Dean&Francis
ISSN 2959-6157

Linear Regression Analysis of Football Player Market Value Fluctuation Factors

Shangmei Yang

Beijing AIDI School, Beijing, China
yangshangmei@aidi.edu.cn

Abstract:
This study employs linear regression analysis to examine the factors influencing football player value fluctuations based on market value data from 44 football players. Through constructing a multiple linear regression model, the research findings reveal: Each unit increase in event intensity leads to an average increase of 324.5 percentage points in value change rate; The Big Five leagues bring an additional 298.7 percentage points of value growth compared to other leagues; Each unit increase in player level results in an average decrease of 156.3 percentage points in value change rate; The linear regression model effectively explains the value fluctuation phenomenon (R^2=0.612). This research provides crucial quantitative insights into football's economic dynamics by identifying key factors affecting player valuation. The findings offer evidence-based guidance for stakeholders in player development, investment strategies, and market regulation, contributing to more efficient resource allocation and strategic decision-making across the football industry. Additionally, the study establishes a methodological framework for analyzing value volatility that can be applied to talent evaluation across sports markets, while enhancing transparency in the increasingly commercialized football transfer ecosystem.

Keywords: Football players; Market value; Transfer market

1. Introduction

As the world's most popular sport, football holds a dominant position in terms of commercial value within the global sports industry. With the increasing commercialization of football, players' market value—including transfer fees, salary levels, and commercial endorsement value—has become an important indicator for measuring their comprehensive strength. However, these values exhibit significant volatility: emerging players may see their value skyrocket tenfold after breakthrough performances, but they may also experience sharp depreciation due to injuries, tactical maladjustment, or off-field issues.

In recent years, research on football player value assessment has increasingly attracted academic at-

ISSN 2959-6157

tention. Various methodological approaches have been developed to understand and predict these values. Yiğit et al. established a football player value assessment model using machine learning techniques based on individual player statistics from over 5,000 players across 11 major leagues. Their study employed advanced supervised learning techniques such as ridge and lasso regressions, random forests, and extreme gradient boosting to support transfer decisions of football clubs. Aznar and Estruch pioneered the application of the Analytic Hierarchy Process (AHP), a multicriteria methodology, to football player valuation [1]. Their approach addressed situations where comparable data is scarce and emphasized the importance of transparent, traceable valuation processes given the significant economic impact of football.

Beyond individual valuation methods, Xarles et al. surveyed the broader field of Action Valuation in sports analytics, introducing a taxonomy with nine dimensions related to data, methodological approaches, evaluation techniques, and practical applications [1]. Their work highlights the increasing importance of assigning value to individual actions based on their contribution to desired outcomes.

The impact of developmental factors on player valuation has been examined by Balliauw et al., who quantified the significant positive impact of youth academy quality on a player's future market value [2]. Through multiple regression analysis of 94 players trained in 13 different academies, they demonstrated that higher quality youth academies lead to better-trained players with higher market values, which correlates with higher future wages and transfer fees.

Tiscini explored the disconnect between football clubs' generally negative financial results and their positive market valuations [1]. Their empirical analysis revealed that a club's value cannot be estimated solely on financial results but requires consideration of overall shareholder benefits, including private control benefits and socio-emotional factors.

The commercial aspect of player valuation was addressed by Su et al., who studied factors influencing athletes' social media following, particularly during team transfer periods [3]. Their research identified the importance of considering the joint influences of related brands at different levels (league, team, and athlete) for understanding how athlete brands develop, which increasingly affects player market value.

Jayawardhana analyzed the influence of education on football clubs' performance, treating clubs as business entities with measured performance indicators, cost efficiency, and education as variables [2]. The study found evidence that both education and club expenditure significantly impact sports performance, though certain educational factors can lead to performance decline.

