

Determinants of Attacking Player Transfer Fees in Europe's Top Five Leagues: A SHAP-Based Analysis

Aakrit Punia¹

Received June 14, 2026

Accepted September 15, 2026

Electronic access October 31, 2026

Background: Transfer fee is the biggest economic decision football clubs make, but its determining factors is only half-understood. Most studies have small samples or few variables and few focus on over represented attacking players among the best transfers. This study examines which player and competition factors strongly determine transfer fees of attacking players in Europe's top five leagues (2010–2025).

Method: Data Collection source was API-Football which contained 3,596 valid transfers with 1,053 transfers involving attacking players. SHAP values were calculated from a Gradient Boosting Regressor fitted to log-transformed fees, with five-fold cross-validation grouped by player. The Human Capital, Market efficiency and Bargaining Theory were used to analyse the impact of 13 variables of players and leagues.

Results: Goals scored and assists were the crucial positive determinants of transfer fees, with age and red cards having significant negative effects. The participation in the Premier League came with a large fee premium over the reference league, while the participation in UEFA competitions had a positive but small impact. The model explained an R-squared of 0.226 (22.6%) of fee variation under five-fold cross validation, below a linear regression baseline (R-squared = 0.344, 34.4%), suggesting a near linear relationship between statistics and fees.

Conclusions: Attacking transfer fees are driven by Scoring goals and league context and not by disciplinary record or prestige. Findings provide practical guidance for clubs, scouts and regulators, while the modest fit suggests , contract terms, agent influence and commercial appeal are unmeasured factors of fee variation.

Keywords: transfer fees, football economics, SHAP, machine learning, player valuation, sports economics

Introduction

The professional football game is now one of the most commercially significant industries in the sports economy. Besides broadcasting revenues and sponsorships, the football transfer market is now a huge economic system in the sport. A Transfer fee refers to the payment made by one club to another to gain the registration rights of a contracted player. This number takes into account not just athletic ability, but also a player's marketability, future potential and perceived economic value.

The theoretical foundation of this study draws on three complementary frameworks. Human Capital Theory¹ treats players as economic assets whose value derives from their skills, experience, and expected future contribution. Market Efficiency Theory² proposes that in an efficient market, transfer fees should fully reflect all observable player characteristics. Empirical evidence consistently shows deviations from this benchmark as valuations are also affected by reputation and commercial appeal, in line with findings on club affiliation and marketing driven premiums³. Bargaining Theory⁴

expands on how bargaining power, the urgency of the contract and the competitive interest of rival clubs can drive the ultimate fee above that justified by objective performance alone.

Transfer fees serve as a critical economic instrument within football. They are capital investments on club balance sheets, they impact on the compliance with financial regulations and they reinforce competitive hierarchies between leagues. There are a number of generally accepted factors that influence transfer value: performance indicators such as goals, assists and appearances; age and long term development potential; disciplinary record; physical attributes; and participation in elite competitions such as the UEFA Champions League and Europa League.

A growing body of literature examines determinants of transfer fee, but significant empirical gaps remain. The majority of existing studies are limited to a single league or a rather short time window³. Even studies based on larger multi league samples tend to find that contextual factors such as team quality and league strength dominate individual performance metrics, or highlight a narrower set of decisive variables such as contract length and age⁵. Few have studied

¹ Amity International School Sector-46, Haryana, India

transfer fee determinants at scale and none have specifically looked at attacking players despite forwards being disproportionately represented among the most expensive transfers in history⁶.

This study deals with the gap in question by examining 3,596 valid transfers involving attackers, midfielders, defenders, and goalkeepers, 1,053 of which concern attacking players who came from the top five professional leagues in Europe over a 15-year period from 2010 to 2025, the data being obtained through the API-Football service. It includes a wide range of variables such as appearances, goals, assists, age, yellow cards, red cards, and the league in which the players' underlying statistics were recorded, covering all of the top five domestic leagues as well as the UEFA club competitions, thus taking into account aspects of player valuation that are ignored by conventional studies. SHAP analysis is used to assess the relative importance and the direction of the effect of each variable, giving a more thorough and easier to interpret explanation of how transfer fees are determined in today's football, as shown in Figure 1.

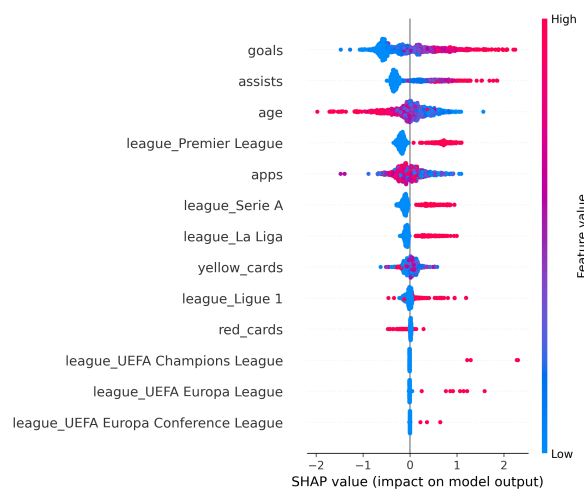


Figure. 1 SHAP summary plot for attacking player transfer fees (full sample). Each data point is a player-transfer observation, with the horizontal position being the SHAP value (impact on the predicted fee) and the colour indicating whether the underlying feature value was high (pink) or low (blue). This allows readers to view both the direction and spread of each variable's effect across the sample.

Research Question and Hypotheses

This study seeks to answer the following research question: which player related and competition related factors most strongly determine the transfer fees of attacking football players in Europe's top five leagues between 2010 and 2025, and in what direction do they operate? In line with the theoretical

frameworks outlined above, four hypotheses are proposed:

H1: Performance related variables (appearances, goals, and assists) will show a positive and meaningful relationship with transfer fees, consistent with Human Capital Theory.

H2: Disciplinary variables (yellow cards and red cards) will show a negative relationship with transfer fees, reflecting their effect on player reliability.

H3: Age will be inversely related to transfer fees, in line with clubs' preference for younger players with greater long term potential.

H4: Participation in elite UEFA competitions will show a positive directional impact on transfer fees but will be limited in overall importance compared with on field performance variables.

Literature Review

Economic and Statistical Approaches

Besson et al.⁵ used an econometric method based on multiple linear regression to examine transfer fees. For this purpose, they compiled a dataset consisting of 2,045 transactions involving money which took place between July 2012 and November 2021 in the English Premier League, the Spanish La Liga, the Italian Serie A, the German Bundesliga, and the French Ligue 1, the data having been checked by sports media and Transfermarkt. The model classified the variables into three categories: player performance, individual characteristics, and contextual factors, applied a logarithmic transformation to the fees, and employed five-fold cross validation. The study finds that an econometric approach is able to account for more than 80 per cent of the variations in transfer fees, contract length, age, and weighted experience being the most important factors; an extra contract year results in a 22 per cent increase in the fee, while being one year younger leads to an increase in valuation of about 12 per cent.

