Term Paper

on

Factors determining the value of football players in the Premier League

by

Chakkarin Laorit	6340020829
Poon Duangpanya	6340080429

This term paper is a part of assessment of
Introduction to Econometrics course.

Second semester, Academic Year 2021
Faculty of Economics, Chulalongkorn University

1. Introductions - Problem and Significance

Football is the most popular sport in the world today. Football games and other activities related to football are also involved in many aspects of the global economy, especially in the travel and leisure sector. Tourists from all over the world travel to watch a match of football at the field and that creates a huge amount of economic activities, such as the tickets for the spectators, the gifts sold, the tours of the football clubs, and the traveling cost to watch football matches.

According to the study of Adam Metelski in 2021, the value of the global football markets in 2018 was as high as 488.5 billion USD. The football clubs are also growing their worth, as there was only one club that was worth more than 2 billion USD in 2012, but it grew to as many as 52 clubs in 2021. Real potential of the football markets is also evident as there were as many as over a billion watchers of the final match of FIFA's world championship in 2018.

The football club is, in fact, an enterprise which seeks for profit. While the football players play an important role in the enterprise's business, they are considered as human resources of the club. (Adiwiyana and Harymawan, 2021) Despite how hard the valuation on the human resources is, the market values of football players that are transferred nowadays are as high as hundred million Euros. (Metelski, 2021). With this skyscraping high transfer value makes it worth the difficulties to try to value football players.

The Premier League is the top-ranked football league in England, and almost the top of the European leagues. Many great performance football players in this league are worth more than a hundred million Euros (Sofifa.com, 2022) Trying to evaluate the market value of these football players based on their empiric performance should make the football clubs decide better when they have to transfer a player.

This term paper study tries to find the relationship between factors determining the Premier League football players' performance, such as age, body size, overall rating score, and the positions they play, and their market transferring value. This should contribute to making the valuation of football clubs in the Premier League easier and more accurate, helps the clubs decide more correctly in transferring each player, and helps the investor decide more precisely in investing in these clubs.

2. Data and Model

2.1 Variables in the Model

Table 1: Variables included in the model

Variables	Description	Source
Dependent Variable		
lnVALUE	Natural logarithm of value of Premier League football players in million Euros	www.sofifa.com on April 13, 2022
Independent Variables		
AGE	Age of Premier League football players in years	www.sofifa.com on April 13, 2022
HEIGHT	Height of Premier League football players in centimeters	www.sofifa.com on April 13, 2022
WEIGHT	Weight of Premier League football players in kilograms	www.sofifa.com on April 13, 2022
OVA	Overall rating score of Premier League football players	www.sofifa.com on April 13, 2022
Dummy Variables		
DEFENDER	Dummy variable of football players in defender position	www.sofifa.com on April 13, 2022
MIDFIELDER	Dummy variable of football players in midfielder position	www.sofifa.com on April 13, 2022
GOALKEEPER	Dummy variable of football players in goalkeeper position	www.sofifa.com on April 13, 2022

2.2 Coefficients in the model

Table 2: Coefficients included in the model

Parameters	Variables	Meaning
β_0	_cons	Expected value of lnVALUE when all dependent variables are zero
β_1	AGE	Percentage change of value of Premier League football players when their age increase by 1 year
β_2	HEIGHT	Percentage change of value of Premier League football players when their height increase by 1 centimeter
β_3	WEIGHT	Percentage change of value of Premier League football players when their weight increase by 1 kilogram
β_4	OVA	Percentage change of value of Premier League football players when their overall rating score increase by 1
β_5	DEFENDER	Difference of expected value of lnVALUE between football player in defender and striker position
β_6	MIDFIELDER	Difference of expected value of lnVALUE between football player in midfielder and striker position
β_7	GOALKEEPER	Difference of expected value of lnVALUE between football player in goalkeeper and striker position

2.3 Expected sign of each coefficients in the model

Table 3: Expected sign of each coefficients in the model

Parameters	Expected sign	Reasons
β_1	-	Younger football players, thus the longer working time, tend to value more.
β_2	+	Football players with larger bodies tend to value more.
β_3	-	Football players with heavier bodies, thus less agility, tend to value less.
β_4	+	Football players with higher overall rating score, thus the higher performance, tend to value more.
β_5	-	Strikers tend to value more than defenders.
β_6	-	Strikers tend to value more than midfielders.
β_7	-	Strikers tend to value more than goalkeepers.