Recent methodological advances include a multiple linear regression model for estimating attacking football players' market value [4]. Using cross-sectional data from 105 Premier League attackers, Lorincz introduced two novel variables: commercial potential and nationality based on homegrown rules. The resulting model achieved an R^2 of 0.65, outperforming previous academic models both in-sample and out-of-sample. Hill et al.proposed a comprehensive framework for football player valuations by integrating Damodaran's valuation typologies (intrinsic, relative, real options, and probabilistic methods) and providing a decision framework for selecting the appropriate approach. Their work addressed the limitations of using proxies and the conflation between price and value in football player valuations.

From a business perspective, Bács examined how player transfers affect the market valuation of publicly traded football clubs, finding that successful transfer policies positively impact a club's market value regardless of transfer balance [5].

Although related research is relatively abundant, existing studies primarily lack quantitative analysis of the linear relationships between various influencing factors and value changes in terms of methodology. To fill this research gap, this study specifically adopts linear regression analysis methods to systematically identify the key linear driving factors of value fluctuations through value volatility data from 44 football players.

2. Methods

This study adopts multiple linear regression analysis as the core research method, based on linear relationship assumptions, to systematically analyze the linear effects of various influencing factors on football player value changes. The choice of linear regression analysis is based on the following considerations. (1) Ability to quantify marginal effects of each factor; (2) Strong model interpretability with intuitive economic meaning of coefficients; (3) Suitability for identifying linear relationships among multiple variables.

2.1 Data Sources and Sample Construction

Table 1 presents representative cases of player market value changes among the research samples. The sample data is sourced from the Kaggle football database platform, covering market value data of 44 football players. The sample composition is as follows: the distribution by competitive level includes 15 world-class players, 20 top-tier European players, and 9 ordinary professional players; the geographical distribution covers 35 players from the five major European leagues and 9 players from other leagues; the situation of value changes consists of 22 cases of negative growth and 22 cases of positive growth. All value data have been standardized to US dollars using the World Bank's annual average exchange rate, which is intended to ensure the comparability of observations in the linear regression analysis.

Table 1. Example of Sample Data Structure

Player	Low/Old Valuation	Peak/New Valuation	Increase	Main Cause	Time Period	Source/Kaggle
Erling Haaland	€20M	€180M	800%	Dortmund→Man City goal spree	2019→2023	kaggle
…………	…………	…………	………..	…………	…………	…………
André Silva	€40M	€12M	-70%	Bundesliga→La Liga adaptation failure	2021→2023	kaggle
Ferran Torres	€55M	€25M	-55%	Barcelona's goal drought	2022→2023	kaggle
Richarlison	€58M	€25M	-57%	Tottenham's opening goal drought	2022→2023	kaggle

This study collected market value data from 44 football players to construct the research sample.

The sample includes 15 world-class players, 20 European first-tier players, and 9 general professional players in terms of competitive level, with geographic distribution covering 35 players from Europe's Big Five leagues and 9 from other leagues, while professional status comprises 22 negative cases and 22 positive cases. Data sourced from the Kaggle football database platform [6]. Value data standardized to USD using World Bank annual average exchange rates to ensure comparability of observations in linear regression analysis.

2.2 Variable Design

Table 2 provides a detailed definition of all variables used in the multiple linear regression model and their statistical characteristics; the dependent variable, Value Change Rate (VCR), adopts the relative change rate calculation method, which can eliminate the impact of differences in the initial values of different players. The key explanatory variables include event intensity, league scale, and player level, and the design of these variables is based on football economics theory and market practical experience; the introduction of control variables is intended to control other factors that may affect changes in player value, thereby improving the statistical reliability of the model. The coding method of variables follows econometric standards, which is convenient for coefficient interpretation and policy implication analysis.