Kirschstein and Liebscher³ examined the extent to which a player's market value is determined by their skills and performance, employing secondary data from about 493 athletes in the German Bundesliga and analysing it across 28 performance indicators using robust regression and multivariate techniques. They found that market values are not entirely determined by performance indicators, some players being either overvalued or undervalued. The market value of a player is strongly influenced by the club they are associated with and by the club's reputation, star players tend to have their values inflated, and marketing opportunities among other external factors also have an effect. Player valuation is thus influenced by both performance and non-performance factors.

Merten⁷ examined the temporal relationship and the immediate return on investment associated with transfer expenditures, using data spanning from 1992 to 2021 which was ob-

tained from historical league tables and financial records supplied by Transfermarkt and Deloitte. In order to establish a link between net transfer spending and the number of points earned, as well as to identify the direction of the relationship between spending and performance, the study applied linear regression and Granger causality analysis, at the same time grouping the clubs into different spending categories. The paper finds that although transfer spending is the most effective means of achieving rapid improvement, it is a high-risk approach which frequently results in unsustainable debt, and therefore clubs need to balance their market aggression with investment in youth development and with wage-to-revenue stability.

Jose et al.⁸ carried out a large-scale, quantitative, cross-sectional study in which they analysed the data of 2,259 professional male football players from the five most valuable European leagues in the 2018–2019 period, using the players' Transfermarkt market values as the dependent variable. To investigate the interactions between age, playing position, team quality and league type, the researchers conducted normality tests, applied the Kruskal-Wallis H test and carried out a multi-factorial ANOVA, with post hoc Bonferroni corrections and calculations of effect sizes. They concluded that market value indicates a hierarchical system in which team quality and league strength are the two most important predictors, with players at top-tier clubs or in the English Premier League enjoying a substantial premium irrespective of their individual statistics.

Depken and Globan⁹ carried out an analysis of 5,760 international player transfers over a period of 14 years and discovered that English clubs pay considerably higher transfer fee premiums than do clubs in France, Germany, Italy, and Spain; they attributed the sharp increase in this English premium to the 2012 Premier League broadcasting agreements, thus giving a causal and league-specific explanation for the type of Premier League fee premium identified in the current study.

Rubio-Martin et al.¹⁰ carried out a study in which they analysed the risk of injury and its impact on market value for 5,336 player observations from seven European leagues between 2006 and 2020, predicting the probability of injury using logistic regression and then using a dynamic panel model to relate that probability to market value; they discovered that a 1 per cent increase in the probability of a serious injury was linked with a 2.29 per cent decrease in market value, a effect which became more pronounced when both the severity of the injury and its recurrence were taken into account.

Di Domizio et al.¹¹ looked at the assessment of transfer market values in the case of Italian Serie A, thereby giving a single-league example that supplements the broad, multi-league emphasis of most of the other studies in this field and providing a useful basis for comparison with the premium for Serie A that is identified in the current study.

Machine Learning Approaches

Li et al.¹² applied SHAP-based interpretable machine learning to over 800 real world transfer records from the Big Five leagues covering the 2017–2018 to 2019–2020 seasons in order to determine the main factors that influence the valuation of professional soccer players. They discovered that a few fundamental human capital factors, for example age, and labour market factors such as the length of the remaining contract, were the key drivers of value, outweighing technical abilities like goal scoring, and that general athletic attributes had more predictive value than soccer-specific skills.

McHale and Holmes¹³ arrived at estimates of the transfer fees of professional footballers by combining advanced performance data for players with machine learning algorithms and discovered that machine learning models exceeded the linear regression benchmarks in terms of out-of-sample prediction accuracy; the model thus obtained was also applied in order to evaluate value for money in individual transfers and for clubs as a whole.

Positional and Career-Stage Analyses

Khalife et al.¹⁴ looked at the way various factors that affect market value differ according to playing position and stage of a player's career, using a dataset related to FIFA that was analysed by means of machine learning, specifically with Gradient Boosting models and cross validation. They discovered that the factors affecting valuation change depending on position, forwards being more dependent on shooting and dribbling, midfielders on passing, and defenders on defensive abilities. The significance of these factors changes with age since younger players are valued more in terms of their future potential and reputation, whereas older players are valued based on their actual performance. Future potential was important for setting market value, and machine learning provided a greater level of accuracy than traditional methods.

In the study by Liu¹⁵, the way in which player market values decrease after their peak years in the top five European leagues was investigated, the data for the 2023-2024 season being used. It was found that there was a steady fall in market value starting at about age 27, the most rapid decline occurring in the Bundesliga and the slowest in Serie A and La Liga, while the Premier League and Ligue 1 exhibited a more moderate rate of decline, showing that the league in which a player competes has an effect not only on their current valuation but also on how quickly that valuation decreases with age.

Valuation Theory and Market Efficiency

Hill et al.¹⁶ claim that value should be based on a player's on-field performance and on factors such as age and experience rather than on secondary considerations like the length

of a contract, employing a conceptual, interdisciplinary review which links sports management research with corporate finance. They divide valuation methods into four categories: intrinsic, relative, real options and probabilistic ones, and examine existing models such as the Statistical Trend Analysis Rating System and the Player Quality Index, thereby creating a decision framework for choosing an appropriate method. They assert that a rigorous framework based on finance is necessary for objectivity, and that although intrinsic valuations are theoretical, relative valuations are still the most practical tool since they make use of standardised performance metrics.

Franceschi et al.¹⁷ carried out a systematic review of 29 peer-reviewed studies concerning the valuation of football players, grouped the independent variables used in the various studies and suggested a common group of core explanatory variables together with the expected direction of their effects. They found that even after thirty years of research there are still considerable gaps to be overcome in reconciling the econometric and machine learning approaches and in extending the analysis to go beyond aggregate or position-agnostic samples.

Franceschi et al.¹⁸ proposed a unified theoretical framework for valuing football players, the framework being based on subjective value theory from the Austrian school of economics and differing between a player's use value to a particular club and their overall market value; they applied this framework in a critical examination of the conceptual validity of crowd-sourced valuation sites such as Transfermarkt.

Coates and Parshakov¹⁹ directly put the wisdom of crowds hypothesis that forms the basis of Transfermarkt valuations against the actual transfer fees and found that although the market values derived from the crowd closely correlate with the real fees, they systematically underestimate the price that is actually paid in the majority of transfers, a point which is relevant to any study that relies on crowd-sourced or aggregated secondary data, such as the present study's use of the API-Football service as a stand-in for transfer fees.

