International Journal of Science and Research (IJSR)
ISSN: 2319-7064
Impact Factor 2025: 7.089

What Factors Influence the Transfer Fees of Professional Football Players, and How Have They Been Modelled?

Reyahn Bantia

Abstract: ***This systematic review examines the factors influencing professional football transfer fees and the statistical methods used to model them. Six representative studies covering multiple linear regression, stepwise regression, machine learning, and systematic review approaches were synthesized in this review. Across the literature, player age, contract length, and performance metrics consistently emerged as the strongest predictors of transfer fees by having the most influence in determining the fee. Machine learning models generally demonstrated higher predictive accuracy than regression models but were used less frequently because of reduced interpretability. This review also identifies important research gaps, including the limited consideration of popularity, time-related variables, and non-European markets in the field as a whole. Overall, the findings of this review suggest that future research should combine broader predictor variables such as popularity and time with more diverse modeling approaches going beyond regression to improve transfer fee estimation.***

Keywords: Football transfer market; Player valuation; Transfer fees; Multiple linear regression; Machine learning; SHAP; Performance metrics; Predictive modeling.

1. Introduction

In 2025 alone, men's football clubs spent approximately $13.08 billion on transfer fees, the most ever in a calendar year. This shows the extent of the growth of the transfer market: it has become massive and financially significant. Accurately valuing players is the main part of the transfer market, and it has become increasingly more important as the market grows. Not only selling and buying clubs, but agents and financial institutions find player valuing essential for their jobs and success too. However, despite the size of the market, there is no universally agreed method for valuing players. Different researchers, agents, and clubs have used different variables and models to value players, meaning the value of a player can differ between stakeholders based on how they calculated it. All of these stakeholders require accurate models, but the literature is fragmented, making it harder to interpret.

This paper aims to synthesize findings across literature regarding the different types of models and variables used, and use it to identify the major determinants of player transfer fees. To achieve that, this review will review 6 different papers, covering both regression models and machine learning models, each of which use different variables with different weights. After this, it will summarize all the variables identified, and conclude which are the most common across regression and machine learning, answering the question: "What factors influence the transfer fees of professional football players, and how have they been modelled." The body is divided into three sections, each covering a type of model.

Papers were selected according to a defined set of inclusion and exclusion criteria. To identify the papers, search terms including "player valuation models" and "transfer fees" were searched on MDPI, ResearchGate, and ScienceDirect. To be included in this review, the papers had to be published in English between 2018 and 2026, and be focused on using a quantifiable modelling approach to model professional football transfer fees. Papers were excluded if they primarily focused on non-European leagues or did not focus on transfer fees. The final 6 papers were chosen because they included a combination of two modelling methods: regression and machine learning, as well as one systematic review of the entire field.

2. Literature Review

Section 1: Econometric and Regression Based Approaches
There are multiple different methods of modeling transfer fees, but Multiple Linear Regression (MLR) is the most dominant one in recent years. MLR allows many different variables to be included when determining the optimal transfer fee for a player, which will be explored further in this section. In this section, I will be covering three different research papers that all have used MLR to model transfer fees. Each paper introduces new variables or model related adjustments, however, with different sample sizes and new variables in each of their models. The models build upon each other, starting off with a European-centric model, moving on to a global model, and finally, a model which implements variables that push the boundaries of regression.

MLR is a technique used in statistics to model the impact of two or more independent (predictor) variables on a single, dependent variable. It fits a line of best fit onto the data, and predicts the dependent variable based on how the predictor variables vary. When dealing with player transfer fees and valuations, the fee (primarily expressed in millions of euros) is the dependent variable, and predictor variables can include things like player age, size of the club they play for, or even popularity. MLR is flexible because different models can utilize different predictor variables, and weigh them differently based on their sample size or scope.

