Money Talks: Predicting FIFA 23 Player Wages from On-Pitch Performance

Matt Suba and Matthew Varela

STATS 401: Applied Statistical Methods II

April 10th, 2025

Word Count: 1319

Background

Professional soccer intrigues its fans not only for the entertainment on the field but also what happens off the field and how the sport operates. Weekly wages, transfer fees, and bonuses often generate headlines and show gaps between perceived talent and the player's actual pay. The analysis uses the publicly available FIFA 23 player database, which includes roughly 7000 players, to investigate which measurable skills best explain a player's weekly wage measured in euros. The database, scraped from the official EA Sports ratings site, records each registered player's primary position, age, primary foot, and more than 30 technical and physical attributes measured on a scale from 1-99. Weekly wage, the response variable, spans widely, from 250 to 560,000 euros, motivating a logarithmic transformation. Observational units are individual professional soccer players, each described by a vector of attributes taken at the start of the 2023-24 season. The importance of creating correct valuations for players stems from the fact that clubs only have a limited amount of money, overvaluing a player would result in worse on-field performance in comparison to their salary, while finding undervalued players allows clubs to get away with less spending while receiving better performance.

The purpose of this analysis includes the potential to improve financial decision making within soccer clubs. By utilizing a linear regression model that incorporates player attributes, specifically age, preferred foot, shooting, pace, attacking-crossing, passing, and dribbling, the study identifies which skills meaningfully correlate with wage structures. This can allow clubs to optimize their funds and strategically recruit their players.

Analysis – Creating the Model

The initial model consists of a multiple linear regression of weekly wage in euros on shooting, pace, attacking crossing, and preferred foot (Figure 1). Although three out of the four variables are highly significant predictors of wage, pace being the only one with a relatively high p-value, there are clear signs that the model violates necessary assumptions. With an R^2 value of 0.1168, the model explained only 12% of the variation in wages. In addition, the residual-versus fitted plot showed a clear funnel shape while the Q-Q plot illustrated a long right hand tail, showing evidence that assumptions of constant variance and normality are violated.

Log Transformation

Because wages are highly right-skewed, a log-transformation was applied to the response variable. Histograms before and after transformation (Figure 2) confirm that log(wage_eur) is far closer to symmetric. Refitting the additive model on the log scale (Figure 3) improved diagnostics, residual spread is more homogeneous and Q-Q deviations are milder, while increasing adjusted R^2 to 0.183. The linearity assumption now appeared reasonable, so subsequent model building proceeded on the log scale. Since all the skill ratings are on a scale of 1-99 and their distributions are all relatively normal and symmetric, there is no need for log transformations to the predictors.

Exploring an Interaction

Skills contributions may differ for left- versus right-footed players, particularly traits closely tied to footedness like attacking-crossing. Therefore, an interaction between attacking crossing and preferred foot was introduced. The fitted-line plot (Figure 4) suggests steeper wage growth for left-footed players as crossing rating increases. The interaction term was highly significant ($p \approx 0.001$), and an ANOVA comparing additive and interaction models (Figure 8) yielded $F = 10.94$, $p < 0.001$, and the interaction model very slightly improved overall fit, raising adjusted R^2 from 0.183 to 0.184 and trimming RMSE by 0.001 log-EUR (a negligible change).

Additional Predictors

To enhance the accuracy of the model, three new quantitative variables were considered– age, passing, and dribbling. Exploring the effect of these predictors, stepwise fitting revealed that dribbling, age, and passing add great explanatory power. In total the incorporation of these variables lifted the model's adjusted R^2 to 0.344 and lowered the RMSE to about 22.4 thousand euros, indicating that adding these predictors explains a greater proportion of the variance in the wages and the model predicts smaller errors. The diagnostic plots showed a largely random residual cloud with little deviation from the line of fit.

Analysis – Interpreting the Model

Diagnostic Checks & Assumptions

Residual-versus-fitted and Q-Q plots for the final model (Figure 5) show approximate linearity and no pronounced curvature. Slight right-tail deviations persist, driven by a few super-earners, yet the pattern is acceptable for large-sample inference. Residual variance is roughly constant across fitted values, satisfying the homoscedasticity assumption. Because each player appears only once and ratings are assigned independently, the independence assumption seems reasonable.

Model Fit

The final model explains about 34% of the variability in log-wages (adjusted $R^2 = 0.344$) and reduces RMSE by roughly 0.13 log-EUR compared with the additive log model (Figure 6). Although compensation is influenced by unobserved factors (endorsements, contract length, league revenue), capturing one-third of the variation with a handful of publicly visible traits represents meaningful explanatory power.

Multicollinearity & Overfitting

Variance Inflation Factors (VIFs) (Figure 7) range from 1.26 (age) to 6.80 (passing). VIFs for the additional predictors model were under 5 for all predictors except passing (VIF = 6.80) and dribbling (VIF = 6.77), indicating slight multicollinearity between those two skills (Figure 7). This would make sense given that these are two very broad and important skills when assessing the value of a player. Multicollinearity mainly inflates the standard errors and can destabilize the individual coefficients, but since both skills are core to player valuation they were retained. Therefore, their coefficients were interpreted with caution while relying on RMSE to judge overall model quality. With $n \approx 7{,}000$ and only eight predictors (including the interaction), the model is unlikely to be over-fit.

Predictor Results

Five terms are statistically significant: age, passing, dribbling, attacking crossing, and the crossing x footedness interaction. Shooting also remains significant but carries a small negative sign while pace is not significant. To gauge the practical magnitude of these effects, a ten-point increment to each of the skills was considered, and their effects were computed in context.

- **Passing** - A 10-point increase in passing rating is associated with a 0.754 rise in log-weekly wage, about a 113% raise, holding other variables fixed.
- **Dribbling** - A 10-point increase raises wages by roughly 0.632 log-EUR (≈ 88%)
- **Attacking-crossing** - A 10-point increase enters with a negative coefficient (-0.028 log-EUR per point), indicating that, all else equal, players rated higher on crossing tend to earn slightly less ($p < 0.001$).

- **Age** - Each additional year corresponds to a 0.0267 increase in log wage, a ≈ 2.7% increase, reflecting experience premiums.
- **Interaction (Attacking-crossing × Right Foot)** – For right-footed players the marginal benefit of crossing is 0.004 log-EUR lower than for left-footed players ($p \approx 0.07$), matching the scarcity value of proficient left-footed crossers.
- **Shooting** – Surprisingly, shooting's coefficient turned slightly negative after controlling for creative metrics (−0.015 per point). This likely reflects overlap with passing and dribbling; once creativity is known, shooting adds little incremental wage value.
- **Pace and the preferred-foot** main effect were not statistically significant at $\alpha = 0.05$, suggesting raw speed and dominant foot, in isolation, do not command extra salary once other skills are accounted for.

Conclusion

Leveraging a streamlined set of FIFA 23 ratings, a log-linear wage model was constructed that meets classical regression assumptions and captures a substantive share of pay variability Technical creativity (passing and dribbling) and experience emerge as the most lucrative traits, whereas raw speed and shooting appear less valuable after other skills are accounted for. Left-footed crossers enjoy a small but noteworthy premium, reflecting positional scarcity.

With additional time no-linear effects (splines for age) would be explored, league or team-level fixed effects would be incorporated, and results on future FIFA editions to assess temporal stability would be validated. Nonetheless, the findings already provide actionable guidance: clubs seeking wage efficiency should scout undervalued right-footed players who excel in passing and dribbling but whose crossing remains average, precisely the profile the model predicts will be paid below market potential.