Table 2. Variable Definition and Value Description

Type	Variable Name	Variable Symbol	Variable Definition	Value Description	Sample Distribution/Statistical Characteristics
Dependent Variable	Value Change Rate	VCR	Relative change in player market value	VCR = [(Final valuation - Initial valuation) / Initial valuation] × 100%	Observed Range: -70% ~ +800% Positive cases: 22 (50%) Negative cases: 22 (50%)
Key Explanatory Variable	Event Intensity	Event_ Intensity	Level of player participation in major events	World Cup/Champions League breakthrough=3; League championship/major transfer=2; Others=1	Values: 1, 2, 3 (three levels) Regression coefficient: 324.5
Key Explanatory Variable	League Scale	League_ Scale	The commercial level of the league where the player plays	Big Five leagues=3; Secondary leagues=2; Other leagues=1	Big Five leagues: 35 (79.5%) Other leagues: 9 (20.5%) Regression coefficient: 298.7

ISSN 2959-6157

Continue Table 2

Key Explanatory Variable	Player Level	Player_Level	Player's status in the transfer market	Logarithmic value based on transfer market ranking	World-class: 15 (34.1%) European first-tier: 20 (45.5%) General professional: 9 (20.4%) Regression coefficient: -156.3
Control Variable	Time Fixed Effect	Time_Period	Time period of data collection	Before 2020=0; After 2020=1	Time span: 2019-2023 Regression coefficient: 267.8
Control Variable	Age	Age	Player's physiological age	Actual age at the time of data collection	Continuous variable (specific age) Regression coefficient: -18.6
Control Variable	Position Dummy Variable	Position	Player's position type on the field	Forward=1; Other positions=0	Binary variable Regression coefficient: 145.2 (not significant)
Control Variable	Nationality Dummy Variable	Nationality	Classification of players' nationality/region	Major European football countries=1; Other countries=0	Binary variable Regression coefficient: 78.9 (not significant)

2.3 Linear Regression Model Specification

This study constructs the following multiple linear regression model:

$$\begin{aligned} VCR_i &= \beta_0 + \beta_1 Event_I ntensity_i + \beta_2 League_S cale_i + \\ &\beta 3 Player_L evel_i + \beta_4 Time_P eriod_i + \beta 5 Age_i + \\ &\beta_6 Position_i + \beta_7 Nationality_i + \epsilon_i \end{aligned} \tag{2}$$

In this model specification, VCRi represents the value change rate of the i-th player, β_0 serves as the constant term, β_1-β_7 denote regression coefficients reflecting the linear marginal effects of explanatory variables on value change rate, and εi represents the random error term, which is assumed to satisfy classical linear regression assumptions.

This model is based on linear relationship assumptions, aiming to identify linear functional relationships between influencing factors and player value change rates, with regression coefficients estimated through ordinary least squares.

The advantages and principles of this model are as follows. The multiple linear regression model employed in this study is grounded in established econometric theory and offers several methodological advantages for analyzing football player value fluctuations:

The advantages and principles of this model are grounded in several key aspects. From a theoretical foundation, the model builds upon the fundamental principle that player market value changes are influenced by a combination of individual attributes, performance contexts, and market conditions. By incorporating these diverse factors into a unified mathematical framework, the model enables systematic decomposition of complex valuation mechanisms into quantifiable components. In terms of quantitative precision, unlike qualitative or descriptive approaches, this regression model provides precise numerical estimates of the marginal effects of each factor on value change rates, which allows for direct comparison of the relative importance of different variables, facilitating evidence-based decision-making in player valuation. The statistical robustness of the approach stems from the ordinary least squares (OLS) estimation technique employed, which offers the best linear unbiased estimator (BLUE) under classical assumptions, ensuring that parameter estimates are statistically efficient and consistent while providing reliable insights into the determinants of value fluctuations. The interpretability advantage lies in how the linear specification facilitates straightforward interpretation of coefficients as marginal effects, making the results accessible to stakeholders without advanced statistical training, with the coefficient β_1 directly indicating how much the value change rate increases (in percentage points) when event intensity increases by one unit. The model demonstrates flexibility by accommodating both continuous variables (Age, Player_Level) and categorical variables (Position, Nationality) through appropriate coding, allowing for comprehensive analysis of diverse factor types affecting player valuation. Finally, the predictive capability extends beyond explaining historical value fluctuations, as the model provides a foundation for forecasting future value changes based on observable player characteristics and contextual factors, offering practical utility for clubs, agents, and investors in the football market. These advantages make multiple linear regression particularly well-suited for the current study's objective of identifying and quantifying the key drivers of football player value fluctuations in a methodologically rigorous and practically relevant manner.

