Soccer Players' Valuation From Top 5 European Leagues

109705010　資財13　羅鵬鎮

109705064　資財13　張俊彥

目錄

Introduction

Over the past decade, the football industry has experienced a profound transformation, and its growth has been remarkable. In Europe alone, the market size expanded by an impressive 32.57% within a five-year period, reaching a staggering 28.9 billion euros in the 2018-2019 season. One particularly captivating aspect of this industry is the transfer market, which plays a pivotal role. It is worth noting that the price tags attached to the most sought-after athletes can reach astronomical figures, surpassing the 100 million euro mark. This trend highlights the escalating financial dynamics and the immense value placed on top-tier players in today's football market.

The environment of a professional football transfer market is rather complex. In lengthy contracts, the relationships between players and their clubs are spelled out. Through the course of the contract, the club is required to pay the player a specific sum of money. The player's contract expires, making him a free agent who may negotiate new terms with other teams. However, the athlete is unable to switch teams until his contract expires without the consent of his current club. The majority of transactions are determined via discussions between the clubs. The transfer can take place if the player and his present club reach a deal with the second party. For the top and most promising players, the purchasing club typically pays a transfer fee, whose value might be quite high.

This report aims to determine the fair market value of football players based on their specific traits. To achieve this, we conducted an in-depth analysis using a comprehensive dataset of players from the top five European football leagues. We examined various player characteristics and investigated their correlation with player market values. As part of our modeling process, we considered factors such as age distribution, player position, and other relevant variables. Through this analysis, we aimed to gain valuable insights into the relative importance of these factors in determining the market value of football players.

In our model analysis, we categorized it based on the four positions in football: Goalkeepers, Defenders, Midfielders, and Forwards. It is expected that each position excels in different traits. For example, a proficient forward would have a higher number of goals or successful dribbles, while a skilled defender would have more team clean sheets or successful tackles.

名詞解釋

Common variables:

- 'games_starts': The number of matches where a player began the game as a starter.
- 'age': The age of the player, indicating how old they are.
- 'CL': Abbreviation for "Champions League," a prestigious soccer tournament.
- 'errors': Mistakes made by the player that negatively impact the team.

Forwards:

- 'Pts': Short for "points," which are earned by a team based on the outcomes of matches. It determines their position in the league standings.
- 'passes': The number of successful passes made by the player, indicating their ability to distribute the ball to teammates.
- 'tackles': The number of successful defensive challenges made by the player, where they dispossess the opponent or regain control of the ball.
- 'goals': The number of goals scored by the player, representing their ability to put the ball into the opposing team's net.

- 'dribbles': The number of successful individual runs with the ball by the player, showcasing their skill in maneuvering past opponents.
- 'touches_att_pen_area': The number of times the player has made contact with the ball inside the opposing team's penalty area.
- 'gca': Abbreviation for "Goal-Creating Actions," which refers to the player's involvement in the sequence of actions leading to a goal, either through scoring or assisting.

Midfielders:

- 'carries': The number of times the player moves the ball while in possession, usually by dribbling or running with it.
- 'pressures_att_3rd': The number of defensive actions initiated by the player in the attacking third of the field, aimed at disrupting the opposition's play.
- 'through_balls': The number of passes played by the player that go through the opposing team's defense and create scoring opportunities for teammates.
- 'goals': The number of goals scored by the player, representing their ability to put the ball into the opposing team's net.
- 'Pts': Short for "points," which are earned by a team based on the outcomes of matches. It determines their position in the league standings.
- 'assists': The number of passes made by the player that directly lead to a goal scored by a teammate.
- 'passes': The number of successful passes made by the player, indicating their ability to distribute the ball to teammates.

Defenders:

- 'clearances': Actions taken by the player to remove the ball from their team's defensive area.

- 'passes_ground': Successful passes made by the player along the ground.
- 'touches_att_pen_area': The number of times the player has made contact with the opposing team's penalty area in an attacking position.
- 'touches_def_pen_area': The number of times the player has made contact with their own team's penalty area in a defensive position.
- 'aerials_won_pct': The percentage of aerial duels won by the player, which are contests for the ball in the air.

