ECO 2223 W26 — Final Data Project

What Determines Premier League Midfielder Weekly Wages?

Season: 2024-25

Group 6
Mirkomil Oblokulov
Idris Ali
Louisy Brian Walter Dzura
Nadila Ainiwaer
Venhuizen Robyn Tina

ECONOMETRICS 2223
King's University College at Western University
April 2026

Part 1: Introduction

Professional football has become one of the most lucrative industries in global sports, with the English Premier League standing as the wealthiest domestic football league in the world. Player wages represent the single largest expenditure for Premier League clubs, yet the determinants of these wages remain an important question for club management, agents, and sports economists. This report investigates the question: *what player-level factors determine weekly wages for Premier League midfielders?*

Midfielders are particularly interesting to study because they occupy a diverse set of roles, from deep-lying defensive midfielders to attacking playmakers and goal-scoring midfielders. Unlike strikers, whose value is often measured primarily by goals, or defenders, whose value is tied to clean sheets, midfielders contribute across multiple dimensions of play, making it less obvious which attributes clubs value most when setting wages.

Economic theory suggests that player wages should reflect marginal revenue product, the additional revenue a player generates through their on-pitch performance. Lucifora and Simmons (2003) were the first to study the correlates of pay and superstar effects using Italian Serie A wage data, finding that individual performance measures such as goals, assists, and appearances are significant determinants of wages. Bryson, Frick, and Simmons (2013) showed that productivity measures like goals and assists drive wage premiums in European football. Frick (2007) employed multiple regression analysis on German football data, identifying player productivity and career longevity as key salary determinants. In a subsequent study, Frick (2011) used panel data from the German Bundesliga spanning up to thirteen seasons and demonstrated that the variance in player salaries is largely explained by individual performance, including career games played, international appearances, and goals scored. He also found evidence of moral hazard, with players reducing effort in the final year of their contracts. Carrieri, Principe, and Raitano (2018) analysed football wages in Serie A and found that player popularity and agent bargaining power also contribute positively to wages, with popularity being particularly important at the top of the salary distribution.

A related strand of literature examines market value rather than wages directly. Sanchez et al. (2022) analysed player market values across the Top 5 European leagues using ANOVA and

found that Premier League players command significantly higher valuations than players in other leagues, that forwards and midfielders are valued above goalkeepers and defenders, and that players aged 21 to 30 with international caps have significantly higher market values. Ferreira (2021) employed fixed effects panel data regression across six European leagues, splitting the analysis by position group. His results showed that assists have a significant positive effect on midfielder market values while goals are the dominant driver for attackers, providing direct support for the hypothesis that creative output is the most valued attribute for midfielders.

More recently, Franceschi, Brocard, Follert, and Gouguet (2024) conducted a systematic review of the determinants of football player valuation across 29 econometric studies. They proposed a set of core explanatory variables, including age, goals, assists, minutes played, international appearances, and contract duration, as the common ground for modelling player value. Our study builds on this literature by focusing specifically on midfielders in the Premier League and testing whether creative output (assists) is valued more highly than goal-scoring for this position group.

Our **treatment variable is assists**, representing creative playmaking ability. We hypothesize that assists are the primary driver of midfielder wages because midfielders are primarily valued for their ability to create chances. Control variables include goals, minutes played, age, yellow cards, red cards, position dummies, and a Big 6 club dummy (Arsenal, Chelsea, Liverpool, Manchester City, Manchester United, Tottenham). We also test whether geographic origin (continent) affects wages after controlling for performance, and whether age has a non-linear (quadratic) effect on wages.

Part 2: Data Source and Regression Model

Our dataset consists of 261 midfielders from the 2024-25 English Premier League season. Player performance statistics were collected from FBRef.com, a comprehensive football statistics database that sources data from StatsBomb and Opta. Weekly wage data were collected from Capology.com, which compiles reported salary information for professional football leagues. All data cleaning, variable creation, and regressions were performed in STATA; the accompanying log file documents the full analytical workflow.