"Econometric Approach to Assessing the Transfer Fees and Values of Professional Football Players" (Poli et al., 2022) categorizes predictor variables into three different archetypes: club related, player related, and the season the transfer took place. In this model, club related variables refer to the competitive level of the player's club in the two years preceding their transfer: the level of their club compared to the top clubs in the world. Clubs are ranked based on the leagues they compete in, with the English Premier League being the highest level. Presence in other competitions, such as UEFA Champions League, are also looked into when evaluating the level of a club. Player related variables are more numerical, such as time remaining on contract, age, position (goalkeeper, defender, midfielder, forward), and minutes played. The season the transfer took place was included in this model to allow the researchers to see the evolution and inflation of prices over time.

The MLR model created in the study by (Poli et al., 2022) had a sample size of 2045 transfers from July 2012 to March 2021 in the top 5 European leagues exclusively (English Premier League, Spanish LaLiga, Italian Serie A, German Bundesliga, and French Ligue 1). It uses 12 distinct variables, each falling into one of the three categories established above. One of the key findings was that contract length remaining, age, and top level experience are the most important variables when determining transfer fees. The model shows that 1 extra year remaining on the contract of a player with 2 years already remaining increases transfer fees paid by other clubs by 22%, and 1 year less in age increases the transfer fee of a player by around 12% if all other factors are kept the same.

This study establishes a foundation for how MLR models work to predict football transfer fees, demonstrating that weighted predictor variables are effective to come to a conclusion on how a player would be valued. However, the model's European-only scope limits the generalizations that can be made by the data it collected. The following two examples build on this foundation by introducing new variables, methodical refinements, and larger scopes.

Building on their 2022 framework, "Statistical Modeling of Football Players' Transfer Fees Worldwide" (Poli, et al., 2024) extends the previous study using advances in statistical modeling and new variables. It introduced 4 new predictor variables that were not present in (Poli, et al., 2022). These new variables included average sporting level of matches played by the footballer over the 365 days preceding the transfer as a single variable, the number of minutes played by the footballer in official international competition games in the year preceding the transfer, the average number of points obtained by the player's teams (two points for a win, one for a draw and zero for a loss) in all official matches played by the footballer during the 365 days preceding the transfer, and percentage of minutes played by the footballer as a starter in official club or national team matches in the 365 days preceding the transfer.

These new variables are simplified as compared to the variables in the previous study, and they put more focus on the players being on the pitch and gaining game experience for their team, as opposed to focus on the fact that the players are a part of the team. This study also expanded the sample size and scope of the model, as a list of 8389 transfers from July 2014 to March 2024 were used to create the MLR model, confirming that the MLR statistical modeling approach is not limited to small scales.

Despite the strengths, both models carry significant limitations. While most of the transfer fees given by the model are robust, there are still some cases where the fee is under- or overestimated. Willem Geubbels' transfer fee from Olympique Lyonnais to AS Monaco was much higher than the model had predicted. This was because Geubbels was a young player (17 years old) who had shone at the youth level so much that he was compared to Kylian Mbappé who moved from Monaco to PSG in the same window for 180 million euros (second most expensive transfer of all time). However, the model underestimated this price because Geubbels had very little experience at the professional level, and the majority of the transactions on which the model was built involved slightly older players with at least a year of professional experience. Mario Balotelli's transfer from Milan to Liverpool in 2014 represents the reverse case, as the fee was around half of the estimated value. This decrease in fee was likely because of Balotelli's disciplinary and behavioral problems. Player behavior is a qualitative variable and thus cannot be quantified, so its impact on transfer fees cannot be accurately captured by regression models. Both (Poli, et al., 2022) and (Poli, et al., 2024) share these weaknesses, highlighting that regression models rely on availability of quantifiable data, and cannot account for intangible qualities that clubs are looking for.

It seems like all regression models aimed at valuing player transfer fees will yield similar results, even if slight changes are made to the variables. This is because while the variables themselves are changed, the categories of variables (such as player age, amount of football played, and level of club) still remain the same. However, a study titled "Determinants of Transfers Fees: Evidence from the Five Major European Football Leagues" (Ante, 2019) challenges this by adding player popularity as a predictor variable. The study also used stepwise regression with backwards elimination in the model, with a smaller sample of 389 football player transfers in the 2018/19 summer transfer window of the five major European football leagues.

