

Forecasting Market Values of Players in Europe's Top Five Leagues Using ARIMA and EA FIFA Attribute-Augmented Models

Chenjia Liu *

Purdue University, West Lafayette, IN 47906, USA

* Corresponding author Email: RicardoLiu10@outlook.com

Abstract: This study examines how professional football players' market values change over time using monthly data from Transfermarkt and performance attributes from FIFA23–FIFA25. It uses two related time-series approaches. First, univariate ARIMA models trace the market value path of each player separately. Second, ARIMAX models build on that baseline by adding FIFA-based attribute groups as exogenous regressors to explain short-term changes in value. Treating each player as an independent time series makes it possible to capture different career paths instead of reducing them to aggregate patterns. The results suggest that most players' market values move in relatively smooth and stable ways, and low-order specifications such as ARIMA (1,0,0) appear most often. Adding FIFA attributes makes the model easier to interpret and shows positional differences. Attacking players are more responsive to attacking and skill-related attributes, whereas defenders and goalkeepers react more to defensive and mentality-related attributes. A few players do not fit these patterns well. Their valuation dynamics seem to be shaped more strongly by transfers, sudden reputation shifts, or career-stage effects. Taken together, the study offers a quantitative framework for linking time-series modeling with performance attributes to understand football player market values more clearly.

Keywords: Football Player Market Value; Time Series Forecasting; ARIMA Model; FIFA 25 Attributes; Europe's Top Five Leagues.

1. Introduction

Clubs and analysts often use the market value of professional football players to evaluate performance and transfer potential[1,2]. As football has become more data-driven, more studies have tried to predict player performance and match results with statistical and machine learning methods[3,4].

This study applies two time-series methods to monthly market value. The first is the ARIMA model, which tracks the valuation trend of each individual player[3-5]. The second is the ARIMAX model. It adds FIFA attribute groups, including attacking, movement, power, and defending, as external regression factors so that changes in value can be estimated more directly. Treating each player as a separate time series makes it possible to follow that player's career path over time rather than collapsing different patterns into a single average.

The results give a quantitative way to relate performance attributes to market value and to spot outliers whose positional profiles and valuation patterns differ from the general trend. This may be useful for scouting and valuation analysis, and it also gives a more precise view of how the modern football economy operates.

2. Data

2.1. Data sources

This study used a dataset that combined two data sources covering November 2023 to August 2025. The first source was Transfermarkt, which reports each player's monthly market value[6]. To reduce scaling differences and make the model more stable, these values were converted to logarithmic form. The second source was FIFA23-FIFA25 player attributes, downloaded during the same period as the

first source[7]. These attributes give quantified indicators of player performance and are commonly used to predict player value[1].

2.2. Variable Construction and Preprocessing

Each player has over 50 raw attributes. For analysis, these were grouped by averaging the relevant underlying indicators into six broader categories: attacking, skill, movement, power, mentality and defending. Missing attribute values were excluded rather than imputed because doing so could have introduced artificial distortion.

After preprocessing, a structured CSV file was generated for each player. Each file contains the following variables:

- Date – Normalized to the first day of each month
- Name and sofifa_id – Unique player identifier
- Transfermarkt value and logarithm of transfer market value
- Raw transfer market value and logarithmic Transfermarkt value
- Age – Calculated based on date of birth
- Position – Cleaned, retaining only primary position information
- Comprehensive attribute groups – Attacking, Skill, Movement, Power, Mentality and Defending

The final processed dataset includes monthly observations for each player over the past two years. This gave enough variation over time to estimate the ARIMA and ARIMAX models separately. A merged summary table was then prepared for the analysis, bringing together model parameters, coefficient estimates, and player positions.

3. Preliminary Analysis

3.1. Descriptive Properties of the Player-Level Time Series

Before fitting the forecasting models, I first examined the cleaned and merged dataset to see how it was structured and what the player-level time series looked like. The combined data contain monthly observations for a panel of elite football players between late 2023 and mid-2025. Each record includes the player's Transfermarkt valuation, log-transformed value, and a set of FIFA game series-based performance indicators.