The data is cross-sectional, covering all 20 Premier League clubs. We include players classified as midfielders (MF), midfielder-forwards (MF/FW or FW/MF), and midfielder-defenders (DF/MF or MF/DF). The sample includes players aged 16 to 40, with weekly wages ranging from 769 to 400,000 pounds. Such a wide range indicates the presence of outliers, which may affect the stability of coefficient estimates. Position was converted into two dummy variables (pos_mf_fw for attacking midfielders, pos_df_mf for defensive midfielders), with pure midfielders as the reference group. Geographic origin was converted into four regional dummies (South America, Africa, North America, Asia), with Europe as the reference group. We also construct a Big 6 club dummy variable (top6) equal to 1 if the player belongs to Arsenal, Chelsea, Liverpool, Manchester City, Manchester United, or Tottenham, and 0 otherwise. This variable captures the wage premium associated with playing for the league's wealthiest and most commercially prominent clubs.

Our proposed regression model:

$$Wage = \beta0 + \beta1(Goals) + \beta2(Assists) + \beta3(Minutes) + \beta4(Age) + \beta5(Yellow) + \beta6(Red) + \beta7(Pos_MF_FW) + \beta8(Pos_DF_MF) + \beta9(Top6) + \varepsilon$$

We also test extended specifications including: (1) a simple bivariate regression of wages on assists, (2) age-squared to capture a non-linear age effect, (3) a log-linear model with ln(Wage) as the dependent variable to reduce heteroscedasticity, and (4) a logit model to examine which factors predict whether a player is a high earner.

Part 3: Data Analysis

Table 1: Summary Statistics of Key Variables

Variable	Obs	Mean	Std. Dev.	Min	Max	r(Wage)
wage_weekly	261	80,750	64,099	769	400,000	1.000
goals	261	1.93	2.45	0	14	0.370
assists	261	1.48	1.80	0	16	0.467
minutes	261	1,208	769	1	2,783	0.285
age	261	25.62	3.88	16	40	0.288
yellow	261	2.56	2.49	0	10	0.121
red	261	0.07	0.25	0	1	0.095

pk	261	0.14	0.56	0	5	0.194
pos_mf_fw	261	0.25	0.44	0	1	0.129
pos_df_mf	261	0.16	0.37	0	1	-0.152
top6	261	0.29	0.46	0	1	0.534

Source: FBRef.com and Capology.com, 2024-25 Premier League season. N = 261.

The average weekly wage for a Premier League midfielder is approximately 80,750 pounds, with substantial variation (standard deviation of 64,099). This demonstrates that wages are highly dispersed, which may affect the precision of regression estimates. The distribution is right-skewed, with a median of 65,000 pounds, reflecting a small number of very high earners. The level of variation suggests that a limited number of high earning soccer players may strongly influence the overall distribution. The average player is about 26 years old, scores roughly 2 goals and provides 1.5 assists. Assists have the strongest simple correlation with wages among performance variables (r = 0.467), followed by goals (r = 0.370) and age (r = 0.288). This provides initial evidence that assists may play the a vital role in explaining wages in the regression analysis. Notably, the Big 6 club dummy has the highest correlation with wages of any variable (r = 0.534), with Big 6 midfielders earning an average of 134,086 pounds per week compared to 58,839 pounds for non-Big 6 midfielders. This large difference highlights the importance of club-level factors beyond individual performance.