Stepwise regressions with backwards elimination is a method of regression that starts off with a regression model with all the existing variables. First, the variable with the lowest explanatory value is eliminated, and the regression is run again. This process repeats, and the variables are eliminated in a step-wise manner until all of the variables reach a specific value of explanatory value. The model shown in this research paper uses an elimination measure of $p \geq 0.2$.

According to the study, the popularity of a player provides football clubs with indicators on how a particular player may have an effect on jersey or ticket sales. It puts emphasis on superstars, for whom talent is only half of their value. Clubs may be willing to spend more on such popular players, because of certain leadership qualities, or even because of the value they bring to fans and to the sport as a whole. This model therefore introduced a completely new variable to the regression model, popularity, which was calculated as the sum of the natural logarithms of Instagram and Twitter followers and Facebook likes.

Taken together, the three studies reviewed in this section – (Poli et al., 2022), (Poli et al., 2024), (Ante, 2019) –

collectively show that regression is a powerful tool to model player transfer fees. Regression is effective across different sample sizes and scopes as well. It was used for a model with 389 samples from only the top 5 leagues, as well as a model with 8389 samples globally.

These studies also show that the choice of variables when valuing players is not settled, however. The predictor variables used in each study are slightly different, with some variables being modified (such as international caps vs international minutes), and others being added (such as player popularity). In all of the studies, all of the quantifiable variables, such as player age, goals, and assists are consistent. However, variables that are hard to quantify, such as player behavior, popularity, and superstardom are not consistent, and each study uses these variables differently, creating complications. These limitations, consistent across each of the regression models, raise questions about whether something different to regression could do better in modeling player transfer fees.

Section 2: Machine Learning Based Approaches

Not as popular as linear regression, machine learning offers an alternative approach to MLR in modeling player transfer fees. As discussed earlier, regression has several limitations such as researcher-determined variables, as well as difficulty in modeling qualitative variables such as behavior. In machine learning models, instead of the researcher deciding which variables matter the most, the model learns patterns from the data and calculates variable weights based on previous trends itself. In this section, I will be covering two more research papers, this time specific to transfer fee models made with machine learning. It will address two key questions: how fees can be predicted using machine learning models, and why machine learning models value players in the way that they do.

Machine Learning is a branch of artificial intelligence. Computers are provided with a list of data, on any topic, and instead of being programmed to interpret the data, they recognize patterns and make decisions based on the data themselves. Machine learning is the opposite of traditional programming. Instead of writing step-by-step instructions for the computer to follow, the model is provided with data, and it independently identifies the rules and trends. In determining transfer fees, a computer would be given data of transfers within a certain time period and scope, along with the transfer cost and player information. Based on the data it is fed, it would predict the costs of players with different attributes in future transfers.

“Estimating transfer fees of professional footballers using advanced performance metrics and machine learning” (McHale & Holmes, 2023) is a research paper that uses machine learning to value transfer fees. The study’s final data set had information on 1946 transfers between 11th August 2016 and 29th September 2020. (McHale & Holmes, 2023) critique previously discussed models because simple metrics such as goals and assists (that are used in the linear regression models) fail to accurately assess the playing ability of a player, which is the most important component in assessing a transfer fee. Goals scored alone fail to capture a player’s actions on the pitch, such as important tackles, chance creation, fouls won, or creativity. Because of this, the study used two specific advanced performance metrics: Plus-Minus Ratings and Goal Impact Metrics.

Much of the data fed into the computer for this study revolves around the two advanced performance metrics and crowd sourced, subjective ratings from sofifa.com and transfermarkt. Firstly, McHale and Holmes completely avoided transfermarkt player transfer valuations in their study because only a small proportion of the valued players actually get transferred, meaning that the valuations do not take into account the fact that a club is actually signing the said player and the player is actually completing a move. Sofifa ratings were a different case, as the study utilized them. Sofifa ratings are subjective, and they offer ratings for things like passing, shooting, and heading. The study uses two specific ratings from sofifa, which are overall rating (estimate of player current ability) and potential rating (estimate of player future ability) for players. Plus-Minus ratings are also simple: essentially, it compares how a team performs with a certain player as opposed to without that player. Goal Impact Metrics are more complex, but in short, they present ratings based on whether a player’s action increases (in the case of attackers) or decreases (in the case of defenders) the probability of a possession ending in a goal. Combined, these metrics and ratings form the core of the machine learning model used to predict transfer fees, which shows that the study is trying to move away from less meaningful numbers such as goals and assists, and move towards capturing a player’s influence on the game.