2.4 Model

The model we used in this term paper's analysis focuses on the value of Premier League football players as a dependent variable. Data used in this analysis is limited to the secondary data summarized in sofifa.com website, which is the website that gathers information about football players in almost every well-known league. We collect data just from the sample of players that their information is available on the website. We collect a sample of 5 players from each of the 20 clubs in the league.

Value of football players should depend mostly on how much they can earn for the club from now to the future, which obviously depends on the performance of the football players and times that the football players can still play for the club. According to this fact, we include ages, sizes of body (as heights and weights), and the overall rating scores, which imply the overall performance of football players in the games, as independent variables to the model.

Moreover, the value of football players depends generally on the position in which the players play in the games. In order to examine the difference of value of football players among positions, we also include the dummy variables for defenders, midfielders, and goalkeepers (strikers are the omitted group).

By these steps, we construct the model as follows.

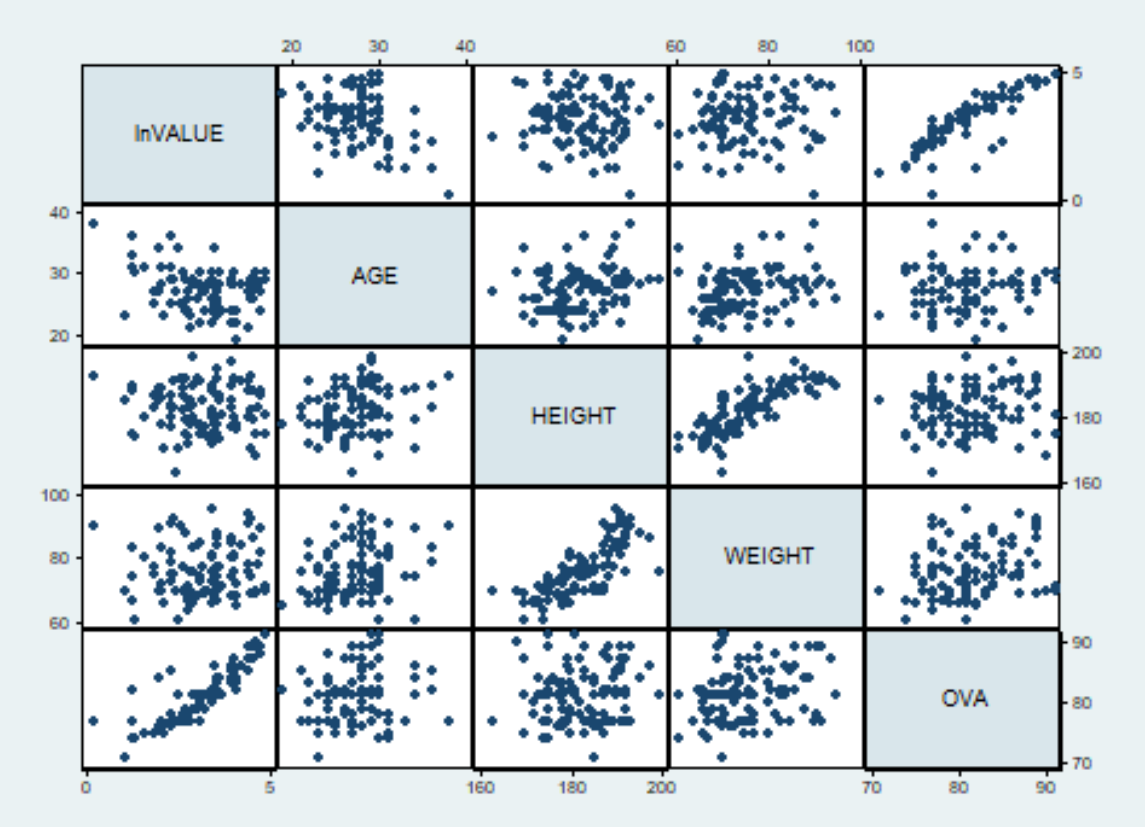
$$\begin{aligned} \ln VALUE_i = {} & \beta_0 + \beta_1 AGE_i + \beta_2 HEIGHT_i + \beta_3 WEIGHT_i + \beta_4 OVA_i \\ & + \beta_5 DEFENDER + \beta_6 MIDFIELDER + \beta_7 GOALKEEPER + \varepsilon_i \end{aligned}$$