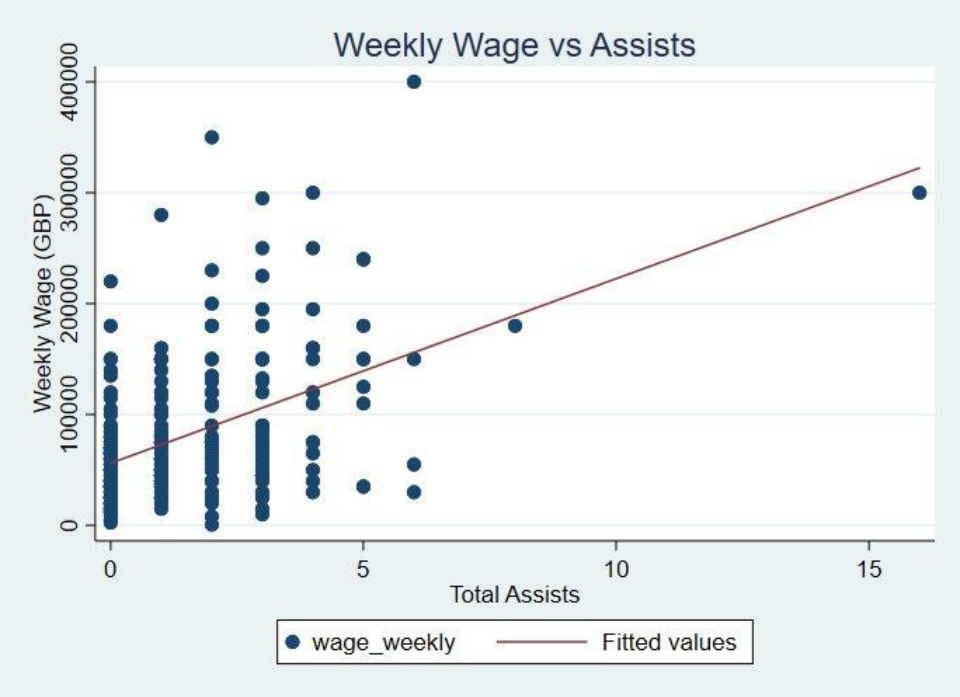

Figure 1: Weekly Wage vs Assists

Figure 1 displays the bivariate relationship between weekly wages and total assists. The fitted line has the steepest slope among all performance variables, consistent with assists being the strongest individual predictor (r = 0.467). The positive relationship is clear, though there is considerable dispersion around the fitted line at lower assist counts, suggesting that other factors such as club status also drive wages. This aligns with the hypothesis that creative output is a primary driver of midfielder wages.



Figure 2: Weekly Wage vs Goals Scored

Figure 2 shows the relationship between weekly wages and total goals scored. The positive slope confirms that goal-scoring midfielders tend to earn more (r = 0.370), though the relationship is weaker than for assists. A cluster of high-wage players appears at moderate goal counts (3 to 7 goals), likely reflecting Big 6 midfielders who contribute goals alongside creative output. However, the weaker relationship compared to assists suggests that goal-scoring may be a less consistent determinant of wages for midfielders.



Figure 3: Weekly Wage vs Age

Figure 3 illustrates the age-wage relationship. The fitted line shows a positive linear trend, but the scatterplot reveals a non-linear pattern: wages rise steeply through the mid-twenties, plateau in the late twenties and early thirties, and decline for players over 34. This motivates the inclusion of an age-squared term in Models 3 and 5. The highest earners are concentrated between ages 28 and 32.