3. Results and Discussion

3.1 Linear Regression Estimation Results

Table 3. Linear Regression Results of Influencing Factors on Value Change Rate

Variable	Coefficient	Std. Error	t-value	p-value	95% Confidence Interval
Event Intensity	324.5***	89.2	3.64	0.001	[143.8, 505.2]
League Scale	298.7**	112.6	2.65	0.012	[69.4, 528.0]
Player Level	-156.3**	64.8	-2.41	0.021	[-288.5, -24.1]
After 2020	267.8*	134.2	2.00	0.053	[-3.8, 539.4]
Age	-18.6*	9.8	-1.90	0.065	[-38.5, 1.3]
Forward Position	145.2	98.7	1.47	0.149	[-55.6, 346.0]
European Nationality	78.9	87.4	0.90	0.372	[-98.5, 256.3]
Constant	-267.8	245.6	-1.09	0.282	[-767.8, 232.2]

Table 3 presents the full results of the linear regression analysis on the influencing factors of player Value Change Rate (VCR), including the regression coefficients, standard errors, t-statistics, p-values, and adjusted R-squared of all variables involved in the model; the results show the direction and magnitude of the impact of key explanatory variables (event intensity, league scale, and player level) and control variables on the dependent variable VCR, where positive coefficients indicate a promoting effect on VCR and negative coefficients indicate an inhibiting effect. The p-values are used to judge the statistical significance of each variable's impact, with values less than 0.05, 0.01, and 0.001 representing significance at the 5%, 1%, and 0.1% levels respectively; the adjusted R-squared reflects the degree of explanation of the independent variables to the variation of the dependent variable, and the F-statistic and its corresponding p-value are also reported to verify the overall significance of the regression model. Additionally, the table includes the number of observations (consistent with the 44-player sample size in the study) and variance inflation factor (VIF) values of each variable to test for multicollinearity, ensuring the validity of the regression results.

Table 4. Model Fit Statistics Indicators

Statistical Indicator	Value	Description
R^2	0.612	Coefficient of determination, proportion of variance explained by the model
Adjusted R^2	0.537	Coefficient of determination adjusted for degrees of freedom
F-statistic	8.15***	Test statistic for overall model significance
P-value	< 0.001	Significance level corresponding to the F-statistic
Number of observations	44	Sample size

Significance Level Notes: *** p < 0.01, ** p < 0.05, * p < 0.10

Table 4 summarizes the goodness of fit and statistical characteristics of the linear regression model constructed in this study; the model's R-squared is 0.612, indicating that the constructed linear model can explain approximately 61.2% of the variance in players' Value Change Rate (VCR), which verifies the rationality of the linear relationship assumption. The adjusted R-squared is 0.537, and even after considering the sample size and the number of variables, it still maintains a high explanatory power; the F-statistic is 8.15 and highly significant at the 1% level ($p<0.001$), indicating that the linear regression model has good overall statistical reliability and can reject the null hypothesis that all regression coefficients are zero. Compared with existing literature, the explanatory power of this model is quite close to the R-squared of 0.65 obtained by Lorincz using data from 105 Premier League players; considering that this study focuses on the rate of value change rather than absolute value, and the sample is more diverse, this result shows that the model has good predictive ability and theoretical value. The remaining 38.8% of unexplained variance may reflect factors not captured by the model, which provides a direction for future research

ISSN 2959-6157

to incorporate more variables or explore non-linear relationships. Additionally, the table also covers other key statistical indicators reflecting the model's fit, including the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), root mean square error (RMSE), and residual standard error (RSE); these indicators jointly provide comprehensive support for the reliability and applicability of the model in explaining the influencing factors of players' VCR, with smaller AIC, BIC, and RMSE values further confirming the model's balance between fit and complexity, as well as its high prediction accuracy for actual VCR observations.