Peeters²⁰ examined whether the crowd-sourced valuations on Transfermarkt accurately reflected the subsequent transfer fees and the results of international matches, thus providing an empirical basis for using Transfermarkt data in academic research and making clear the conditions under which crowd valuations match true player quality.

Financial, Ownership, and Commercial Factors

Sauer et al.²¹ carried out an analysis of the way financial investors generate value in football as compared with the principles of conventional investment theory, employing a Straussian grounded theory method. The investors were chosen on the basis of their previous investment activities, the use of third party equity, their active investment behaviour, and the fact

that their holding periods were short. The data were gathered in two stages: first from secondary sources such as Mergermarket, the investors' websites, press releases and articles, and then through 16 semi-structured interviews with investors, advisors, regulators and club executives. With the aid of Atlas.ti, the researchers performed open coding, obtained strong inter-coder reliability, and reached theoretical saturation before carrying out axial coding, which resulted in the identification of four core themes showing how these investors differ from traditional private equity. The findings assist club managers in understanding the motivations of investors, enable investors to see that football can be a viable market, and aid regulators in designing suitable safeguards.

Metelski⁶ looked at the main factors that influence financial aspects of the football transfer market by looking at the top 100 most expensive player transfers in history, using data from Transfermarkt. The players were categorised as goalkeepers, defenders, midfielders, and forwards in order to find out whether offensive players are more frequently included among the highest-value transfers. To examine age-related trends in valuation, descriptive statistics and the Pearson correlation coefficient were used. The results show a strong emphasis on offensive players, since attackers are more often found amongst the most valuable transfers, which shows the value placed on goal-scoring ability. The high prices paid for young players suggest that clubs are willing to spend money on their future potential, and the study points out financial inequalities in that a small number of well-off teams regularly spend more than the rest.

Richau et al.²² studied the effect of ownership concentration and the origin of investors on the transfer fees of individual players in the English Premier League, looking at 1,124 transfers from the 2012/13 to the 2018/19 period. Ownership concentration was divided into three categories: dispersed, majority, and single ownership, and the largest shareholder was classified as either being domestic or foreign. The researchers used OLS models with the transfer fees transformed using a logarithm and took into account factors such as the players' age, position, and performance, finding that there was no clear evidence to show that greater ownership concentration by itself increases the willingness to pay. On the other hand, single ownership might have a negative impact on transfer fees, while foreign investors, in particular those who are controlling shareholders, tend to pay a premium for players, especially when considering the higher parts of the fee distribution.

Hoey et al.²³ investigated whether the football transfer system acts as a pro-competitive measure by redistributing financial resources among the clubs, employing industrial organization techniques to analyse transfer data drawn from the financial statements of 202 European clubs. They found that the transfer system only slightly reduces financial inequality

rather than significantly improving competitive balance; mid-sized clubs derive the greatest benefit from transfer income whereas smaller clubs receive only a small amount, and the large clubs are able to absorb transfer losses without losing their financial leading position. Consequently, the system has only a small pro-competitive effect and does not equalise the competitive environment.

Uzik et al.²⁴ developed a mathematical model which linked transfer prices to a player's fan base and their presence on social media, claiming that a player who is widely followed can help a club to recover part of the transfer fee by means of merchandise and sponsorship income, and thus extend the valuation of a player beyond their on-field performance into the commercial area.

Malagon-Selma et al.²⁵ have proposed a new method for measuring player popularity by means of Google Trends search data; they developed six popularity indices and demonstrated that including these indices reduced the error in transfer fee predictions in the cases of linear regression, random forest, and gradient boosting models, thus supporting the idea that off-field popularity has measurable predictive value together with on-field performance.

Synthesis and Research Gap

When the studies reviewed are considered as a whole, they reach three main conclusions. Firstly, performance-related factors such as goals, assists, number of appearances, and age consistently appear as key determinants of a player's market value, regardless of the methodological approach used from econometric models⁵ to machine learning techniques¹⁴. Secondly, non-performance factors such as the club's reputation, the ownership structure, the origin of the investors, and the strength of the league also have a significant impact, showing that the football transfer market differs from the strict forecasts of Market Efficiency Theory^{3,8,22}. Thirdly, while some studies have looked at transfer fees at an aggregate level or have focused on the top 100 transfers⁶, few have combined a large multi-league dataset with an interpretable machine learning method (even though Li et al.¹² are close to this, using a similar multi-league SHAP-based approach without concentrating on attackers), and no study has specifically focused on attacking players even though attackers are well known to be overrepresented among the most expensive transfers. This study fills that gap by applying SHAP-based interpretation to a large dataset that is focused on attackers and covers 15 years across five major European leagues.

Theoretical Framework

Market Efficiency Theory holds that in an efficient market, prices fully reflect all available information². If the football

transfer market were perfectly efficient, a player's transfer fee would equal the true market value based on performance, age, potential, and other factors, with no systematically undervalued or overvalued players. However, evidence suggests deviations from efficiency, as player valuations are influenced by both performance and non performance factors, leading to systematic overvaluation and undervaluation³. This study examines whether observable player characteristics significantly influence transfer fees, using SHAP to evaluate how efficiently the market prices player attributes.

Human Capital Theory holds that individuals possess skills, knowledge, experience, and abilities that enhance productivity and economic value¹. Applied to professional sport, athletes are economic assets whose value depends on expected future contribution. Player characteristics such as goals, assists, appearances, experience in elite competitions, and technical abilities can be interpreted as forms of human capital, and clubs invest significant fees in players expected to deliver sporting success and long term performance benefits.

Bargaining Theory holds that bargaining outcomes depend on the relative negotiating power of the parties⁴. The final negotiated price is influenced by factors beyond the intrinsic value of the asset. Bargaining power, alternatives, urgency, and strategic position all play a part. Two players of similar performance might command different fees because selling clubs ask higher fees for strategically important players, while buying clubs might pay more because of urgency, competition from rivals or commercial factors.

Methods

Data and Variables

Forwards, who are also known as attackers, are outfield players who play further up the pitch than midfielders or defenders. Because of their position further forward and the relatively small amount of defensive duty they have, forwards generally score more goals than any other kind of player on the team. There are a number of sub-positions within this attacking group, such as strikers, wingers (on the left and right), central forwards, false nines, and second strikers. However, the data in question does not identify these sub-positions individually; instead, the players are grouped under the more general categories of Attacker and Forward, these categories together covering all the attacking roles mentioned above and forming the basis for the position filter used in this study.