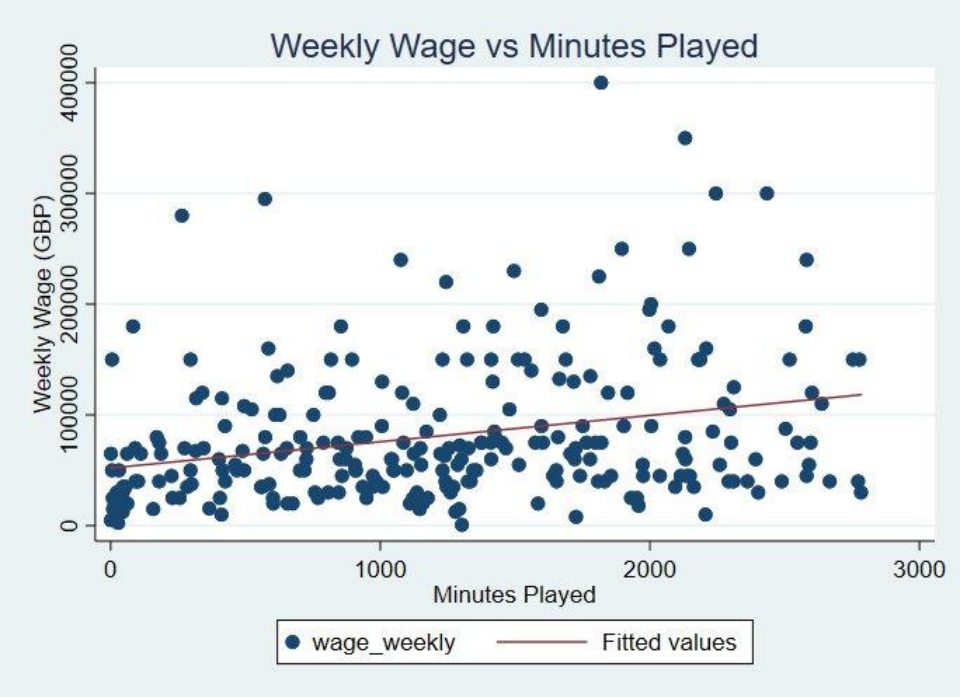

Figure 4: Weekly Wage vs Minutes Played

Figure 4 plots wages against minutes played. The relationship is notably weak (r = 0.285) with an almost flat fitted line. This suggests that wages are largely determined before the season begins through contracts reflecting expected ability and club status, rather than by in-season playing time.

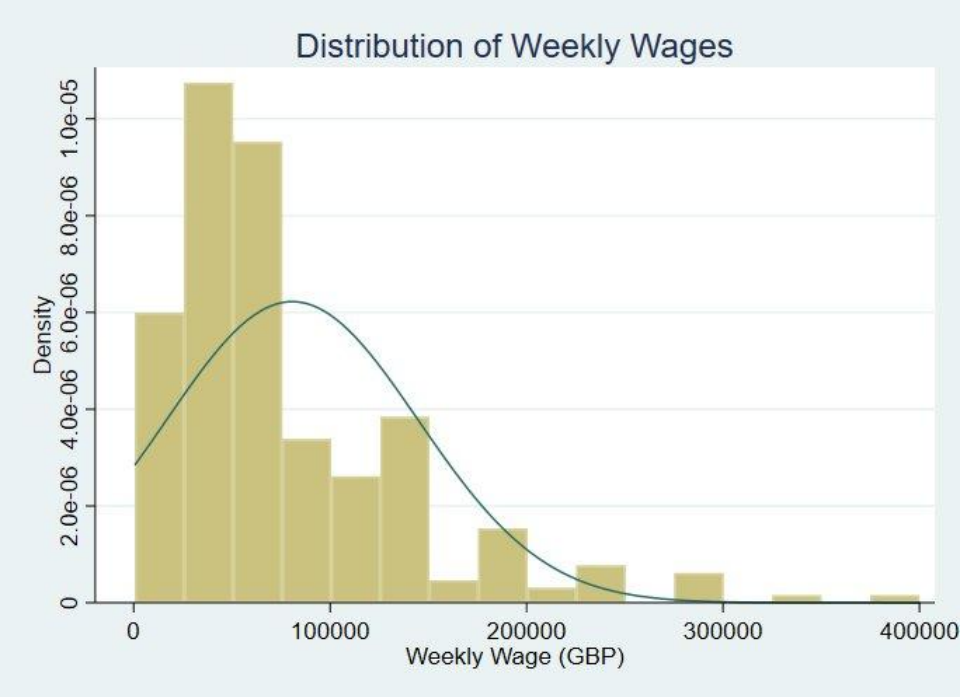

Figure 5: Distribution of Weekly Wages

Figure 5 shows the distribution of weekly wages. The distribution is clearly right-skewed, with the mode between 30,000 and 60,000 pounds and a long right tail past 300,000. This skewness motivates the log transformation used in Models 4 and 5.

Table 3: Correlation Matrix of Explanatory Variables

	goals	assists	mins	age	yellow	red	mf_fw	df_mf	top6
goals	1.000	0.471	0.542	0.082	0.244	0.096	0.294	-0.216	0.163
assists	0.471	1.000	0.536	0.089	0.219	0.059	0.031	-0.072	0.237
mins	0.542	0.536	1.000	0.082	0.658	0.226	-0.035	0.125	0.050
age	0.082	0.089	0.082	1.000	0.127	0.062	0.018	-0.010	-0.142
yellow	0.244	0.219	0.658	0.127	1.000	0.272	-0.198	0.171	-0.028
red	0.096	0.059	0.226	0.062	0.272	1.000	-0.046	0.134	0.036
mf_fw	0.294	0.031	-0.035	0.018	-0.198	-0.046	1.000	-0.258	0.112
df_mf	-0.216	-0.072	0.125	-0.010	0.171	0.134	-0.258	1.000	-0.057
top6	0.163	0.237	0.050	-0.142	-0.028	0.036	0.112	-0.057	1.000