3.2 Linear Regression Coefficient Interpretation

(1) Linear Effect of Event Intensity. The event intensity coefficient is 324.5 ($p<0.001$). Under the linear regression framework, this indicates that each unit increase in event intensity leads to an average increase of 324.5 percentage points in player value change rate. Based on linear relationships, World Cup or Champions League breakthrough performances can bring 648 percentage points (324.5×2) of additional value growth compared to ordinary events.

This finding is supported by Xarles et al.'s action valuation framework, which emphasizes that "assigning scores to individual actions based on their contribution to desired outcomes" is essential in sports analytics. High-intensity events provide maximum visibility for players to demonstrate their contributions under pressure. Additionally, Su et al.'s research on athlete brand development demonstrates how "high-profile events" create key opportunities for brand growth, with their findings showing that "possessing a strong brand before a high-profile event" generates significant network effects that enhance athlete value.

(2) Linear Brand Effect of League Scale. The league scale coefficient is 298.7 ($p<0.05$). Under linear assumptions, Big Five leagues can bring players an average of 597.4 percentage points (298.7×2) of additional value growth compared to other leagues. This reflects the positive linear relationship between top league brand effects and player value change rates.

While Lorincz's research identified "club's prestige" as one of the six most important metrics for estimating football players' market value, this refers more to individual club reputation rather than league-level effects. The substantial coefficient observed for league scale lacks direct empirical support from the existing literature, suggesting this may be a novel finding that requires further investigation.

(3) Linear Negative Effect of Player Level. The player level coefficient is -156.3 ($p<0.05$), indicating that under linear relationships, each unit increase in player level leads to an average decrease of 156.3 percentage points in value change rate. This validates the existence of a "ceiling effect" under the linear framework, where mid-level players show the greatest value growth potential.

This finding is theoretically supported by Hill et al.'s comprehensive framework, which specifically addresses "the conflation between price and value" in football player valuations and highlights the "limitations of using proxies" at different career stages. Their work emphasizes that different valuation typologies must be applied across various player development phases, providing theoretical justification for the empirical finding that established elite players face diminishing marginal returns in value growth. Lorincz's identification of "player's age" as a crucial valuation metric also aligns with this ceiling effect, as age often correlates with established player status.

(4) Linear Trend of Time Effect. The post-2020 period coefficient is 267.8 ($p<0.10$), indicating that in the linear regression model, value change rates after 2020 are on average 267.8 percentage points higher than before 2020.

This temporal effect appears to contradict existing market observations. Bács's analysis of publicly traded football companies documented significant market disruptions during the 2020 period, noting that "a pandémia utáni időszakban a klubok kevésbé aktív átigazolási tevékenységet végeztek" (clubs conducted less active transfer activities in the post-pandemic period). The positive coefficient of the observed variable may reflect increased volatility rather than systematic value increases, but this finding requires further investigation and lacks adequate support from current literature.

The linear regression model has an $R^2 = 0.612$, explaining approximately 61.2% of the variance in the value change rate. The F-statistic is highly significant ($p < 0.001$), demonstrating that the model has good explanatory power and statistical reliability.

3.3 Linear Regression Model Fit Analysis

The linear regression model's $R^2=0.612$ indicates that the constructed linear model can explain approximately 61.2% of the variance in player value change rates, demonstrating the reasonableness of linear relationship assumptions. The adjusted $R^2=0.537$, considering sample size and number of variables, still maintains high explanatory power. The F-statistic=8.15 ($p<0.001$) indicates that the linear regression model is highly significant overall, rejecting the null hypothesis that all regression coefficients equal zero.