The model combines the above metrics, as well as some other simple statistics such as contract status, player age, and player position. The researchers split their data by time, using 80% of it for training (1557 transfers), and the remaining 20% for testing (389 transfers) to avoid data leakage. Three different machine learning algorithms are used in the study: glmnet, xgbTree, and xgbDART. The technical details of each algorithm are not central to this review, but the xgbTree and xgbDART models are the best fitting, as they achieve higher R^2 values (0.77 and 0.76 respectively) than other regression models they were tested against. This finding shows a significant gain in predictive accuracy over regression.

Having addressed how machine learning models predict player transfer fees, this section now turns to the second key question: why models value players in the way that they do. “When interpretable machine learning meets the beautiful game: a predictive analytics approach to soccer player valuation in the transfer market” (Li, et al. 2026) seeks to answer this question. It is comparable to the previous research paper (McHale & Holmes) because of the similar scope (both cover big five leagues) and time frame (both overlap between 2017 to 2020).

Before analyzing how the machine learning models work, the study first compares certain variables and determines the most important variables. For example, the study identified that certain sport-general skills (e.g. composure, aggression, reaction) are more valuable to the model than football-specific skills such as dribbling. Football-specific skills are also overshadowed by other player-specific factors including age and team rating. Machine learning models therefore tend to value sport-general psychological and physiological attributes more than football specific skills.

In order to explain the results of the model (how it predicts fees), the study utilizes SHAP (SHapely Additive Explanations). A common pitfall of machine learning models is that despite generating accurate predictions, they do not explain the reasoning behind the prediction. This leaves the user with many possibilities: was the player valued high because they were young, because they played for a strong club, or because their composure rating is high? SHAP solves this problem by calculating how much each variable increased or decreased the transfer fee. For instance, there is a player valued at €50 Million and the average is €20 Million. SHAP calculates the proportion of each variables' impact in the €30 Million difference. Put simply, SHAP assigns a magnitude to each variable for each prediction, revealing which factors contributed to the player being valued as they were. This is explored by Li et al. (2026) in various machine learning models.

In this section, I have analyzed two studies that aim to value players using machine learning models. Machine learning models improve the accuracy of the regression models, as shown in "Estimating transfer fees of professional footballers using advanced performance metrics and machine learning" (McHale & Holmes, 2023). This study (and other machine learning models) move beyond basic statistics such as goals scored which are dominant in MLR models, and instead consider advanced performance metrics that accurately assess a player's playing ability, such as Plus-Minus ratings. Additionally, machine learning models are centered around the computer interpreting the data for itself, so it decides the weights of each of the quantifiable variables, adding consistency based on data which was not present in MLR models.

A parallel between regression and machine learning is that age and contract length are the most important factors to consider when modeling the value of a player. This shows that while machine learning is a step up in accuracy, it still converges with regression findings in certain ways. Despite this, the use of crowd-sourced ratings such as sofifa ratings in machine learning models are subjective, so any machine learning model's data depends partly on the data which can have its own biases in valuing a player. Due to this, both regression and machine learning can struggle to model a player's value due to missing variables or data bias. These limitations suggest that neither regression nor machine learning have achieved a complete evaluation of a player's transfer fee, and that the field as a whole may benefit from broader variable sets and modeling approaches.

Section 3: Synthesis and Gaps

Having examined regression and machine learning models individually, this section uses Franceschi et al. (2024) to evaluate the broader picture of player valuation research to identify gaps in the field. "Determinants of football players' valuation: A systematic review" (Franceschi et al., 2024) addresses this, as it is a systematic review of 29 different papers on football player transfer fee modelling methods. The study reviewed 111 trained models across the literature, giving their conclusions credibility. This section aims to use the conclusions from (Franceschi et al., 2024) to review what the player valuation field as a whole has found, as well as gaps that are still present.