Table 3 presents the pairwise correlations between all explanatory variables. The highest correlation is between minutes and yellow cards (r = 0.658), which is expected since players who play more minutes accumulate more bookings. Goals and assists are moderately correlated (r = 0.471), as are goals and minutes (r = 0.542) and assists and minutes (r = 0.536). Critically, the top6 dummy has low correlations with all variables (highest is r = 0.237 with assists), confirming

that it captures a distinct club-level dimension rather than proxying for player quality. No pairwise correlation exceeds 0.70 among the variables in our core specification, and all VIF values are below 3.1. This suggests that multicollinearity is unlikely to be a critical concern in the regression study.

Part 4: Regression Study

Table 2: OLS Regression Results — Dep. Var: Weekly Wage (M1-M3), Log(Weekly Wage) (M4-M5)

	M1: Simple	M2: Core+Top6	M3: Age²+Top6	M4: Log+Top6	M5: Log+Age²
goals		2,337	2,228	0.025	0.023
		(1,540)	(1,513)	(0.018)	(0.017)
assists	16,637***	8,218***	8,355***	0.036	0.038
	(1,955)	(1,969)	(1,935)	(0.024)	(0.023)
minutes		6.73	4.11	0.0003**	0.0002**
		(6.24)	(6.19)	(0.0001)	(0.0001)
age		5,355***	27,183***	0.078***	0.411***
		(729)	(6,884)	(0.009)	(0.072)
age_sq			-410**		-0.006***
			(129)		(0.001)
yellow		-622	-683	-0.029	-0.030
		(1,571)	(1,543)	(0.022)	(0.021)
red		9,373	11,811	-0.004	0.034
		(11,651)	(11,460)	(0.174)	(0.174)
pos_mf_fw		1,441	-797	-0.093	-0.127
		(7,030)	(6,941)	(0.129)	(0.126)
pos_df_mf		-15,964*	-16,724**	-0.218**	-0.230**
		(8,080)	(7,940)	(0.091)	(0.090)
top6		70,200***	71,072***	0.839***	0.852***
		(6,393)	(6,285)	(0.089)	(0.088)
const	56,145***	-98,452***	-378,900***	8.470***	4.193***
	(4,551)	(19,420)	(90,000)	(0.254)	(0.944)
N	261	261	261	261	261
Adj R^2	0.216	0.523	0.540	0.424	0.447
SER	56,776	44,275	43,475	0.630	0.617

BP p-val	0.0009	0.0000	—	0.3103	0.3437

*Notes: Standard errors in parentheses. * p<0.10, ** p<0.05, *** p<0.01. M1 is a bivariate regression of wage on assists only. M4-M5 use robust (HC1) standard errors. Reference groups: pure MF (position), Europe (region), non-Big 6 club. F-test for joint significance of position dummies: F(2, 251) = 2.08, p = 0.127. F-test for joint significance of region dummies: F(4, 247) = 0.24, p = 0.914. Top6 = Arsenal, Chelsea, Liverpool, Manchester City, Manchester United, Tottenham.*

Model 1: Simple Regression (Assists Only). We begin with a bivariate regression of weekly wages on assists to establish the baseline relationship. Assists is highly significant (b = 16,637, p < 0.001), with each assist associated with 16,637 pounds higher weekly wage. However, the adjusted R^2 of 0.216 indicates assists alone explain only 22% of wage variation. This estimate likely suffers from omitted variable bias, as it does not control for age, club status, or other performance measures.

Model 2: Core Model with Top 6 Dummy (9 variables). Adding the Big 6 club dummy to the core specification dramatically improves the model. The top6 coefficient is 70,200 (t = 10.98, p < 0.001), meaning that midfielders at Big 6 clubs earn approximately 70,200 pounds more per week than comparable players at other clubs, holding performance constant. This single variable increases the adjusted R^2 from 0.297 (without top6) to 0.523. Assists remains highly significant (b = 8,218, t = 4.17, p < 0.001), though its coefficient drops from 16,637 (M1) to 8,218, confirming the simple estimate was upward biased. Age is the second strongest predictor (b = 5,355, t = 7.35, p < 0.001). Goals are no longer significant (p = 0.130). This highlights the importance of controlling for omitted variables such as club status and reputation.