Goalkeepers:

- 'psnpxg_per_shot_on_target_against': Expected goals (xG) per shot on target faced by the goalkeeper, which measures the quality of shots they face.
- 'clean_sheets': The number of matches in which the goalkeeper and their team prevented the opposing team from scoring any goals.
- 'wins_gk': The number of matches won by the goalkeeper's team.
- 'passes_pct_launched_gk': The percentage of passes made by the goalkeeper that are long, typically launched towards the other end of the field.

Hypothesis

We set four hypotheses for each different position:

For a Forward we wondered if the number of goals and goal creations would affect the value of a player the most, since we assumed that the main task of a attacking player should be to score as many goals as possible and to create chances for the team to win, at the end on the day, in order to win a football match you need to score more goals than your rivals.

For Midfielders we decided to set the more assists and passes the more valuable the player is, because midfielders are essentially the players who connect the defenders with the attacker, so they should have contact with the ball more often and provide assists to the Forwards.

The Defenders are the players on the pitch who try to stop the opposing attackers from scoring, naturally our first thoughts were, the more clearances and passes_blocked the more valuable the player is.

As for the Goalkeepers, they are the only players allowed to use their hands, and their main job is to stop the ball from going into the goal, so naturally we would wonder if the number of saves have a positive impact on the player's market price?

Research Method

(一)研究資料定義與來源

本文研究目的為探討影響歐洲足球五大聯賽中球員身價的重要因子，因此我們使用球員的身價(以歐元計並取 log)作為應變數，球員在場上的各種表現數據做為自變數，放入 OLS model 中預測球員的身價。

由於不同位置的球員之間身價有顯著的差異，要關注的表現數據也大不相同，因此我們會按照位置分成四個部分來分別討論。而針對上述提到的球員表現數據與身價，我們是大多是從 Transfermarket 上取得，有一些更進階的表現數據則是透過 Kaggle 上喜愛足球的粉絲整理好的資料集取得，同樣也是透過 Transfermarket 來取得，不過經過了相對複雜許多的整理才方便使用。

資料涵蓋的部分則是包含 2017~2020 三個賽季的數據，並且只取身價在100萬歐元以上的球員，因為我們認為身價在100萬歐元以下的球員，他們場上的表現數據其實不能完全的反映在身價上，因為他們上場的時間不多，表現也不太穩定，對於球隊並不是很重要，而且他們的數量很多，也會影響到模型的結果，所以我們主要關注的群體為有一定身價的球員。

(二)研究變數

對於不同位置的球員，需要關注的數據表現都不一樣，前面名詞解釋中就是研究中有使用到的變數。為避免重複論述，這裡只會提到我們對於所有球員都有放置的共同變數：

先發場數(game_starts)、失誤次數(errors)、年齡(age)以及所屬球隊是否參與歐冠比賽(CL)

前三項對於球員的身價影響都是顯而易見的，另外多放所屬球隊是否參與歐冠比賽(CL)的原因是，假如所在球隊能夠進入歐冠，代表整支球隊都相當有競爭力，球員的水平都有一定保證，我們認為在這樣的情況下，場上的表現數據更能直接的反映在球員身價上，而較不會有因為隊友表現差而跟著被拖累的情況發生。

Result Analysis

(一)前鋒 Forwards

OLS Regression Results

Dep. Variable:	value	R-squared:	0.624
Model:	OLS	Adj. R-squared:	0.619
Method:	Least Squares	F-statistic:	131.1

	coef	std err	t	P>|t|	[0.025	0.975]
const	16.0910	0.182	88.238	0.000	15.733	16.449
Pts	0.0212	0.002	10.146	0.000	0.017	0.025
passes	0.0004	0.000	2.569	0.010	8.7e-05	0.001
tackles	-0.0075	0.002	-3.151	0.002	-0.012	-0.003
goals	0.0250	0.008	3.128	0.002	0.009	0.041
dribbles	0.0017	0.001	1.889	0.059	-6.65e-05	0.003
touches_att_pen_area	0.0033	0.001	3.565	0.000	0.001	0.005
gca	0.0095	0.008	1.193	0.233	-0.006	0.025
games_starts	0.0144	0.006	2.440	0.015	0.003	0.026
age	-0.0815	0.006	-13.089	0.000	-0.094	-0.069
CL	0.3657	0.083	4.391	0.000	0.202	0.529
errors	-0.0563	0.080	-0.700	0.484	-0.214	0.102

The Forwards' model gave us an adjusted R squared of 0.619, we decided to use seven different variables and four common ones in all models giving us a total of eleven variables for the model.