Table 1 summarizes the 6 papers reviewed in this paper across methodological and analytical factors:

As table 1 illustrates, performance based factors and regression models dominate the field, despite the widely varying scope and method.

	Poli et al. (2022)	Poli et al. (2024)	Ante (2019)	McHale & Holmes (2023)	Li et al. (2026)	Francheschi et al. (2024)
Modeling Method	MLR	MLR	Stepwise Regression	Machine Learning	Machine Learning	Literature Review
Sample Size	2045 Transfers	8389 Transfers	389 Transfers	1946 Transfers	831 Transfers	29 Papers, 111 Models
Scope	2012-2021, Top 5 Leagues	2014-2024, All Leagues	2018-2019, Top 5 Leagues	2016-2020, Top 5 Leagues	2017-2020, Top 5 Leagues	1990s-2022Global
Key Variables	Age, Contract Length	Sporting Level of Matches and Team	Player Popularity	Goal Impact Metric, Plus-Minus Ratings	Why Each Variable is Used	Performance Metrics, Time, Popularity
Explanatory Power	$R^2 = 0.80$	$R^2 = 0.85$	R^2 = N/A	$R^2 = 0.77$	R^2 = N/A	R^2 = N/A
Key Finding	Regression is an accurate method for predicting transfer fees	Regression approach holds at a global scale.	Social Media popularity is a significant variable	Machine learning models outperform regression in accuracy	Sport general skills outweigh football specific skills	Performance variables dominate literature, time and popularity are not represented

Franceschi et al. (2024) divides independent variables into 6 different categories: Popularity, Player Characteristics, Club Characteristics, Performance, Labour, and Time. According to the review, Performance related variables are the most common category in all of the papers they analyzed. This means that most of the models used to value player transfers use some kind of performance metric, with some even using advanced performance metrics. This finding aligns with the papers I have reviewed, as all of them utilize performance based metrics in some form. For example, Poli et al. (2022) and Poli et al. (2024), two of the MLR models, utilize goals scored and assists in their valuations, which is a basic performance metric. The machine learning models also use performance metrics, but they use more advanced ones. McHale & Holmes (2023) uses Goal Impact Metrics and Plus-Minus Ratings, which are more detailed and better for evaluating a player's impact on the pitch. Overall, a parallel between all of the studies I have reviewed (and the studies reviewed by Franceschi et al., 2024) is that performance based metrics and variables are the most common and widely utilized in valuing football player transfer fees.

On the other hand, literature as a whole has neglected or ignored several variables. According to Franceschi et al. (2024), seemingly important variables including time, popularity, and labour were underrepresented in the 29 papers reviewed. However, popularity and time were two variables that were demonstrably underrepresented in the papers reviewed in this study. Social media popularity appeared in Ante (2019), where number of followers of a player were used in the regression model to predict fees, but the popularity or "superstar" factor did not appear in any of the other studies. Popularity remains an important factor, however, because it influences why transfers such as that of Cristiano Ronaldo to Juventus in 2018 went with a €100 million price tag despite the player being 33 years old.

Time seems like an abstract variable, but it is relevant in valuing player transfers as much as popularity. There are only 3 variables present across the 29 papers reviewed by (Franceschi et al. 2024) that account for time passing. Time is an important variable because as time goes by, the value of money decreases and players go for much higher fees. For example, the highest transfer fee of all time before 2005 was Zinedine Zidane's €77.5 million move from Juventus to Real Madrid. Now, 21 years later, the highest transfer fee of all time is €222 million, showing the rapid evolution of transfer fees over time. Despite the massive impact of time on transfer fees, in the papers I have reviewed, only 1, (Poli et al. 2022), accounts for it, showing its underrepresentation similar to popularity.