The study makes use of secondary data obtained through the API-Football service and looks at the factors that influence transfer fees from 2010 to 2025 in the top five professional leagues in Europe. Although the entire dataset includes 3,596 valid transfers involving all positions, the SHAP analysis is carried out on the subgroup consisting of 1,053 transfers

by players (768 different players), since the study concentrates on forwards. As the dependent variable, the player's transfer fee is used, having first been log transformed in order to standardise the distribution and minimise skewness, in accordance with previous research. The independent variables consist of six numeric performance and disciplinary indicators (appearances, goals, assists, age, yellow cards, and red cards) as well as seven league dummy variables which represent the competition in which the player's underlying statistics were collected (Premier League, Serie A, La Liga, Ligue 1, UEFA Champions League, UEFA Europa League and UEFA Europa Conference League), with the Bundesliga serving as the reference category.

The encoding procedure chose the Bundesliga to be the reference category since it comes first alphabetically out of the eight league categories, not for any actual reason; this decision has no impact on model performance and affects only which league's coefficient is directly shown in Table 1, as all the other leagues are given in relation to the one that has been left out.

It is also a reasonable rather than a distorting choice since the Bundesliga is not a small or unusual league when compared with the others in the sample; it involves 225 attacker transfers, the second highest number after the Premier League's 227 and well ahead of Ligue 1 (211), Serie A (197), and La Liga (177), thus making it an appropriate baseline rather than an artifact of limited data. The idea of using the Premier League as the reference category had been considered but discarded, as that would mean excluding the Premier League's own coefficient from Table 1 and would require its premium—the single largest league effect found in this study—to be reported only indirectly as a negative coefficient for all the other leagues. By keeping the Bundesliga as the reference category, it is possible to state the Premier League premium directly, which is the way the most recent comparable study in this field presents its own results.

The data were obtained via the API-Football service (api-football.com), this service offering transfer, appearance, and disciplinary information in a structured format. As part of the data cleaning process, exact duplicate rows were deleted since the original export included repeated entries for most of the transfer records; the position field as reported was standardised into the Attacker and Forward categories mentioned above; and the transfer fee values, which were given as text strings in formats such as 95K or 5.5M €, were converted into numerical euro values. Any records that had a missing or zero transfer fee or missing values for any of the features considered by the model were excluded from the analysis.

The following fields were taken from the source data: Transfer Date, Transfer Fee, Team From, Team To, Player ID, Player Name, Age During Season, Stat League, Stat Season, Position, Appearances, Goals Total, Assists, Yellow Cards,

and Red Cards; fields available in the source such as minutes played, passing, tackling, and dribbling statistics were not included in the feature set of the model. In order to be included, the data had to have a Transfer Fee that was greater than zero and not missing, a Position value of either Attacker or Forward, and non-missing values for all the features in the model; any records that failed to meet any of these conditions were excluded, together with the exact duplicate rows mentioned above.

The transfer fees, as provided by the source, are given in euros and are used in their nominal form; during the period from 2010 to 2025 no adjustment has been made for inflation or for fluctuations in currency value. This approach is a simplification because a nominal euro in 2010 does not have the same purchasing power or indicate the same market conditions as a nominal euro in 2025, a fact which is addressed again in the section below dealing with temporal effects.

The dataset does not separately classify loan moves, free transfers, or fees which include performance-based add-ons and sell-on clauses, and it also does not indicate mid-season transfers any differently from transfers made at the end of the transfer window. It is therefore necessary to include all these types of transfer in the sample since the original data is not reliable in distinguishing between them; particularly loan fees and any undisclosed or nominal free-transfer fees cannot be directly compared with permanent-move fees that have a fully disclosed value, and this is regarded as a limitation of the data available rather than a decision made in the modelling process.

The values for each performance predictor—such as appearances, goals, assists, yellow cards, and red cards—show the data from the player's most recent completed season before the transfer date, since this is the information a club would realistically have seen when discussing the transfer. Statistics from seasons after the transfer are not included in order to prevent lookahead bias in the features of the model.

Many of the variables which had been identified in previous literature as relevant to transfer valuation—such as contract length, wage level, international caps, injury history, marketability, release clauses, nationality, and height—are not available via the data source employed in this study and are therefore not included in the model. This point is examined again in the Limitations section, where the effects it has on the model's unexplained variance are further discussed.

SHAP and Estimation Procedure

SHAP (SHapley Additive exPlanations) is a technique in the field of interpretable machine learning which is used to show the contribution of each individual variable to a model's prediction. It makes use of Shapley values from cooperative game theory: each feature is considered to be a player in a game and the predicted transfer fee is the amount that the player re-

ceives. The SHAP value indicates the amount by which each feature contributes to either increasing or decreasing the predicted transfer fee in comparison to the average prediction. The framework offers both global interpretability, by determining which variables are most important as a whole, and local interpretability, by illustrating how a particular variable influences a single prediction.

In the study, SHAP values were obtained from a tree-based ensemble (Gradient Boosting Regressor) that had been trained on the entire log transformed transfer fee sample, and the model's generalisation ability was evaluated separately by means of cross validation, as described below. The SHAP value for each feature shows its average marginal contribution to the predicted transfer fee for all possible combinations of features. The Direction of Impact (DOI) given in the results refers to the correlation between the values of the features and their respective SHAP values, positive values showing that higher feature values lead to higher predicted transfer fees and negative values showing the opposite effect. Overall importance is calculated as the mean absolute SHAP value, with the values being normalised in order that they can be compared across different features. The Gradient Boosting Regressor was set with `n_estimators = 500`, `learning_rate = 0.05`, `max_depth = 5`, `subsample = 0.8`, and `random_state = 42`, the hyperparameters having been chosen by manual tuning rather than through an automated search method such as grid search or Bayesian optimisation.

Cross-Validation and Baseline Comparison

In order to examine how well the model generalises beyond a single fit on the entire sample, five-fold cross validation was carried out, grouping the data by player, in such a way that a player who had more than one transfer in the sample would not appear in both the training and the test parts of the same cross-validation split. In order to determine whether the extra complexity introduced by the Gradient Boosting Regressor is worthwhile, its cross-validated performance was also compared with that of a linear regression baseline which was estimated using the same features and the same folds. The results for both models are given and discussed in Table 2.

Ethical Considerations

This study used only publicly available, aggregated secondary data from the API-Football service and did not involve human participants, personal data, or animal subjects. No ethical approval was therefore required.

Results

The main results are reported in Table 1, which presents the direction of impact and overall importance of each variable.