All players follow the same monthly timeline, and no forward-filling was needed. This means the dataset is a balanced panel, which makes the later modeling stage more straightforward.

A visual inspection of the raw market-value trajectories shows that most players change fairly smoothly over time, with gradual increases or decreases rather than abrupt jumps. In practice, player valuation appears to move like a low-frequency economic signal, which makes ARIMA-type modeling[3,5] a reasonable choice. The log transformation also produces a more stable distribution with lower skewness, which helps stabilize variance across players with very different nominal values.

3.2. Temporal Behavior of FIFA Attributes and Positional Composition

The covariates based on FIFA attributes, overall rating,

attacking, skill, movement, power, mentality, defending, and age, vary moderately across players, while changing only slowly for the same player over time. Month-to-month movement in these variables is limited. That makes them suitable external regressors for ARIMAX models, because they add useful information without adding much noise. The distribution of playing positions also shows that the sample is heterogeneous, with defenders, midfielders, attackers, and goalkeepers all included. This allows position-based comparisons at a later stage of the modeling.

These initial observations indicate that the dataset meets the basic assumptions for both univariate ARIMA and multivariate ARIMAX modeling and support the use of time-series forecasting for player-specific market-value dynamics.

4. ARIMA Model Selection

4.1. Model Specification and Selection Criterion

To simulate changes over time in the market valuations of football players, we fitted an ARIMA (p, d, q) model for each player[5]. For model selection, we used the Akaike Information Criterion as the main criterion so the different parameter combinations could be compared on the same basis.

4.2. Distribution of AIC Values and Parameter Frequencies

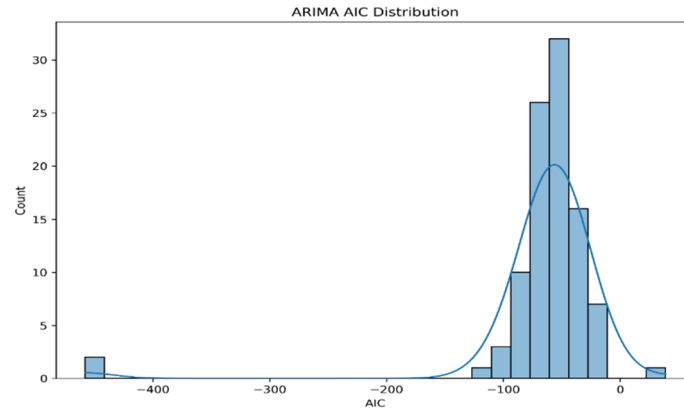


Figure 1. ARIMA AIC Distribution

Figure 1 shows the distribution of AIC values for all fitted ARIMA models. For most players, the AIC values fall between -120 and -40, which suggests that model fit is fairly

consistent across the sample. A few outliers appear, likely linked to unusual valuation trajectories, but they are relatively rare and do not dominate the dataset.

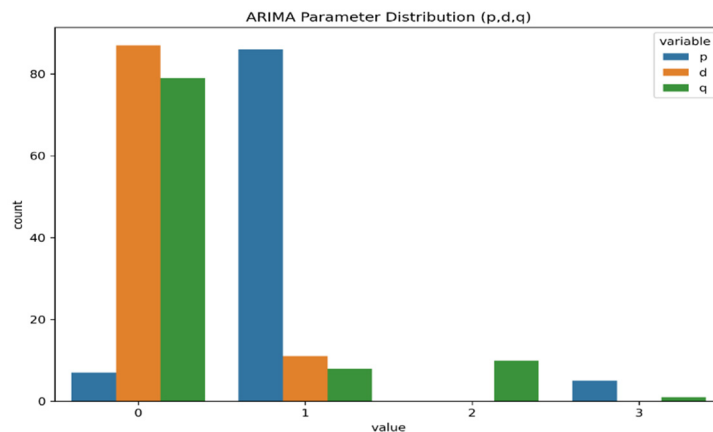


Figure 2. ARIMA Parameter Distribution(p,d,q)

Figure 2 summarizes the frequency of the selected parameters p , d , and q for all players. The results reveal clear patterns:

- $p = 1$ is the most common autoregressive order.