Another finding concerning the field of football player values is that a vast majority (85% of the 111 models evaluated by Franceschi et al., 2024) of the models use regression, commonly ordinary least squares (OLS). This shows that despite the improvements in accuracy shown by machine learning models, as demonstrated in section 2, the field has been slow in moving away from regression. This pattern suggests that regression models are easier to build and interpret. Widespread use of regression models despite machine learning models being shown to have higher degrees of accuracy also implies that the field as a whole is prioritizing ease of usage and interpretation of their models over predictive accuracy.

Taken together, the findings of Franceschi et al. (2024) and the papers reviewed in this study identified that across all papers, a small set of variables including age, contract length, and performance consistently appear as the most "valuable" when determining transfer fees. This is the case in both the regression and machine learning models I reviewed. Evident in recent high-profile transfers, the players in question generally are either young, have come off good seasons, or both. For example, the €145 million transfer of Alexander Isak to Liverpool followed a season in which Isak scored 23 goals in the Premier League season, leading to a fifth place finish for his current team, Newcastle United. Isak was also relatively young at 25, and had 3 years left on his Newcastle contract. This transfer itself shows how important age and performance are in valuing players, suggesting that existing models have accurately identified the most influential variables. Despite this, there are various variables such as popularity of players and labour and models that diverge from regression that remain unexplored due to over-reliance on regression models and performance metrics. The evidence reviewed across the 6 studies therefore suggests that future research would benefit from broader variables, diverse modeling methods, and focus on underrepresented categories such as time and popularity.

3. Discussion

Regression emerges as the most popular and widely used method of modeling player transfer fees, with around 85% of all models using it. According to the studies based on regression models, they explain approximately 80% of fee variation. This means that 80% of the differences between player fees can be explained by the regression models. The evidence reviewed demonstrates that machine learning outperforms regression in accuracy, shown in McHale & Holmes xgbTree R^2 vs regression R^2. Despite this clear finding, the field has not shifted towards machine learning, meaning that data interpretability is being prioritized. I think that regression being dominant despite lower accuracy shows that researchers are more inclined towards producing models with easier to interpret findings and data instead of accuracy.

Across all the models, machine learning and regression, a recurring limitation emerges. Both the regression and machine learning models struggled to value players when certain, unquantifiable variables were involved. Across studies, these included player behavior, superstardom, and young potential. Balotelli's fee was overestimated by the model because it failed to account for the player's behavioral issues and unprofessionalism. Ronaldo's fee was underestimated because the model did not take into account his superstar quality and extraordinary popularity with both fans and other players. Geubbels' fee was underestimated because he was a young player with an excellent record at the youth level: the model only accounted for games at the professional level, and could not see potential in a young player like a scout could. These three cases sum up the limitations in all player transfer fee models: they struggle when faced with intangible and unquantifiable aspects to a player, which are taken into consideration by buying and selling clubs when valuing a player.

Regarding the factors that influence the transfer fees of professional football players, the models correctly identified the most influential ones. According to the studies examined in this paper, objective player and club related variables including player age, player contract length, player position, and sporting level of selling club were the highest weighted and had the most influence in the models' predictions of value. In the transfer market, these are indeed the most sought-after attributes of football players, causing clubs to pay more for younger players playing for better clubs. The papers I reviewed align with this, even if they struggle to integrate subjective variables. Because of this, prediction models in this field are better used as tools, instead of definite answers. The models are accurate most of the time, but cannot account for certain human elements no matter how they are programmed.

Overall, the most important factors that have not been studied well in the 6 papers I have reviewed are player popularity and time (as identified in section 3). Factors such as superstardom, social media following, and behavior attributes all come under popularity. These are the exact variables that models have a hard time modeling, stated in previous paragraphs. It is because these factors havae not been studied as well by researchers relative to performance based variables. To improve accuracy, future researchers could dedicate more time towards experimenting with modeling player popularity, so superstardom and following would be accurately modelled, potentially reducing the large prediction errors observed in the cases of Ronaldo to Juventus and Balotelli to Liverpool.