For the common variables shown in the last four lines, we can see that Forward errors made are not relevant, on the other hand, the other three common variables show a strong correlation with a player's market value, with age being negatively correlated and CL positively. Moreover, for our hypothesis we specifically wanted to add 'goals' and 'gca' because we felt like they would be the main indicators for a Forward. As seen in the figure above, only goals show a significant relation, whereas gca showed to be not significant surprisingly.

We also wanted to see other variables, and as expected both dribbles and touches at penalty areas are important variables for a forward. Although their coefficients are not as big of a determinant, they are still positively correlated with the player's market value.

For tackles, passes, and pts, we can see that the three variables show a strong correlation with the model, although tackles relate negatively with the model, we think that this is because the usual job for a forward is to attack, so if they are helping defensively that could mean that the team is being surpassed by the other team.

（二）中場 Midfielders

OLS Regression Results			
Dep. Variable:	value	R-squared:	0.578
Model:	OLS	Adj. R-squared:	0.574
Method:	Least Squares	F-statistic:	142.3

	coef	std err	t	P>|t|	[0.025	0.975]
const	15.9918	0.174	91.889	0.000	15.650	16.333
carries	0.0004	0.000	2.400	0.017	8.11e-05	0.001
pressures_att_3rd	0.0005	0.001	0.724	0.469	-0.001	0.002
through_balls	0.0244	0.010	2.425	0.015	0.005	0.044
goals	0.0549	0.012	4.479	0.000	0.031	0.079
Pts	0.0256	0.002	12.760	0.000	0.022	0.030
assists	-0.0010	0.013	-0.073	0.942	-0.027	0.025
passes	0.0002	0.000	0.957	0.339	-0.000	0.000
games_starts	0.0050	0.006	0.862	0.389	-0.006	0.016
age	-0.0864	0.006	-14.275	0.000	-0.098	-0.074
CL	0.1845	0.082	2.250	0.025	0.024	0.345
errors	-0.0004	0.041	-0.009	0.993	-0.080	0.079

The Midfielders' model gave us an adjusted R square of 0.574, and similar to the Forwards' model we used seven different variables and four common ones.

For the common variables in the bottom we can see that 'age' and 'CL' are clearly significant while 'game_starts' and 'errors' are not significant, different from the Forward's model, starting a game is not as important for Midfielders for some reason, we are not entirely sure why this happened, but we think this is due to a midfielders performance being most of the time all rounded even when entering onto the game later on.

Contrary to our initial hypothesis, both assists and passes a midfielder does are not significant, even showing negative relation with the model in assists. This goes completely against our knowledge on football which may imply a problem in the model or in the data. We think that the main reason is because most assists may be done by forwards, since a team might play with three or even four people in the attacking area, making them even more likely to have an assist.

To our surprise 'through_balls' and 'goals' resulted significantly in determining the market value of a Midfielder. This might be due to midfielders being mostly the position who can perform this type of passes.

Finally we can see in the rest of the variables, 'Pts', 'carries', and 'pressures_att_3rd', that the first two mentioned are significant while the last one is not. Pts being the only variable out of these three with a greater coefficient.

(三)後衛 Defenders

假說:清球(clearances)次數越多，對球員身價有正面影響

OLS Regression Results

Dep. Variable:	value	R-squared:	0.535
Model:	OLS	Adj. R-squared:	0.532
Method:	Least Squares	F-statistic:	199.9
Date:	Sat, 27 May 2023	Prob (F-statistic):	1.02e-252
Time:	15:28:57	Log-Likelihood:	-1598.4
No. Observations:	1576	AIC:	3217.
Df Residuals:	1566	BIC:	3270.
Df Model:	9		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	16.7727	0.141	119.243	0.000	16.497	17.049
games_starts	-0.0235	0.004	-5.720	0.000	-0.032	-0.015
age	-0.0816	0.005	-17.227	0.000	-0.091	-0.072
CL	0.8336	0.049	16.973	0.000	0.737	0.930
errors	-0.0202	0.018	-1.109	0.268	-0.056	0.016
clearances	0.0015	0.001	2.206	0.028	0.000	0.003
passes_ground	0.0010	7.96e-05	12.286	0.000	0.001	0.001
touches_att_pen_area	0.0136	0.001	10.945	0.000	0.011	0.016
touches_def_pen_area	0.0014	0.001	2.701	0.007	0.000	0.002
aerials_won_pct	0.0020	0.001	1.421	0.156	-0.001	0.005