- $d = 0$ is dominant, indicating that most valuation time series are stationary and do not require differencing.

- $q = 0$ is also quite common, while higher-order moving average terms appear only in a minority of participants.

Taken together, these patterns suggest that the simple ARIMA (1,0,0) model fits most participants best, capturing short-term persistence without requiring differencing or higher-order moving average components. Only a very small number of participants needed more complex model settings, usually because of structural breaks or abrupt changes in valuation.

This consistent structure among participants lays a solid foundation for subsequent extensions of the ARIMAX model,

in which participant attributes are incorporated as exogenous predictors.

5. ARIMAX Model Results and Interpretation

5.1. Model Fit and Order Selection

To test whether player attributes improve short-term forecasting, the baseline ARIMA model is extended to an ARIMAX framework with eight external regressors from FIFA: overall rating, age, attacking, skill, movement, power, mentality and defending.

This section presents a comprehensive analysis of model fit, parameter behavior, and positional differences using all five diagnostic plots.

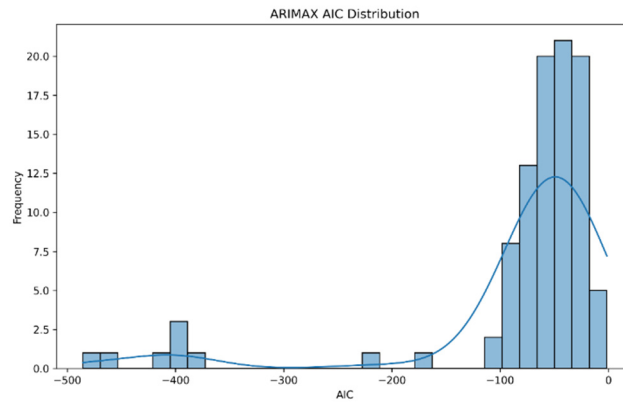


Figure 3. ARIMAX AIC Distribution

Figure 3 shows two main patterns. Most players have AIC values concentrated between approximately -120 and -20 , which suggests a fairly similar level of goodness-of-fit across most series. A smaller group shows extreme AIC values below -300 . These cases match valuation series with strong deterministic trends or very smooth patterns, where the ARIMAX specification aligns unusually well with the

observed values.

The distribution of AIC values therefore suggests that adding FIFA attributes usually gives stable model performance across heterogeneous player trajectories. At the same time, it separates out a few cases in which the exogenous information improves model fit by a large margin.

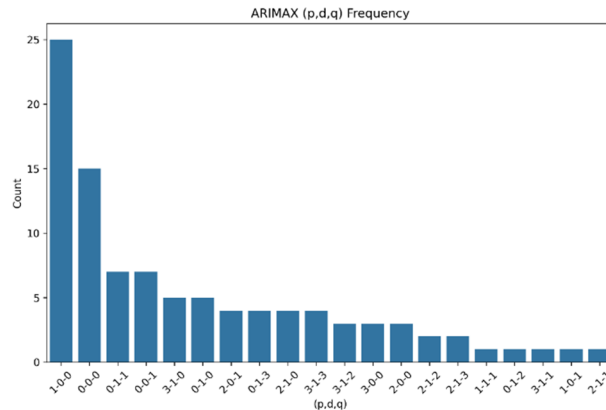


Figure 4. ARIMAX(p,d,q) Frequency

The pattern of selected model orders points to relatively simple model structures. According to figure 4, the most common model is (1,0,0), followed by (0,0,0), (0,1,1), and (0,0,1). Most of the fitted models are low-order. This suggests that autoregressive dependence is generally weak, so most players do not need long time memory in the model.

Differencing is sometimes needed, especially for players

whose series show asymptotically monotonic trends. Moving-average terms are less common, but they appear in some cases where player values show short-term volatility. After exogenous features are added, the valuation series look statistically simpler than traditional financial or macroeconomic time series.