Table 1 SHAP results for transfer fee (attacking players)

Variable	Direction of Impact	Overall Importance
Goals	0.937	0.495
Assists	0.874	0.393
Age	-0.810	0.301
Premier League	0.970	0.297
Appearances	-0.373	0.222
Serie A	0.932	0.187
La Liga	0.929	0.134
Yellow cards	-0.008	0.096
Ligue 1	0.422	0.067
Red cards	-0.734	0.025
UEFA Champions League	0.960	0.015
UEFA Europa League	0.936	0.015
UEFA Europa Conference League	0.930	0.003

Source : author's calculations

The SHAP results for the transfer fee (with regard to attacking players from 2010 to 2025) are given in Table 1. The Direction of Impact shows the correlation between the values of the features and the associated SHAP values; Overall Importance refers to the mean absolute SHAP value, expressed on a normalised scale. The league variables are coded in relation to the Bundesliga, which is the reference category left omitted. The Overall Importance figures do not add up to one when summed over all the features; instead, each value has been independently normalised as the mean absolute SHAP value for that variable and must therefore be interpreted as indicating a relative ranking rather than as representing a proportion of the total explained variation.

The goal statistic has a direction of impact of 0.937 and an importance of 0.495, which means it is by a clear margin the most significant factor in the model. The number of goals scored is very strongly associated with increased transfer fees – consistent with the view that attackers are mainly valued for their ability to convert chances into goals.

The second most important factor is the variable Assist with an impact value of 0.874 and an importance of 0.393. There is a connection between players who create more goals for their teammates and higher transfer fees, suggesting that clubs value creative play in addition to direct goal scoring.

Age has a negative impact of -0.810 and an importance of 0.301, which means it is the third most important factor; older players are linked with considerably lower transfer fees, this being in keeping with clubs charging less for players who have less resale value and shorter careers remaining.

The Premier League has a direction of impact of 0.970 and an importance of 0.297, which is the highest league effect in the model; compared with the Bundesliga, which is the reference category, a player whose statistics were collected in the

Premier League is linked to a significant fee premium, this reflecting the league's higher broadcasting revenues and spending power.

The impact of appearances is -0.373 and its importance is 0.222. Rather counterintuitively, the more appearances a player has, the lower their predicted fee becomes when goals, assists, and age have already been taken into account. This is probably due to multicollinearity: appearances are strongly correlated with goals ($r = 0.71$, see Diagnostic Testing), so after the model has already assigned value to goal output, further appearances without a corresponding increase in goals may indicate that the player is part of the squad rather than a key performer. This finding is discussed again in the Discussion.

The fact that playing in Serie A (with a direction of impact of 0.932 and an importance of 0.187) and La Liga (with a direction of impact of 0.929 and an importance of 0.134) is linked to considerable positive fee premiums when compared to the Bundesliga—although these are smaller than the effect seen in the Premier League—shows the relative commercial strength of these leagues.

Yellow cards have an impact direction of -0.008 and an importance of 0.096; the direction is near zero, which shows that there is no consistent relationship between disciplinary bookings and fees once the effects of the other variables have been taken into account, even though the variable as a whole has moderate importance.

The impact score for playing in Ligue 1 is 0.422 and its importance is 0.067, which is a positive but smaller premium than that of the other top five leagues, this being in line with Ligue 1's relatively lower broadcasting revenues.

The red card factor has a value of -0.734 and an importance of 0.025. The more frequently players are sent off, the lower their transfer fees tend to be, which is in line with the disciplinary risk and the lost playing time connected with receiving a red card.

There is a consistent pattern when it comes to participation in UEFA club competitions: the Champions League (with a direction of impact of 0.960 and an importance of 0.015), the Europa League (with a direction of impact of 0.936 and an importance of 0.015), and the Europa Conference League (with a direction of impact of 0.930 and an importance of 0.003) are all positively linked to fees, though they make only a small contribution to the overall model importance. Although taking part in continental competition is linked to higher fees, this effect is mostly offset by the domestic league and performance variables when these are taken into account.

Local Explanations

Although Table 1 shows the global importance of various features, SHAP also enables each individual prediction to be broken down into the specific contributions made by each feature

for that particular player. Waterfall plots were looked at for three typical attacking players selected from the lower, middle, and upper groups of transfer fees in the sample: one with a transfer fee of €1.0 million, another with €7.0 million, and a third with €46.5 million. In the case of the player with the lowest fee, the biggest negative factors were zero goals and zero assists, which pulled the prediction well below the average, an effect only partially counteracted by the positive impact of having played in Ligue 1. In the case of the player with the highest fee, the most significant positive contribution came from a high number of goals (13 goals), this being supported by playing in the Premier League, the number of appearances, the number of yellow cards, and age. These local explanations show that the factors identified in Table 1 do not work in the same way across the entire range of fees: goal output seems to have the greatest influence on the prediction at the top end of the market, while league context and the lack of output are the most important factors at the lower end.

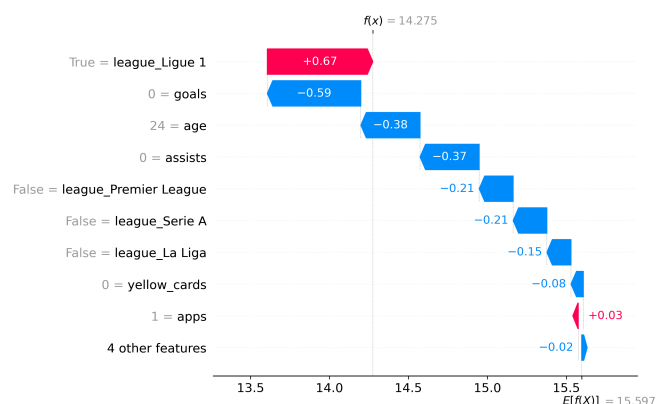


Figure. 2 Local SHAP waterfall plot, low-fee example (€1.0 million transfer). Bars show how each feature pushed the prediction for this individual player up (pink) or down (blue) from the model's average predicted fee, $E[f(X)]$, to the final predicted value $f(x)$ shown at the top.

Table 2 reports model performance under five fold cross validation grouped by player, rather than a single train test split, giving a more reliable estimate of how well the model generalises to unseen players. The Gradient Boosting Regressor achieved a mean R-squared of 0.226 (SD = 0.087), a mean RMSE of 1.208, and a mean MAE of 0.908 across the five folds. A linear regression baseline, estimated on the same features and the same folds, achieved a mean R-squared of 0.344, a mean RMSE of 1.116, and a mean MAE of 0.857, outperforming the more complex Gradient Boosting Regressor. This indicates that the relationship between the observed player statistics and transfer fees is close to linear in this sample, and that the added flexibility of gradient boosting may be overfitting to noise given the limited number of attacking

Table 2 Cross-Validated Model Performance

Model	R-squared (mean $\pm$ SD)	RMSE (mean)	MAE (mean)
Gradient Boosting Regressor	0.226 $\pm$ 0.087	1.208	0.908
Linear Regression (baseline)	0.344 $\pm$ 0.062	1.116	0.857

Sources: Author’s calculations. Five-fold cross-validation grouped by player.