除了失誤次數(errors)和空中爭球勝率(aerials_won_pct)統計上不顯著外，其餘皆顯著。而較令人意外的是先發場數的係數是負的，代表先發越多次反而對身價是負面的影響，我們認為可能是後衛這個位置，負責的事情是面對對方的進攻，而阻止進攻本身就是十分耗力的事情，因為次數多而且都需要不斷奔跑跟上對手，因此常常需要使用替補上場，而考量到對手的體力也會消耗，所以最優秀的後衛選手不一定會在一開始就上場，而是會用輪替的方式來防守對方。

而關於我們想探討的假說，也可以看到清球次數的係數為正，符合我們預期的假說內容。雖然清球可能給予對手其他好的進攻機會而間接造成球隊失分，但整體而言還是對後衛球員的身價有正面的影響。

(四)守門員 Goalkeepers

假說:撲救的次數越多，對球員的身價有正面的影響

從下圖的結果來看，可以發現只有年齡、是否參加歐冠(CL)和上場時球隊獲勝次數在統計上是顯著的，其餘皆不顯著，而且我們想要關注的撲救次數也是不顯著。

OLS Regression Results

Dep. Variable:	value	R-squared:	0.615
Model:	OLS	Adj. R-squared:	0.603
Method:	Least Squares	F-statistic:	50.91

	coef	std err	t	P>|t|	[0.025	0.975]
const	17.1543	0.361	47.530	0.000	16.444	17.865
games_starts	0.0159	0.011	1.487	0.138	-0.005	0.037
age	-0.0993	0.011	-9.341	0.000	-0.120	-0.078
CL	0.4067	0.135	3.018	0.003	0.141	0.672
saves	-0.0008	0.002	-0.369	0.713	-0.005	0.004
clean_sheets	0.0258	0.018	1.463	0.145	-0.009	0.061
wins_gk	0.0644	0.014	4.672	0.000	0.037	0.092
errors	-0.0442	0.041	-1.070	0.286	-0.126	0.037
passes_pct_launched_gk	0.0010	0.005	0.188	0.851	-0.010	0.012

因此我們嘗試將撲救次數換成撲救成功率，觀察是否會有變化。從下圖可以看到只有先發次數變得稍微顯著一些，包含撲救成功率在內的其他表現依然是不顯著。

	coef	std err	t	P>|t|	[0.025	0.975]
const	17.3075	0.406	42.625	0.000	16.508	18.107
games_starts	0.0124	0.007	1.749	0.082	-0.002	0.026
age	-0.0994	0.011	-9.359	0.000	-0.120	-0.078
CL	0.4086	0.134	3.049	0.003	0.145	0.673
save_pct	-0.3103	0.479	-0.648	0.518	-1.253	0.633
clean_sheets	0.0300	0.018	1.632	0.104	-0.006	0.066
wins_gk	0.0648	0.014	4.755	0.000	0.038	0.092
errors	-0.0460	0.041	-1.112	0.267	-0.127	0.035
passes_pct_launched_gk	0.0025	0.006	0.409	0.683	-0.009	0.014

為了解決不顯著的問題，我們將想要探討的撲救更改為射門射正時被進的球數(psnpxg_per_shot_on_target_against)，因為這兩個數據表示的都是守門員的防守能力，而經過查詢資料後，我們認為使用射門射正時被進的球數應該可以讓模型有更好的表現。

更換變數後的回歸結果如同下圖，可以看到R-square有增加至0.623，讓模型的解釋力增加了一些，而且射門射正時被進的球數統計上也是呈現顯著，其係數為負也符合常理，代表被進越多球對於守門員的身價有負面的影響。