5.2. Distribution of ARIMAX Coefficients

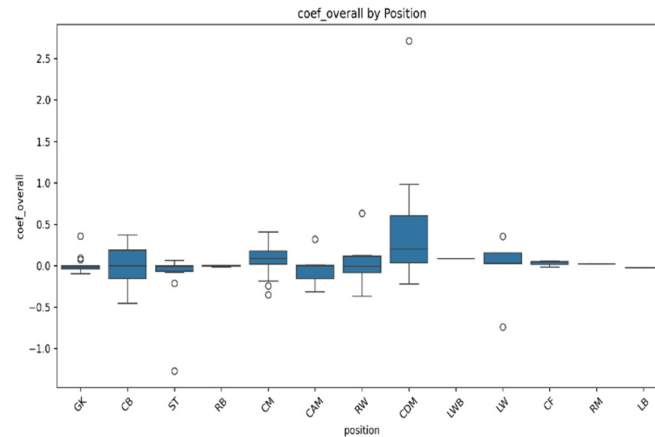


Figure 5. coef_overall by Position

Figure 5 shows clear structural differences across tactical roles. The coefficients for CDM, CM, CAM, and RW are more concentrated, which suggests that the value fluctuations of midfielders and transitional players are more closely tied to the overall rating. By contrast, the coefficients for GK, CB, RB, and LB are tightly clustered around zero. For these positions, player value appears to depend more on specific attributes, such as reaction speed and tackling ability, which

are not fully captured by the overall "general" rating.

The pattern for attacking roles is less stable. The coefficients for ST and CF are more dispersed, which fits the higher volatility and performance-driven value fluctuations usually seen among forwards. Taken together, this position-based decomposition suggests that role-specific modeling is necessary because different tactical positions respond differently to FIFA's overall rating.

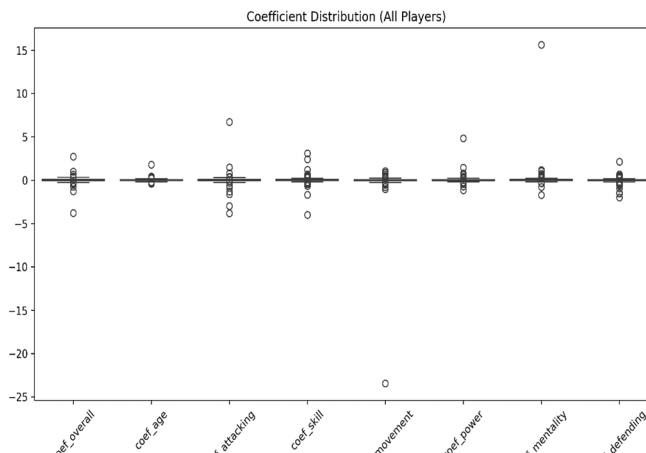


Figure 6. Coefficient Distribution (All Players)

Figure 6 shows the distribution of all estimated ARIMAX coefficients across players and summarizes how FIFA attributes relate to short-run changes in market value. Most coefficients are tightly grouped around zero, which suggests that, for most players, changes in attributes have only small marginal effects on value dynamics. A few coefficients, however, lie far from the center. This matters. These extreme outliers point to clear heterogeneity across players, as a small group shows unusually large positive or negative sensitivities. In some cases, this may reflect unusual valuation patterns, while in others it may come from anomalies in their rating histories.

The attributes linked to attacking, movement, and mentality have a slightly wider spread than the rest. That pattern suggests more variation in how these qualities feed into transfer-market volatility. Even so, the main pattern in the figure is still the same: ARIMAX effects are usually small and fairly stable across players, though some idiosyncratic deviations remain and are worth examining further at the positional level.