4. Conclusion

This review began by examining 3 papers centered towards modeling player values using MLR models. Current research shows that regression models are currently dominant in the field due to ease of interpretation. They are also accurate, accounting for 80% of the factors influencing a transfer fee on average. After this, the review shifted focus to machine learning models. Research shows that they are more accurate than regression models, and are also more explainable due to SHAP. However, they are not dominant because they are difficult to interpret and rely on subjective performance metrics. After reviewing regression and machine learning, Franceschi et al. (2024) was used to broaden the perspective and to objectively view the gaps of the field as a whole.

The evidence collected in this review points to a clear answer to the research question: What factors influence the transfer fees of professional football players, and how have they been modelled? The factors most consistently identified as highly influential across the papers were player age (years), player contract length (years), and rating using performance metrics. Additionally, advanced performance metrics such as Goal Impact Metrics demonstrated stronger predictive value than basic statistics such as goals scored. The dominant method of modeling is regression, but machine learning is also used by some studies. Player transfer fee models have significantly improved over time, but as stated previously, no model covers every single quantifiable and non-quantifiable variable that influences player transfer fees, meaning that models are better used as tools instead of validations.

While this review has addressed the research question, several limitations remain. Firstly, it only covered data from the top 5 European leagues. It did not cover any transfer data from smaller leagues, lower divisions, and non-European markets. This makes the findings of this review specific and applicable only to European top 5 leagues instead of the football transfer market as a whole. Additionally, it did not cover women's football, which has its own transfer market, vastly different from the men's. Lastly, the review only covered 6 research papers across regression, machine learning, and field gaps. Given the amount of existing literature, 6 papers offer a partial picture. Reviewing more papers on each type of model would have uncovered more information and more data in order to draw parallels and answer the research question with more conviction.

This review demonstrates that player age, contract length, and performance metrics are the most consistent determinants of professional football transfer fees across existing studies. Multiple linear regression remains the dominant modeling approach because of its interpretability, while machine learning methods generally achieve higher predictive accuracy and provide promising alternatives. Despite these advances, current models still struggle to capture subjective factors such as player popularity, behavior, and perceived potential. Future research should expand beyond regression models, incorporate underrepresented variables, and examine transfer markets outside the major European leagues to produce more comprehensive and generalizable valuation models.

References

[1] Ante, Lennart. "Determinants of Transfer Fees: Evidence from the Five Major European Football Leagues." *SSRN Electronic Journal* (2019). https://doi.org/10.2139/ssrn.3433529

[2] Franceschi, Maxence, Jean-François Brocard, Florian Follert, and Jean-Jacques Gouguet. "Determinants of Football Players' Valuation: A Systematic Review." *Journal of Economic Surveys* 38, no. 3 (2024): 577–600. https://doi.org/10.1111/joes.12552

[3] Hastie, Trevor, Robert Tibshirani, and Jerome Friedman. *The Elements of Statistical Learning*. 2nd ed. New York: Springer, 2009.

[4] Li, Shaobo, Anteneh Ayanso, and Kaveepan Lertwachara. "When Interpretable Machine Learning Meets the Beautiful Game: A Predictive Analytics Approach to Soccer Player Valuation in the Transfer Market." *Sport, Business and Management* 16, no. 2 (2026): 240. https://doi.org/10.1108/SBM-05-2025-0108

[5] McHale, Ian G., and Benjamin Holmes. "Estimating Transfer Fees of Professional Footballers Using Advanced Performance Metrics and Machine Learning." *European Journal of Operational Research* 306, no. 2 (2023): 687–698. https://doi.org/10.1016/j.ejor.2022.06.033

[6] Montgomery, Douglas C., Elizabeth A. Peck, and G. Geoffrey Vining. *Introduction to Linear Regression Analysis.* 5th ed. Hoboken: Wiley, 2012.

[7] Poli, Raffaele, Roger Besson, and Loïc Ravenel. "Econometric Approach to Assessing the Transfer Fees and Values of Professional Football Players." *Economies* 10, no. 1 (2022): 4. https://doi.org/10.3390/economies10010004

[8] Poli, Raffaele, Roger Besson, and Loïc Ravenel. "Statistical Modeling of Football Players' Transfer Fees Worldwide." *International Journal of Financial Studies* 12, no. 3 (2024): 93. https://doi.org/10.3390/ijfs12030093