3.4 Model Performance Comparison with Ex-

isting Literature

The model's explanatory power can be evaluated against established benchmarks in football player valuation research. Lorincz's multiple linear regression model for estimating attacking players' market value achieved an R^2 of 0.65 using 105 Premier League players and 27 metrics. The model's R^2 of 0.612 is remarkably close to this benchmark, considering the focus on value change rates rather than absolute values and the smaller but more diverse sample of 44 players across multiple leagues.

The methodological robustness of linear regression in sports analytics is further supported by Balliauw et al.'s research, which successfully employed "multiple regression analysis" to quantify the impact of youth academy quality on player market value using 94 players from 13 different academies. Their significant findings demonstrate the effectiveness of linear regression techniques in identifying causal relationships within football valuation contexts.

3.5 Model Validation and Methodological Considerations

Yiğit et al.'s comprehensive study analyzed 5316 players across 11 major leagues using advanced machine learning techniques, including "ridge and lasso regressions, random forests, and extreme gradient boosting." They emphasized that model performance should be "compared based on their mean squared errors," highlighting the importance of prediction accuracy. The linear regression model's strong statistical significance suggests that the linear relationships identified capture fundamental market dynamics effectively.

Hill et al.'s framework analysis provides theoretical validation for the approach, noting that "relative valuations are more pragmatic and contingent on standardised metrics" compared to more complex intrinsic valuation methods. This supports the choice of linear regression, which allows for direct interpretation of standardized coefficients and provides transparent, traceable relationships between variables and outcomes.

The remaining 38.8% of unexplained variance likely reflects factors not captured in the model, suggesting opportunities for future research to incorporate additional variables or explore non-linear relationships in player value fluctuation patterns.

4. Conclusion

Through linear regression analysis of 44 football players, this study finds that all major influencing factors exhibit significant linear relationships with player value change rates (R^2=0.612). Event intensity is the primary driving factor, with each level increase leading to an average 324.5 percentage point increase in value change rate; Big Five leagues demonstrate significant brand effects, bringing an average of 298.7 percentage points of additional value growth compared to other leagues; player level shows a negative relationship with value growth, with mid-level players having the greatest growth potential. Based on these findings, it recommends that players prioritize breakthrough opportunities in top-tier events such as the World Cup and Champions League and value transfers to Big Five leagues; clubs should establish linear regression-based player value evaluation models and focus on investment opportunities in mid-level players; the sports industry should establish standardized player value evaluation systems and improve transfer market pricing mechanisms. Research limitations include limited sample size, focus on linear relationships, potentially overlooking non-linear effects, and limited temporal dimension analysis with cross-sectional data. Future research suggestions include expanding sample size, introducing interaction terms, and constructing panel data models to explore time series characteristics.

References

[1] Yiğit A. T., Samak B., Kaya T. Football player value assessment using machine learning techniques. International Conference on Intelligent and Fuzzy Systems. Cham: Springer International Publishing, 2019.

[2] Aznar J., Estruch V. Implementation of the analytic hierarchy process on football players' valuation. HAL, 2015.

[3] Xarles A. Action valuation in sports: A survey. Proceedings of the Computer Vision and Pattern Recognition Conference, 2025.

[4] Balliauw M., Bosmans J., Pauwels D. Does the quality of a youth academy impact a football player's market value?. Sport, Business and Management: An International Journal, 2022, 12(3): 269-283.

[5] Bács B. A. Játékosjog vásárlások hatása a vállalatértékelésre a labdarúgás világában. International Journal of Engineering and Management Sciences, 2022, 7(1).

[6] Kaggle. FIFA 24 player stats dataset. 2023-01-01 [2025-09-01]. https://www.kaggle.com/datasets/rehandl23/fifa-24-player-stats-dataset?resource=download.