OLS Regression Results

Dep. Variable:	value	R-squared:	0.623
Model:	OLS	Adj. R-squared:	0.611
Method:	Least Squares	F-statistic:	52.66
Date:	Sat, 27 May 2023	Prob (F-statistic):	8.68e-50
Time:	15:28:59	Log-Likelihood:	-241.12
No. Observations:	264	AIC:	500.2
Df Residuals:	255	BIC:	532.4
Df Model:	8		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	17.5536	0.385	45.567	0.000	16.795	18.312
games_starts	0.0132	0.007	1.888	0.060	-0.001	0.027
age	-0.0994	0.011	-9.455	0.000	-0.120	-0.079
CL	0.3826	0.133	2.873	0.004	0.120	0.645
psnpxg_per_shot_on_target_against	-1.9996	0.849	-2.356	0.019	-3.671	-0.328
clean_sheets	0.0269	0.017	1.548	0.123	-0.007	0.061
wins_gk	0.0627	0.014	4.634	0.000	0.036	0.089
errors	-0.0388	0.041	-0.948	0.344	-0.119	0.042
passes_pct_launched_gk	0.0058	0.006	1.038	0.300	-0.005	0.017

(五)加入聯賽固定效果

因為各個位置的球員狀況相似，所以我們只列出在足球中最具代表性的位置，前鋒(Forwards)的回歸結果。

serieA:義大利足球甲級聯賽

從下圖的結果可以看到，iserieA是不顯著的，而其他的變數只有盤球成功數(dribbles)統計量變成了不顯著。

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  value   R-squared:                       0.628
Model:                            OLS   Adj. R-squared:                  0.624
Method:                 Least Squares   F-statistic:                     133.4
Date:                Sat, 27 May 2023   Prob (F-statistic):          8.03e-194
Time:                        15:28:55   Log-Likelihood:                -965.33
No. Observations:                 959   AIC:                             1957.
Df Residuals:                     946   BIC:                             2020.
Df Model:                          12
Covariance Type:            nonrobust
========================================================================================
                           coef    std err          t      P>|t|      [0.025      0.975]
----------------------------------------------------------------------------------------
const                   16.0924      0.181     88.772      0.000      15.737      16.448
Pts                      0.0216      0.002     10.443      0.000       0.018       0.026
passes                   0.0004      0.000      2.517      0.012    8.03e-05       0.001
tackles                 -0.0084      0.002     -3.480      0.001      -0.013      -0.004
age                     -0.0409      0.003    -13.363      0.000      -0.047      -0.035
goals                    0.0205      0.008      2.640      0.008       0.005       0.036
dribbles                 0.0012      0.001      1.373      0.170      -0.001       0.003
touches_att_pen_area     0.0035      0.001      3.866      0.000       0.002       0.005
gca                      0.0089      0.008      1.147      0.252      -0.006       0.024
games_starts             0.0167      0.006      2.831      0.005       0.005       0.028
age                     -0.0409      0.003    -13.363      0.000      -0.047      -0.035
CL                       0.3669      0.082      4.454      0.000       0.205       0.529
errors                  -0.0576      0.080     -0.717      0.474      -0.215       0.100
isSerieA                -0.0702      0.057     -1.226      0.221      -0.183       0.042
```

PremierLeague:英格蘭足球超級聯賽

從下表的結果可以看到，isPremierLeague是顯著的，而且其他變數的統計量也都沒有太大的變化。

OLS Regression Results

Dep. Variable:	value	R-squared:	0.676
Model:	OLS	Adj. R-squared:	0.672
Method:	Least Squares	F-statistic:	164.8
Date:	Sat, 27 May 2023	Prob (F-statistic):	5.14e-222
Time:	15:28:55	Log-Likelihood:	-899.15
No. Observations:	959	AIC:	1824.
Df Residuals:	946	BIC:	1888.
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	16.0695	0.169	94.975	0.000	15.737	16.402
Pts	0.0204	0.002	10.571	0.000	0.017	0.024
passes	0.0004	0.000	2.780	0.006	0.000	0.001
tackles	-0.0107	0.002	-4.781	0.000	-0.015	-0.006
age	-0.0418	0.003	-14.631	0.000	-0.047	-0.036
goals	0.0260	0.007	3.581	0.000	0.012	0.040
dribbles	0.0016	0.001	1.882	0.060	-6.83e-05	0.003
touches_att_pen_area	0.0022	0.001	2.592	0.010	0.001	0.004
gca	0.0114	0.007	1.568	0.117	-0.003	0.026
games_starts	0.0198	0.005	3.608	0.000	0.009	0.031
age	-0.0418	0.003	-14.631	0.000	-0.047	-0.036
CL	0.4127	0.077	5.385	0.000	0.262	0.563
errors	-0.0151	0.075	-0.201	0.841	-0.162	0.132
isPremierLeague	0.5767	0.048	11.905	0.000	0.482	0.672

Ligue1:法國足球甲級聯賽

從下圖的結果可以看到，isLigue1是顯著的，而其他變數的統計量也都沒有太大的變化。