3. Descriptive Statistics

Table 4: Summary of statistical value of each variable

Variable	Observations	Mean	Standard deviation	Minimum	Maximum
VALUE	100	35.767	31.33588	1.2	129
lnVALUE	100	3.167116	0.980725	0.1823216	4.859812
AGE	100	27.19	3.428034	19	38
HEIGHT	100	182.38	7.151718	163	199
WEIGHT	100	75.82	8.021914	61	95
OVA	100	81.25	4.416166	71	91

Table 5 : Scatterplot matrix between each variable in the model



4. Regression Analysis

We use the OLS (ordinary least square) method by STATA software to estimate the coefficients of each variable in the model. The results are as follows.

Source	SS	df	MS	Number of obs =	100
				F(7, 92) =	277.26
Model	90.9109734	7	12.9872819	Prob > F =	0.0000
Residual	4.30935805	92	.046840848	R-squared =	0.9547
				Adj R-squared =	0.9513
Total	95.2203315	99	.96182153	Root MSE =	.21643

lnVALUE	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
AGE	-.1181743	.007186	-16.45	0.000	-.1324462	-.1039023
HEIGHT	.0029312	.0052719	0.56	0.580	-.0075394	.0134017
WEIGHT	-.0002419	.004507	-0.05	0.957	-.0091932	.0087093
OVA	.2030182	.0051873	39.14	0.000	.1927158	.2133206
1.DEFENDER	-.1712935	.0659601	-2.60	0.011	-.3022959	-.0402912
1.MIDFIELDER	-.1435215	.0611281	-2.35	0.021	-.2649272	-.0221158
1.GOALKEEPER	-.4978143	.0836035	-5.95	0.000	-.663858	-.3317705
_cons	-10.46155	.9310997	-11.24	0.000	-12.31079	-8.612305

5. Regression problems detection and correction

5.1 Model specification errors

First, we check whether the model is misspecified or not in order to go on to check for other possible problems. For this problem, we will mainly use the Ramsey RESET test.

```
Ramsey RESET test using powers of the fitted values of lnVALUE
       Ho:  model has no omitted variables
                   F(3, 89) =      7.18
                   Prob > F =      0.0002
```

From the p-value result of the test, we can confidently reject the null hypothesis, so there is a model specification problem in our model. This problem caused positive and negative biases which depend on correlations between omitted variables and independent variables. We will discuss the problem later in the next chapter.

5.2 Multicollinearity

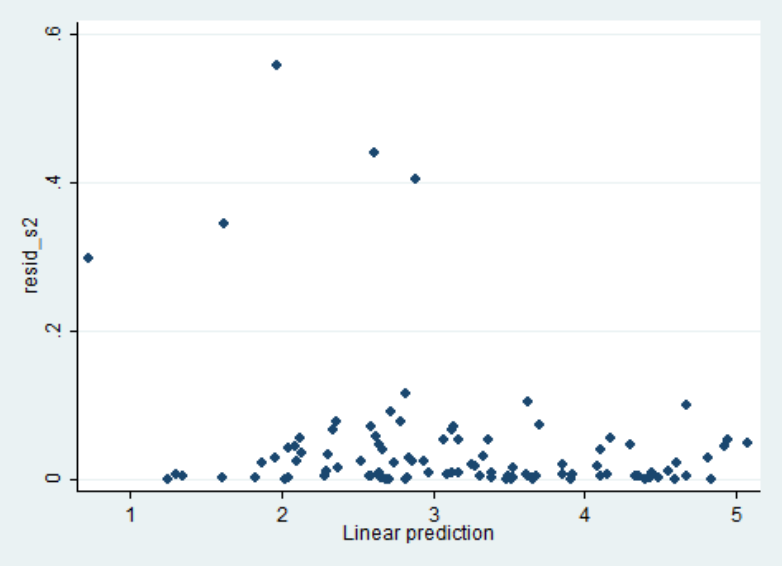
From the scatterplot matrix, we cannot observe any strong relationship between the independent variables, thus we expect no multicollinearity problem. We confirm this conclusion with a formal test, the variance inflation factor (VIF). The result implies that the model is really without multicollinearity problems.

```
. vif
```