Model 3: Quadratic Age with Top 6. Adding age-squared further improves the model (Adj R^2 = 0.540). The positive age coefficient (27,183, p < 0.001) and negative age-squared coefficient (-410, p = 0.002) indicate wages peak at approximately age 33, then decline. The top6 premium remains virtually unchanged at 71,072 (p < 0.001). Defensive midfielders now earn significantly less (−16,724, p = 0.036).

Models 4 and 5: Log-Linear with Top 6. The Breusch-Pagan test rejects homoscedasticity in level models, so we estimate log-linear specifications. In the preferred Model 5, the top6

coefficient is 0.852 ($p < 0.001$), implying Big 6 midfielders earn approximately 134% more than non-Big 6 midfielders ($\exp(0.852) - 1 = 1.34$). Age shows a quadratic peak at age 33. Defensive midfielders earn approximately 23% less ($p = 0.010$). Minutes played becomes significant ($p = 0.009$). The log transformation resolves heteroscedasticity (BP $p = 0.34$), and the adjusted R^2 rises from 0.252 (without top6) to 0.447.

Logit Model. As a robustness check, we created a binary variable "high_wage" equal to 1 if a player earns above the median wage (65,000 pounds/week) and estimated a logit model. The results are consistent: top6, assists, and age are the strongest predictors of being a high earner.

Part 5: Specification Checks

Heteroscedasticity. The Breusch-Pagan test on Model 2 yields $chi^2 = 44.27$, $p < 0.001$, rejecting constant variance. The log transformation in Models 4 and 5 resolves this (BP $p = 0.31$ and $p = 0.34$ respectively). We report robust (HC1) standard errors for all log specifications.

Multicollinearity. The VIF test on Model 2 shows all values below 3.1 (goals 1.89, assists 1.67, minutes 3.06, age 1.06, yellow 2.03, red 1.10, pos_mf_fw 1.24, pos_df_mf 1.20, top6 1.12; Mean VIF = 1.60). The correlation matrix (Table 3) confirms no pairwise correlation exceeds 0.66 among core regressors. The top6 dummy's low VIF (1.12) indicates it captures a distinct dimension not already represented by performance variables.

Omitted Variable Bias. The dramatic improvement in adjusted R^2 from 0.297 to 0.523 when adding top6 strongly suggests that the original specification without club controls suffered from omitted variable bias. Club financial power is a major wage determinant that, when omitted, inflated the coefficients on assists and goals. The top6 dummy addresses this. Our model still explains only 45-54% of wage variation, indicating that important unobserved factors such as contract length, transfer fee, international caps, and commercial value remain omitted.

Joint Significance Tests. An F-test on position dummies yields $F(2, 251) = 2.08$, $p = 0.127$; position type does not jointly affect wages. An F-test on four region dummies yields $F(4, 247) = 0.24$, $p = 0.914$; geographic origin does not affect wages after controlling for performance and club.

Residual Analysis. The residual-vs-fitted plot (Figure 6) shows a fan-shaped pattern consistent with heteroscedasticity in the level model. The residual histogram (Figure 7) shows approximate normality with slight right skewness. The Durbin-Watson statistic is approximately 2.02, confirming no autocorrelation.