OLS Regression Results

Dep. Variable:	value	R-squared:	0.641
Model:	OLS	Adj. R-squared:	0.637
Method:	Least Squares	F-statistic:	141.0
Date:	Sat, 27 May 2023	Prob (F-statistic):	5.11e-201
Time:	15:28:55	Log-Likelihood:	-948.44
No. Observations:	959	AIC:	1923.
Df Residuals:	946	BIC:	1986.
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	16.2633	0.180	90.174	0.000	15.909	16.617
Pts	0.0203	0.002	9.949	0.000	0.016	0.024
passes	0.0003	0.000	2.365	0.018	5.72e-05	0.001
tackles	-0.0077	0.002	-3.279	0.001	-0.012	-0.003
age	-0.0421	0.003	-13.987	0.000	-0.048	-0.036
goals	0.0235	0.008	3.077	0.002	0.009	0.039
dribbles	0.0017	0.001	1.907	0.057	-5.01e-05	0.003
touches_att_pen_area	0.0029	0.001	3.321	0.001	0.001	0.005
gca	0.0099	0.008	1.290	0.197	-0.005	0.025
games_starts	0.0170	0.006	2.956	0.003	0.006	0.028
age	-0.0421	0.003	-13.987	0.000	-0.048	-0.036
CL	0.3827	0.081	4.747	0.000	0.224	0.541
errors	-0.0581	0.079	-0.736	0.462	-0.213	0.097
isLigue1	-0.3401	0.057	-5.957	0.000	-0.452	-0.228

LaLiga:西班牙足球甲級聯賽

從下表可以看到，isLaLiga是顯著的，但是盤球成功次數也從顯著變成了不顯著。

OLS Regression Results

Dep. Variable:	value	R-squared:	0.633
Model:	OLS	Adj. R-squared:	0.628
Method:	Least Squares	F-statistic:	135.7
Date:	Sat, 27 May 2023	Prob (F-statistic):	4.53e-196
Time:	15:28:55	Log-Likelihood:	-960.05
No. Observations:	959	AIC:	1946.
Df Residuals:	946	BIC:	2009.
Df Model:	12		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	16.0572	0.181	88.914	0.000	15.703	16.412
Pts	0.0220	0.002	10.679	0.000	0.018	0.026
passes	0.0004	0.000	2.444	0.015	6.92e-05	0.001
tackles	-0.0088	0.002	-3.670	0.000	-0.013	-0.004
age	-0.0399	0.003	-13.071	0.000	-0.046	-0.034
goals	0.0192	0.008	2.477	0.013	0.004	0.034
dribbles	0.0014	0.001	1.542	0.123	-0.000	0.003
touches_att_pen_area	0.0031	0.001	3.392	0.001	0.001	0.005
gca	0.0084	0.008	1.085	0.278	-0.007	0.024
games_starts	0.0198	0.006	3.354	0.001	0.008	0.031
age	-0.0399	0.003	-13.071	0.000	-0.046	-0.034
CL	0.3781	0.082	4.634	0.000	0.218	0.538
errors	-0.0546	0.080	-0.684	0.494	-0.211	0.102
isLaLiga	-0.1774	0.051	-3.464	0.001	-0.278	-0.077

Bundesliga:德國甲級足球聯賽

從下表可以看到，isBundesliga是不顯著的，一樣是盤球成功次數從顯著變成不顯著。

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  value   R-squared:                       0.629
Model:                            OLS   Adj. R-squared:                  0.624
Method:                 Least Squares   F-statistic:                     133.5
Date:                Sat, 27 May 2023   Prob (F-statistic):          5.87e-194
Time:                        15:28:55   Log-Likelihood:                -965.02
No. Observations:                 959   AIC:                             1956.
Df Residuals:                     946   BIC:                             2019.
Df Model:                          12
Covariance Type:            nonrobust
=======================================================================================
                          coef    std err          t      P>|t|      [0.025      0.975]
---------------------------------------------------------------------------------------
const                  16.1533      0.186     86.997      0.000      15.789      16.518
Pts                     0.0210      0.002     10.088      0.000       0.017       0.025
passes                  0.0003      0.000      2.409      0.016    6.45e-05       0.001
tackles                -0.0079      0.002     -3.296      0.001      -0.013      -0.003
age                    -0.0415      0.003    -13.492      0.000      -0.048      -0.035
goals                   0.0217      0.008      2.783      0.005       0.006       0.037
dribbles                0.0012      0.001      1.269      0.205      -0.001       0.003
touches_att_pen_area    0.0033      0.001      3.665      0.000       0.002       0.005
gca                     0.0107      0.008      1.372      0.170      -0.005       0.026
games_starts            0.0167      0.006      2.851      0.004       0.005       0.028
age                    -0.0415      0.003    -13.492      0.000      -0.048      -0.035
CL                      0.3900      0.083      4.724      0.000       0.228       0.552
errors                 -0.0520      0.080     -0.647      0.518      -0.210       0.106
isBundesliga           -0.0853      0.058     -1.461      0.144      -0.200       0.029
```