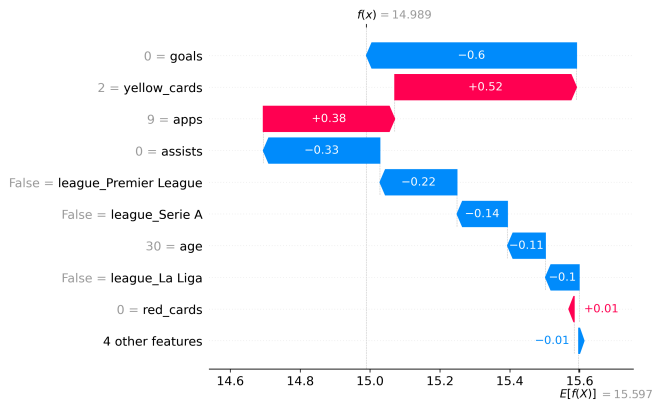


Figure. 3 Local SHAP waterfall plot, mid-range example (€7.0 million transfer). As in Figure 2, bars show the feature-by-feature contributions that move this individual player’s prediction away from the model’s average toward its final predicted value.

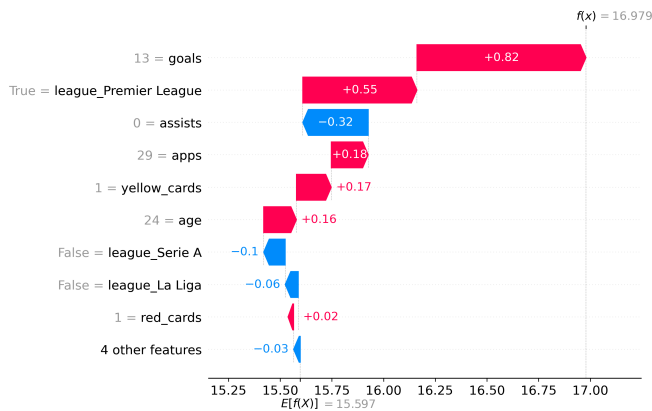


Figure. 4 Local SHAP waterfall plot, high-fee example (€46.5 million transfer). As in Figures 2 and 3, bars show how each feature contributed to this individual player’s prediction, illustrating how the balance of drivers shifts at the top end of the fee distribution.

player transfers. Overall, the model explains under a quarter of the variation in transfer fees on unseen data, meaning that the large majority of fee variation reflects factors not captured here, such as agent influence, contract terms, club specific negotiation dynamics, and commercial appeal.

To interpret the SHAP results, the pairwise correlation coefficients were calculated to assess the multicollinearity among

the numeric predictors. The strongest correlations were between appearances and goals ($r = 0.71$), appearances and yellow cards ($r = 0.64$), and goals and yellow cards ($r = 0.53$); all other correlations were less than $r = 0.40$. The appearances-goals correlation approaches, but does not exceed, the conventional $r = 0.8$ threshold typically used to flag serious multicollinearity. This is something to bear in mind when interpreting the relative rankings in Table 1, as SHAP is able to split importance between correlated features as opposed to crediting it to a single variable, although tree based models such as the Gradient Boosting Regressor are relatively robust to correlated predictors.

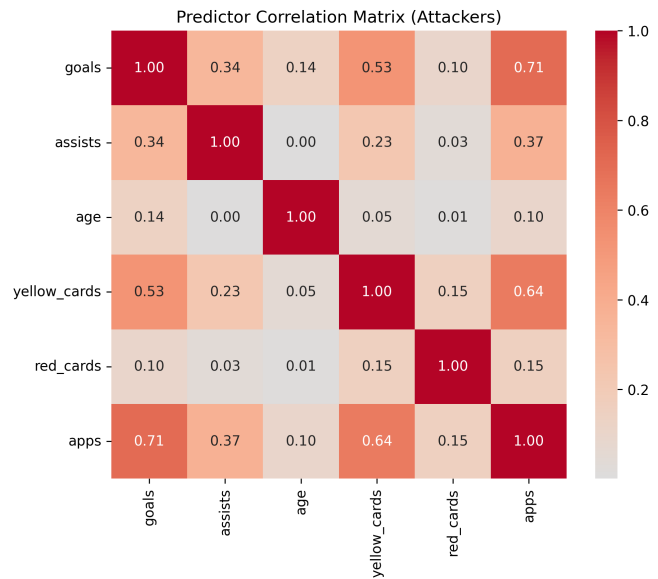


Figure. 5 Predictor correlation matrix (attacking players) Each cell reports the Pearson correlation coefficient between a pair of numeric predictors, with darker red indicating stronger positive correlation; readers should note the comparatively high correlation between appearances and goals discussed in the text.

To formally test for heteroscedasticity rather than relying on visual inspection of Figure 6 alone, a Breusch-Pagan test was conducted on the model residuals. The test returned an LM statistic of 81.49 with a p-value below 0.001, indicating statistically significant heteroscedasticity at the 5 percent level. This suggests that, despite the visually even scatter in Figure 6, prediction errors become less consistent at higher transfer

fees, and results for the most expensive transfers should be interpreted with additional caution.

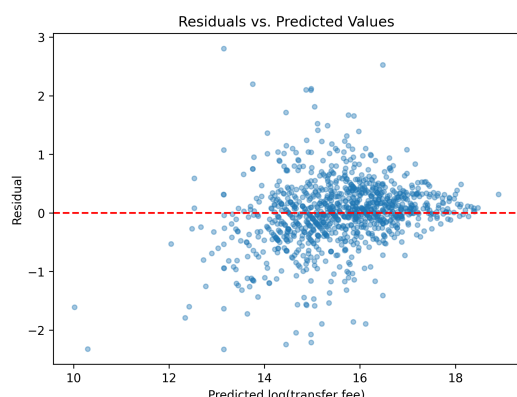


Figure. 6 Residuals vs. predicted values, cross-validated model. Each point is one player's residual (actual minus predicted log transfer fee) plotted against its predicted value; a roughly even, patternless scatter around the dashed zero line would indicate a well-specified model, while any funnel-shaped widening would indicate heteroscedasticity, formally tested in the text.

Discussion

It is worthwhile reemphasising a methodological point before moving on to a substantive interpretation: the SHAP values given throughout this study explain the way in which the trained Gradient Boosting Regressor makes use of each variable when producing predictions. These values measure the association and relative importance within the model, not causal effects that have been proven in real-world transfer pricing, and the discussion that follows should therefore be understood with this difference in mind.