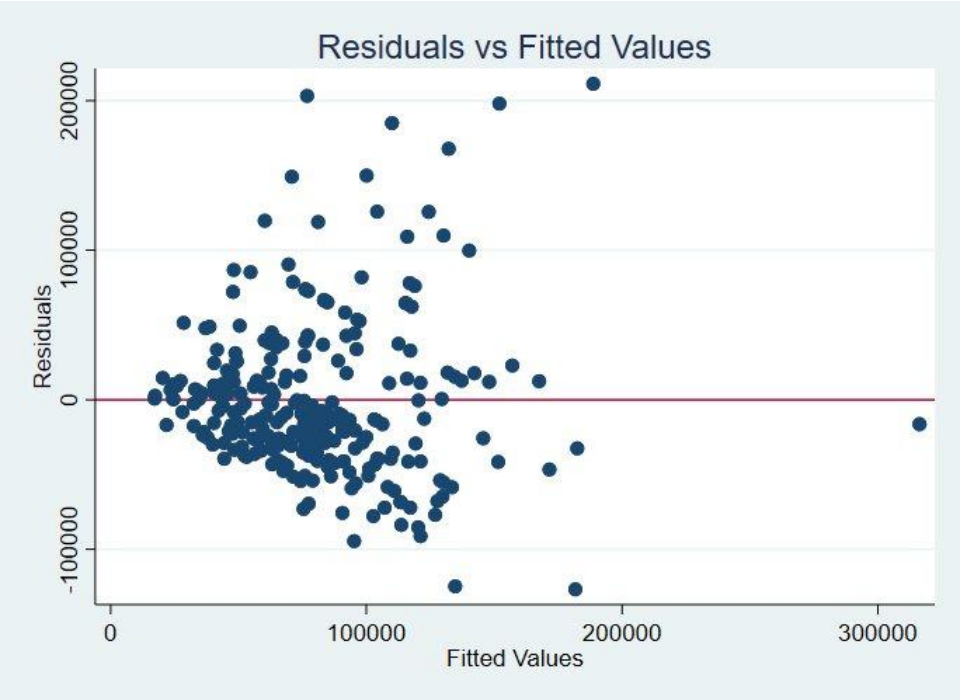

Figure 6: Residuals vs Fitted Values

Figure 6 displays residuals plotted against fitted values. The fan-shaped pattern, where residual variance increases with fitted values, is characteristic of heteroscedasticity. This motivates the log transformation in Models 4 and 5.

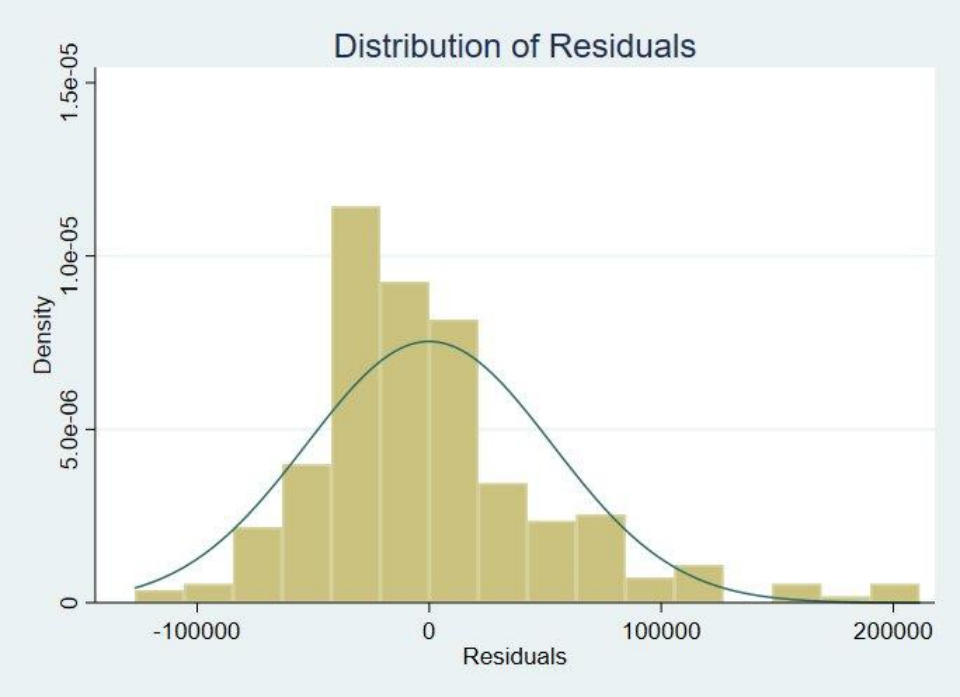

Figure 7: Distribution of Residuals

Figure 7 shows the distribution of OLS residuals with a normal density overlay. The residuals are approximately centred at zero with slight right skewness. The log transformation improves normality.

Part 6: Conclusion

This study examined the determinants of weekly wages for 261 Premier League midfielders in the 2024-25 season using OLS regression with five model specifications, building from a simple bivariate model to a log-quadratic specification with club controls.