Conclusion

(一)前鋒 Forwards

進球數對於前鋒球員的身價有正面的影響，符合我們假說預期；但創造射門機會統計上並不顯著，我們無從判斷和解釋。

(二)中場 Midfielders

助攻和傳球次數越多，無法判斷是否有影響，因為兩的變數的統計量皆為不顯著；但我們發現進球數和空檔傳球數越多則會對球員身價有正面的影響，原本我們認為進球得分的任務應該是前鋒的任務，中場球員應該較少直接參與到球門前的進攻，結果令人意外的是進球數對於中場球員的身價具有影響。

(三)後衛 Defenders

清球次數越多，對後衛的身價具有正面的影響，和假說一致。雖然清球可能會給予對手其他進攻機會，但從結果上推論，先解決當下緊急的情況是更為重要的事情，即使把球踢出界外，也可以透過重新發球的時間來重整防守布陣，抵擋進攻。

(四)守門員 Goalkeepers

撲救次數越多，由於統計量不顯著，無法判斷是否有影響；但面對射門射正時沒能攔下的球數越多，對於身價有負面的影響。

很可惜沒有考慮到球員在球隊中和其他球員以及教練之間的關係為何，因為這也可能會影響到球員在場上的表現，進而影響到身價。另外足球場上的統計數據其實還有不少，但是礙於篇幅限制和避免共線性，我們是從不同面向(進攻、防守、傳球......)的數據中，依據其他參考資料、個人判斷以及回歸結果的變數統計量，挑選出具代表性的表現數據來放入模型中。

另外我們有多加入不同聯賽的固定效果，來觀察模型結果的變化，我們發現英超、法國、西班牙聯賽的變數統計量皆為顯著，義大利和德國聯賽的統計量為不顯著。而其他變數的統計量變化都不大。

Reference

https://www.transfermarkt.com/

https://www.kaggle.com/datasets/kriegsmaschine/soccer-players-values-and-their-statistics

https://www.whoscored.com/Statistics

https://www.transfermarkt.com/marktwertetop/wertvollstespieler

https://www.footballbenchmark.com/library/player_valuation_update_young_talent_dominate_the_ranking_of_most_valuable_football_players

https://theathletic.com/3085749/2022/01/27/premier-league-how-do-you-value-a-player/

https://www.colossusbets.com/blog/market-value-of-football-players/