The aim of this study was to find out what factors are linked to the transfer fees of attacking football players in the top five leagues in Europe from 2010 to 2025. The results indicate that direct goal scoring is the most significant positive factor affecting transfer fees, after that come assists, with age showing a strong negative relationship. The context of the league also plays a significant role: there is the highest positive fee premium for playing in the Premier League, followed by Serie A and La Liga, whereas Ligue 1 has a smaller premium and involvement in UEFA club competitions, although it is consistently in a positive direction, makes only a minor contribution to the overall impact. With regard to disciplinary records, the findings are mixed: as one would expect, red cards are linked to lower fees, but yellow cards show almost no consistent directional effect. Surprisingly, when goals and assists have already been taken into account, appearances are associated with lower predicted fees, probably because of the strong

correlation between appearances and goals rather than because there being a real penalty for the amount of playing time. Hypothesis H1 is therefore only partly supported since goals and assists act as expected but appearances do not. H2 is also only partly supported since red cards behave as expected but yellow cards have no meaningful effect. H3 is strongly supported since age shows both a negative direction and a considerable level of importance. H4 is supported in terms of direction, as all the league dummies including those for the UEFA competitions are positive compared to the reference category, but the domestic leagues, especially the Premier League, have a great deal more influence than the prestige of participating in continental competitions, going beyond what was originally anticipated by H4.

It is worth examining on its own account the low SHAP importance of the variables relating to UEFA competitions, even though they have a consistently positive effect, since a great deal of the current literature regards participation in top-level continental competition as a meaningful valuation signal in itself. A reasonable explanation for this is the overlap with the performance variables that are already included in the model: the players who take part in the Champions League or the Europa League come mostly from clubs which also achieve the highest numbers of goals and assists in this sample, so after taking into account goals, assists, and league context, the contribution of competition participation offers only a relatively small amount of new information for the model to use. Moreover, feature interactions may exacerbate this effect, because a tree-based model such as the Gradient Boosting Regressor can pick up part of the competition signal implicitly via combinations of league and performance variables rather than attributing it directly to the competition dummies.

The results confirm what economic theory would predict, but they also reveal the points at which the football transfer market differs from that theory. The importance attached to goals and assists is in line with Human Capital Theory¹, since players are assessed according to the productive output they contribute to a club. Yet the finding that league context—especially playing in the Premier League—has more influence than taking part in elite UEFA competitions, and the fact that appearances show a negative rather than a positive relationship once goals have been taken into account, indicate that the market for forwards does not always price observable characteristics in the way that would be expected². The negative effect of age and receiving a red card is in accordance with Bargaining Theory⁴, as clubs might use these factors as bargaining chips to lower the fees when negotiating. At the same time, the model only accounts for 0.226 (22.6 per cent) of the variation in transfer fees when tested using cross validation, which is considerably less than the 0.344 (34.4 per cent) obtained by the simple linear regression baseline, implying that more than half of the variation in fees is due to factors not included in

this analysis, such as agent involvement, club rivalries, media attention, or commercial branding³.

From a policy point of view, the findings have practical relevance for clubs, scouts, and regulators. Clubs that want to spend their money efficiently should concentrate on obtaining players who directly contribute to goals, rather than paying excessive amounts just to improve their reputation, appearances, or the prestige of the competition. Scouts could make use of SHAP-based models to spot players who are currently undervalued, as the analysis clearly indicates which attributes the market prizes. Nevertheless, the magnitude of the league premiums serves as a reminder that some of the effect appearing to be due to the league itself might in fact be acting as a proxy for unobserved factors such as the financial resources of the club, the level of media exposure it receives, and the extent of its marketing reach, rather than simply reflecting a true valuation of the competition. This study provides a predictive model of transfer fees and not an assessment of regulatory policy, so any suggestions it may have for governing bodies such as UEFA and FIFA must be treated as speculative and not as formal policy recommendations. With that qualification, the difference between players' measurable performance and the size of the fee premiums might merit further investigation by regulators who are concerned with transfer market transparency, although establishing this would require evidence that goes well beyond the scope of the current predictive analysis.

The study has a number of limitations. It makes use of secondary data from one source, the API-Football service, and the methodology employed by this service for collecting and validating the data is not completely available in the public domain. Other research involving data from different aggregated football data platforms, particularly the crowd-sourced valuations on Transfermarkt, shows that these valuations are strongly correlated with the actual transfer fees but tend to systematically underestimate the final price in the majority of transfers¹⁹, and the accuracy of the individual valuations can depend on the player's profile and the level of visibility in a league²⁰. This means that there is a kind of measurement error in the source that is hard to assess from outside the platform and affects the dependent variable itself, not just the independent variables, and therefore readers should regard the stated fees as estimates of what the market consensus is rather than as verified records of the transactions. The study does not take into account contract length, wages, agent fees, international caps, injury history, release clauses, or the player's pre-transfer market value and concentrates solely on attacking players, so the results might not be applicable to other positions. The SHAP analysis only reveals associations and relative importance rather than establishing strict causal effects, and the counterintuitive negative relationship found for appearances is probably due to multicollinearity with goals

rather than indicating a real causal effect; this problem could be dealt with in future work by means of feature selection or regularisation.

The transfer year was not included as a predictor, and all the fees have been analysed in nominal euro terms over the period from 2010 to 2025 without making any adjustments for inflation in the football market or for currency fluctuations during that time; this aspect still represents an unaddressed source of bias rather than a modelled one, and the year-on-year changes that occur in the transfer market are not separated from the player and league effects that are reported here. Since minutes played were not available in the source data, they have not been converted into per-90 rates nor have they been included as a predictor. It is left for future research to test the per-90 versions of the performance variables and to compare the different specifications with and without the inclusion of appearances, as this might help to determine whether the counterintuitive negative relationship with appearances is due to multicollinearity with goals or reflects a real effect. Height, which had been associated with a counterintuitive valuation pattern in an earlier version of this dataset, was not available in the API-Football data that was used for the final analysis; its omission in this case is therefore a deliberate result of the data source and not an error.

There are several avenues along which future research could build upon this study. One possibility would be to extend the model to take into account non-performance factors such as the number of followers on social media, the influence of agents, the length of the contract, and the player's commercial value, as this might help to account for the remaining variation in transfer fees. Another option would be to carry out similar SHAP-based studies on midfielders, defenders, and goalkeepers in order to compare how valuation varies according to position. Thirdly, future research could investigate whether these patterns remain valid during times of financial difficulty, for example during the COVID-19 pandemic, or under new financial regulations such as UEFA's Financial Sustainability Rules; it would also be useful to include the transfer year or season as a direct predictor in the model in order to separate out market-wide fee inflation from the player and league effects described in this study. Lastly, combining SHAP with time series analysis could show how the importance of the different variables has changed over the 15-year period, thus providing a more detailed understanding of how the football transfer market is developing.

