensive? Analyzing the transfer activity of the richest soccer. *Augsburg Honors Review* 11(1): 5.
- Depken CA and Globan T (2021) Football transfer fee premiums and Europe's big five. *Southern Economic Journal* 87(3): 889–908.
- FIFA (2019) *Global transfer market report 2019: Men professional football: A review of international football transfers worldwide*. <https://digitalhub.fifa.com/m/248987d86f2b9955/original/x2wrqjstwjoianncnod-pdf.pdf>.
- Poli R, Ravanel L and Besson R (2017) "How to evaluate a football player's transfer value?" CIES.
- Kirschstein T and Liebscher S (2019) Assessing the market values of soccer players—a robust analysis of data from German 1. and 2. Bundesliga. *Journal of Applied Statistics* 46(7): 1336–1349.
- McHale G and Holmes B (2023) Estimating transfer fees of professional footballers using advanced performance metrics and machine learning. *European Journal of Operational Research* 306(1): 389–399.
- Poli R, Besson R and Ravanel L (2021) Econometric approach to assessing the transfer fees and values of professional football players. *Economies* 10(1): 4.
- UEFA (1992) *Principles of co-operation between member states of UEFA and their clubs*. Berne, Switzerland: Union des Associations Européennes de Football.

Appendices

Appendix 1. Overview of recent studies.

Study	Sample Size	Focus	Methodology
Kirschstein and Liebscher (2019)	500	Market value predictions using FIFA data	Machine learning (multiple algorithms)
Müller et al. (2017)	300	Market value based on player statistics	Regression models
Poli et al. (2021)	2000	Transfer fees using historical data	Multiple linear regression

Appendix 2. Descriptive statistics.

Variable	Obs	Mean	Std. dev.	Min	Max
Lfee	57,736	1.670	4.601	0	19.11283
freetransfer	57,736	0.290	0.454	0	1
tt_loan	57,736	0.234	0.423	0	1
lage	57,710	3.231	0.161	2.70805	3.78419
lage2	57,710	10.463	1.050	7.333536	14.32009
ldur	50,831	5.874	0.813	0	7.53743
lheight	56,397	0.587	0.037	0.4574249	0.7178398
lpace	19,038	4.261	0.162	3.332205	4.59512
ldribbling	15,446	4.163	0.218	2.70805	4.60517
lshooting	19,038	4.040	0.264	2.564949	4.584968
ldefending	19,038	3.965	0.339	2.772589	4.510859
lpassing	19,038	4.108	0.180	2.944439	4.543295
lphysicality	19,038	4.195	0.149	3.295837	4.543295
lgt	57,736	2.773	1.834	0	4.60517
lgoo	57,736	14.315	2.777	0	23.51749
footer_both	57,736	0.034	0.182	0	1
footer_left	57,736	0.250	0.433	0	1
footer_right	57,736	0.671	0.470	0	1
pos_forward	57,736	0.315	0.465	0	1
pos_mid	57,736	0.303	0.460	0	1
pos_def	57,736	0.291	0.454	0	1
lco_cuclu	57,736	4.159	1.595	0	6.113682
lco_preclu	57,736	4.399	1.400	0	6.102559
ltot_preclu	57,736	14.114	6.415	0	20.25701
ltot_cuclu	57,736	12.276	7.639	0	20.93291
nat_cont~sia	57,736	0.047	0.211	0	1
nat_~_africa	57,736	0.063	0.243	0	1
nat_cont~lia	57,736	0.005	0.068	0	1
nat_cont_e~e	57,736	0.343	0.475	0	1
nat_cont_s~a	57,736	0.568	0.495	0	1
y2008x	57,736	0.011	0.103	0	1
y2009x	57,736	0.085	0.278	0	1
y2010x	57,736	0.105	0.307	0	1
y2011x	57,736	0.104	0.305	0	1
y2012x	57,736	0.098	0.297	0	1
y2013x	57,736	0.108	0.310	0	1
y2014x	57,736	0.105	0.307	0	1
y2015x	57,736	0.109	0.312	0	1
y2016x	57,736	0.116	0.320	0	1
y2017x	57,736	0.115	0.320	0	1
y2018x	57,736	0.041	0.199	0	1