5.3. Positional Heterogeneity and Attribute Sensitivity

Figure 7 summarizes the average influence of each attribute across all positions. The matrix shows several clear structural patterns. LM has especially extreme values, with highly positive attacking coefficients and highly negative movement coefficients, which suggests that valuation changes for wide midfielders are very sensitive to these two dimensions. CDM has a clearly positive coefficient for overall rating, which fits the role's dependence on general ability more than on any single skill attribute. GK also stands out, with higher power and mentality coefficients, in line with the physical and psychological demands of goalkeeping. The pattern is fairly consistent. Attacking positions such as RW, LW and ST show moderate positive sensitivity to attacking and skill attributes, which matches their tactical roles. Overall, this heatmap offers a direct comparison of how playing roles relate to FIFA skill dimensions and works as a "characteristic-position" map for interpreting the ARIMAX structure.

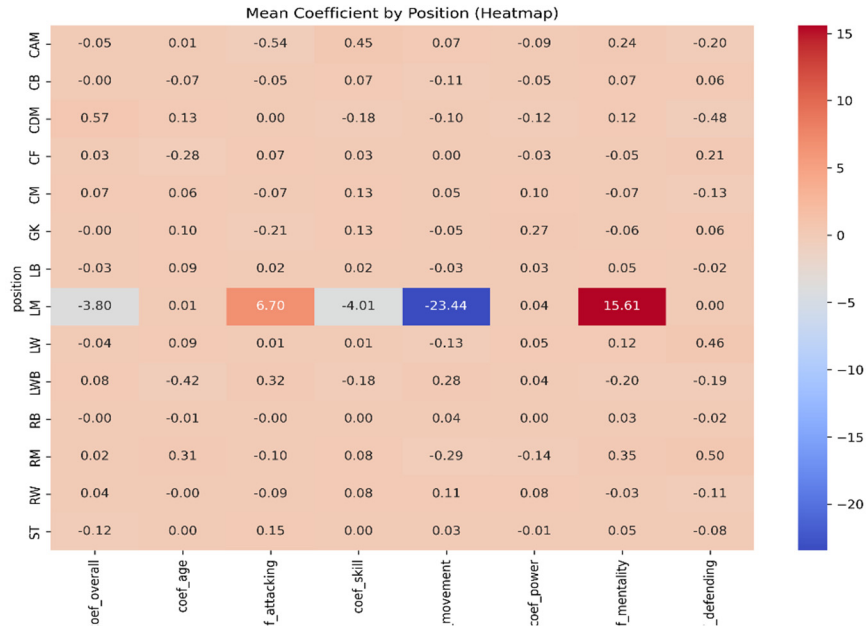


Figure 7. Mean Coefficient by Position (Heatmap)

6. Outlier Analysis

A small number of players exhibit coefficient patterns and information criteria that deviate substantially from the rest of the sample.

6.1. Joan Garcia

Joan Garca is unusual because his ARIMAX coefficients are very large and their signs do not follow a clear pattern. His model uses a simple (0,1,1) specification, but the exogenous responses are extreme, for example overall 0.16, skill +2.39, power +4.82, and attacking 2.98. In this case, the FIFA attributes do not seem to track his market-value path in a stable linear way.

The timing of his rise is probably one reason. Garca seems to be a recent breakout player whose visibility increased quickly, ending in his summer transfer to FC Barcelona. His historical market-value series is short and includes sharp jumps over a limited period. That matters here. With less history to work grammarians have less room to estimate coefficients that remain stable, and that likely helps explain the unusual parameter values.

6.2. Nico Williams

Nico Williams is also an outlier because his ARIMAX coefficients are unusually large in several attributes, especially attacking (+6.70), movement (-23.44), and mentality (+15.61). These values are much higher or lower than those found for the other players in the sample, which suggests that the relationship between his FIFA characteristics and monthly market value changes is less stable.

He is also the only LM in the data set. This has a direct effect on the results. Since there is no other LM to offset the pattern, his coefficients strongly shape the LM averages shown in the positional heatmaps and boxplots, which creates visible differences from the other roles. One likely explanation is how quickly his performance level and transfer valuation rose at a young age. If that rise happened faster than the model could capture cleanly, the estimated responses would be less stable than those of both his positional group

and the wider sample.