References

- 1 G. S. Becker. Human capital: A theoretical and empirical analysis with special reference to education. National Bureau of Economic Research, 1964.
- 2 E. F. Fama. Efficient capital markets: A review of theory and empirical

- work. *The Journal of Finance*. Vol. 25, pg. 383–417, 1970, <https://doi.org/10.2307/2325486>.
- 3 T. Kirschstein, S. Liebscher. Assessing the market values of soccer players - a robust analysis of data from German 1. and 2. Bundesliga. *Journal of Applied Statistics*. Vol. 46, pg. 1336–1349, 2018, <https://doi.org/10.1080/02664763.2018.1540689>.
- 4 J. F. Nash. The bargaining problem. *Econometrica*. Vol. 18, pg. 155–162, 1950, <https://doi.org/10.2307/1907266>.
- 5 R. Besson, R. Poli, L. Ravenel. Econometric approach to assessing the transfer fees and values of professional football players. *Economies*. Vol. 10, pg. 1–14, 2021, <https://doi.org/10.3390/economies10010004>.
- 6 A. Metelski. Transfer market in football: Analyzing the top 100 most expensive players. *Humanities and Social Sciences Quarterly*. Vol. 31, pg. 109–120, 2024, <https://doi.org/10.7862/rz.2024.hss.22>.
- 7 B. Merten. The impact of transfer spending in expediting improvement of on-field performance of English Premier League clubs. *Scholar Commons*, 2021.
- 8 F. Jose, J. L. Felipe, A. Fernandez-Luna, P. Burillo, R. Luis, J. Sanchez-Sanchez, J. Garcia-Unanue. Money talks: Team variables and player positions that most influence the market value of professional male footballers in Europe. *Sustainability*. Vol. 12, pg. 3709, 2020, <https://doi.org/10.3390/su12093709>.
- 9 C. A. Depken, T. Globan. Football transfer fee premiums and Europe's big five. *Southern Economic Journal*. Vol. 87, No. 3, pg. 889–908, 2021, <https://doi.org/10.1002/soej.12471>.
- 10 G. Rubio-Martin, F. Gonzalez Sanchez, C. M. Manuel-Garcia, M. Cabezon Manchado. The effect of injury risk on players' value: Evidence from the main European Leagues. *Humanities and Social Sciences Communications*. Vol. 13, Article 223, 2026, <https://doi.org/10.1057/s41599-026-06511-w>.
- 11 M. Di Domizio, R. Caruso, B. Frick. The appraisal of players' transfer market values: Empirical evidence from Italian Serie A. *International Journal of Sport Finance*. Vol. 19, No. 1, pg. 39–51, 2024, <https://doi.org/10.32731/ijssf/191.022024.03>.
- 12 Y. Li, A. Ayanso, S. Yuan, M. Kusy, S. Kerwin. When interpretable machine learning meets the beautiful game: A predictive analytics approach to soccer player valuation in the transfer market. *Sport, Business and Management: An International Journal*. 2025, <https://doi.org/10.1108/SBM-05-2025-0108>.
- 13 I. G. McHale, B. Holmes. Estimating transfer fees of professional footballers using advanced performance metrics and machine learning. *European Journal of Operational Research*. Vol. 306, No. 1, pg. 389–399, 2023, <https://doi.org/10.1016/j.ejor.2022.06.033>.
- 14 D. Khalife, J. Yammine, E. Chbat, C. Zaki, N. Jabbour. Dynamic financial valuation of football players: A machine learning approach across career stages. *International Journal of Financial Studies*. Vol. 13, pg. 111, 2025, <https://doi.org/10.3390/ijfs13020111>.
- 15 J. Liu. Post-prime football player valuations: Depreciation difference between the English Premier League and the top European leagues. *International Journal of Financial Studies*. Vol. 13, No. 1, pg. 17, 2025, <https://doi.org/10.3390/ijfs13010017>.
- 16 D. F. Hill, J. Skinner, A. Grosman. A review of football player metrics and valuation methods: A typological framework of football player valuations. *Managing Sport and Leisure*. pg. 1–24, 2025, <https://doi.org/10.1080/23750472.2025.2459727>.
- 17 M. Franceschi, J.-F. Brocard, F. Follert, J.-J. Gouguet. Determinants of football players' valuation: A systematic review. *Journal of Economic Surveys*. Vol. 38, No. 3, pg. 577–600, 2024, <https://doi.org/10.1111/joes.12552>.
- 18 M. Franceschi, J.-F. Brocard, F. Follert, J.-J. Gouguet. Football players in light of economic value theory: Critical review and conceptualisation. *Managerial and Decision Economics*. Vol. 45, No. 2, pg. 896–920, 2024, <https://doi.org/10.1002/mde.4039>.
- 19 D. Coates, P. Parshakov. The wisdom of crowds and transfer market values. *European Journal of Operational Research*. Vol. 301, No. 2, pg. 523–534, 2022, <https://doi.org/10.1016/j.ejor.2021.10.046>.
- 20 T. Peeters. Testing the wisdom of crowds in the field: Transfermarkt valuations and international soccer results. *International Journal of Forecasting*. Vol. 34, No. 1, pg. 17–29, 2018, <https://doi.org/10.1016/j.ijforecast.2017.08.002>.
- 21 T. Sauer, C. Anagnostopoulos, H. Zulch, L. Werthmann. Creating value in football: Unveiling business activities and strategies of financial investors. *Managing Sport and Leisure*. pg. 1–21, 2024, <https://doi.org/10.1080/23750472.2024.2314568>.
- 22 L. Richau, F. Follert, M. Frenger, E. Emrich. The impact of investors on transfer fees in the English Premier League: A study of the ownership structures. *Corporate Ownership and Control*. Vol. 18, pg. 241–256, 2021, <https://doi.org/10.22495/cocv18i3siart2>.
- 23 S. Hoey, T. Peeters, F. Principe. The transfer system in European football: A pro-competitive no-poaching agreement? *International Journal of Industrial Organization*. Vol. 75, pg. 102695, 2021, <https://doi.org/10.1016/j.ijindorg.2020.102695>.
- 24 M. Uzik, R. Warias, J. Glova. Management of transfer prices in professional football as a function of fan numbers. *Mathematics*. Vol. 10, No. 16, pg. 2982, 2022, <https://doi.org/10.3390/math10162982>.
- 25 P. Malagon-Selma, A. Debon, J. Domenech. Measuring the popularity of football players with Google Trends. *PLOS ONE*. Vol. 18, No. 8, pg. e0289213, 2023, <https://doi.org/10.1371/journal.pone.0289213>.