Key Finding 1: Club status is the dominant wage determinant. The Big 6 dummy is the single most powerful predictor across all specifications. In the core model, Big 6 midfielders earn approximately 70,200 pounds more per week ($p < 0.001$), and in the log specification, approximately 134% more. Adding this variable alone increases the adjusted R^2 from 0.297 to 0.523. This is consistent with Sanchez et al. (2022), who found that Premier League club affiliation significantly drives market values.

Key Finding 2: Creative output remains the most important performance driver. Assists are consistently significant across all specifications with top6 included (b = 8,218 in the core model, $p < 0.001$). This confirms our hypothesis that playmaking ability commands the highest performance-based wage premium for midfielders, consistent with Ferreira (2021). The coefficient drops from 16,637 (M1) to 8,218 (M2), indicating the simple estimate was upward biased due to omitted club quality.

Key Finding 3: Age has a non-linear effect peaking at 33. Wages increase with age up to approximately 33, then decline. This is consistent with human capital theory. Frick (2011) similarly showed that career experience is a key salary determinant.

Key Finding 4: Geography and position type do not matter. Neither geographic origin ($F = 0.24$, $p = 0.914$) nor position type ($F = 2.08$, $p = 0.127$) jointly affects wages. Clubs pay for output and club status, not background.

Key Finding 5: Defensive midfielders earn less. The pos_df_mf dummy is significant in log specifications (−23%, $p = 0.010$ in Model 5), suggesting defensive roles command lower wages even after controlling for observable performance.

Limitations and Improvements. Our model explains 45-54% of wage variation. Possible omitted variables include contract length, transfer fee or market value, international appearances, social media following, and club revenue. Endogeneity may be present if higher-paid players receive more playing time. Future research could use panel data across multiple seasons, include transfer market values, and employ instrumental variable techniques.

Were our expectations confirmed? Yes, more strongly than before. Assists remain the primary performance-based wage driver for midfielders, and the Big 6 dummy reveals that club financial power is the single largest factor. Goals become insignificant once club status is controlled for, suggesting the original marginal significance ($p = 0.049$ without top6) was driven by omitted variable bias. Geographic origin does not affect wages, suggesting the Premier League labour market is relatively efficient. Overall, the results highlight that wages for midfielders are driven by both performance and club-level factors, with club status emerging as the most dominant influence.

Part 7: References

Bryson, A., Frick, B., & Simmons, R. (2013). The Returns to Scarce Talent: Footedness and Player Remuneration in European Soccer. Journal of Sports Economics, 14(6), 606-628.

Capology.com (2025). Premier League Salaries 2024-25. Retrieved from https://www.capology.com/uk/premier-league/salaries/

Carrieri, V., Principe, F., & Raitano, M. (2018). What Makes You 'Super-Rich'? New Evidence from an Analysis of Football Players' Wages. Oxford Economic Papers, 70(4), 950-973.

FBRef.com (2025). 2024-25 Premier League Player Statistics. Retrieved from https://fbref.com/en/comps/9/Premier-League-Stats

Ferreira, M. (2021). The Impact of Performance Measures in Football Players' Transfer Market Value [Master's dissertation, NOVA Information Management School, Universidade Nova de Lisboa].

Franceschi, M., Brocard, J.F., Follert, F., & Gouguet, J.J. (2024). Determinants of Football Players' Valuation: A Systematic Review. Journal of Economic Surveys, 38(3), 577-600.

Frick, B. (2007). The Football Players' Labor Market: Empirical Evidence from the Major European Leagues. Scottish Journal of Political Economy, 54(3), 422-446.

Frick, B. (2011). Performance, Salaries and Contract Length: Empirical Evidence from German Soccer. International Journal of Sport Finance, 6(2), 87-118.

Lucifora, C., & Simmons, R. (2003). Superstar Effects in Sport: Evidence from Italian Soccer. Journal of Sports Economics, 4(1), 35-55.

Sánchez, M., Orgaz, B., Ramirez-Campillo, R., Nakamura, F.Y., Luis-Pereira, J.M., Carretero, M., & Sánchez-Sánchez, J. (2022). Factors associated to the market value of professional soccer players. *Cuadernos de Psicología del Deporte, 22*(3), 140-150.