7. Discussion

For both ARIMA and ARIMAX, the results show that monthly Transfermarkt values for most players change gradually and can be modeled with simple time-series structures, usually (1,0,0) or (0,1,1). The AIC values are also fairly consistent across the sample, which suggests that forecasting performance does not vary much for most players. A few cases were less regular.

Adding FIFA-based attributes to ARIMAX makes the model easier to interpret because it shows which skill dimensions are linked most clearly to month-to-month value changes[1]. Predictive performance also differs across models, and interpolation or machine learning methods can sometimes do better than time-series models, depending on the structure of the data[4]. Earlier work reports a similar pattern, where model choice has a strong effect on valuation accuracy [2].

Most coefficients are still fairly small, but the differences by position make sense. Attack-related variables matter more for forwards and wingers, while defensive attributes have greater relevance for center-backs and goalkeepers. The coefficient distributions also show some uneven patterns. For example, mentality and movement have a wider variance, which suggests that their effects differ more across players.

8. Future Possibilities

Several directions remain for later work. Using a wider sample, across more seasons or additional leagues, would allow longer-term patterns to be tracked and make comparisons between periods more reliable. It would also make it easier to check whether the trends found here hold in a broader set of cases.

The models could be extended by including transfer-related events such as injuries, club changes, and managerial shifts. These factors may explain breaks in player valuation that the current specification does not capture well.

Another step would be to test machine-learning methods such as gradient boosting, LSTM models, or mixed-effects

frameworks. These approaches may improve predictive accuracy. They may also detect nonlinear interactions between player attributes and market valuation that are harder to identify in the present model.

9. Conclusion

This study examines monthly Transfermarkt market values for 98 elite players with two time-series approaches, a univariate ARIMA model and a multivariate ARIMAX model that adds FC25-derived performance attributes. ARIMA provides a baseline for how values change over time, and for most players the series fit stable, low-order forms such as (1,0,0) or (0,1,1). When FIFA attributes are added, the pattern changes. ARIMAX accounts for more variation in player valuations, shows positional differences, and identifies several exogenous sensitivities. For example, defenders respond more strongly to defensive and mentality attributes, while attacking players tend to have larger coefficients for skill and attacking categories.

The analysis also identifies several outlier players, including Joan Garca and Nico Williams, whose coefficient profiles differ sharply from what is typical for their positions. Their market values may have depended more on context, shifted quickly with changes in reputation, or been affected by club-specific shocks. In these cases, attribute information alone does not explain the observed market value. Timing matters too. Transfer events and media exposure also appear to have contributed.

ARIMA and ARIMAX therefore provide a useful way to study both time-based patterns and the attribute-related

drivers of football player valuations. In this study, combining real-world market data with synthetic game-based performance indicators appears to show meaningful differences across playing positions and career paths.

References

- [1] Anjum, S., & Fatima, A. (2023). Predictive analytics for FIFA player prices: An ML approach. *Journal of Scientific Research and Technology*, 1(6), 204–212.
- [2] He, Y. (2015). Predicting market value of soccer players using linear modeling techniques [Undergraduate honors thesis]. University of California, Berkeley. https://www.stat.berkeley.edu/~aldous/Research/Ugrad/Yuan_He.pdf.
- [3] Jurman, G. (2019). Seasonal linear predictivity in national football championships. *Big Data*, 7(1), 21–34. <https://doi.org/10.1089/big.2018.0076>.
- [4] Kontos, C., & Karlis, D. (2023). Football analytics based on player tracking data using interpolation techniques for the prediction of missing coordinates. *Italian Journal of Applied Statistics (Statistica Applicata)*, 35(2). <https://doi.org/10.26398/IJAS.0035-010>.
- [5] Gupta, A. (2017). Time series modeling for dream team in fantasy Premier League. In *Proceedings of the International Conference on Sports Engineering*. arXiv:1909.12938. <https://arxiv.org/abs/1909.12938>.
- [6] Transfermarkt. (n.d.). Player market values. <https://www.transfermarkt.Com>.
- [7] SoFIFA. (n.d.). FIFA player attributes database. <https://sofifa.Com>.