Variable	VIF	1/VIF
AGE	1.28	0.779702
HEIGHT	3.00	0.332834
WEIGHT	2.76	0.361964
OVA	1.11	0.901612
1.DEFENDER	1.69	0.590252
1.MIDFIELDER	1.93	0.518211
1.GOALKEEPER	1.80	0.556608
Mean VIF	1.94	

5.3 Heteroscedasticity

There are many tests to identify heteroscedasticity problems. In this study, we use 3 tests, which are the informal test, Breusch – Pagan – Godfrey (BPG) test, and White test. We start with the informal test. By observing the graphical relationship between the values of residual squared (u) and the values of predicted dependent variable (yhat), the result is as follows.



Next, we use the Breusch – Pagan – Godfrey (BPG) test.

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of lnVALUE

         chi2(1)      =    14.16
         Prob > chi2  =   0.0002
```

Then, with the last test is the White test.

```
White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(29)     =     72.03
         Prob > chi2  =    0.0000
```

In the last two tests, the results we got are statistically significant in which we can reject the null hypothesis. There is a heteroskedasticity problem in our model. There are two ways to fix the problem, weighted least square and heteroscedasticity-robust standard error. Since we could not find the functional form of the variances, we should not use the weighted least square method. Therefore, we have to use the robust standard error instead to fix the problem. The results after the correction will be shown in the next chapter.

5.4 Autocorrelation

Since the data and the model we used are both in cross-sectional style, the autocorrelation problem seems not to be a big problem in our model estimation.

6. Conclusion

Linear regression

Number of obs	= 100
F(7, 92)	= 286.97
Prob > F	= 0.0000
R-squared	= 0.9547
Root MSE	= .21643

lnvalue	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
AGE	-.1181743	.0108109	-10.93	0.000	-.1396457	-.0967029
HEIGHT	.0029312	.0048075	0.61	0.544	-.0066169	.0124793
WEIGHT	-.0002419	.0039893	-0.06	0.952	-.008165	.0076811
OVA	.2030182	.0050878	39.90	0.000	.1929133	.2131231
1.DEFENDER	-.1712935	.057522	-2.98	0.004	-.2855371	-.05705
1.MIDFIELDER	-.1435215	.0478204	-3.00	0.003	-.2384969	-.0485461
1.GOALKEEPER	-.4978143	.1023337	-4.86	0.000	-.7010579	-.2945707
_cons	-10.46155	.7786266	-13.44	0.000	-12.00797	-8.91513

First, this model has an omitted variable bias. We should find omitted variable and add it to our model. But it seems to be the problem since we could not find and add the omitted variables to this model. However, we know that there are other variables which should affect the value of football players, such as goal scoring, number of assists, number of yellow cards and red cards, for instance. This data we used comes from sofifa.com which is the website providing data about football players based on a videogame named EA sports FIFA 22. While we can collect the data about age, body size, overall rating score, and the positions the football players are playing from this website, further variables mentioned above are not available on sofifa.com, as they are real-world data. In case we merge two datasets from different sources which are collected by different methodology, our data will be unusable and the model we will get will be unreliable. So, please be careful when using this model. Keep in mind that this model is based on the data from the video-game-reflected website, which in fact, does not totally reflect reality.

Secondly, to analyze the result of the model, the coefficient of AGE is negative and statistically significant at 5%. We can interpret that the age of Premier League football players has a negative effect on the value of Premier League football players. Next, the coefficient of OVA is positive and statistically significant at 5%. However, the coefficients of HEIGHT and WEIGHT are statistically insignificant at all

levels. It is not surprising because the Premier League is the most popular football league which mostly focuses on football playing performance. We can see football players like N'Golo Kanté who is only 168 cm in height, but he is a keyman that leads Chelsea Football Club to a fourth place in the 2020-2021 season.

Lastly, all coefficients on dummy variables are negative and statistically significant at 5%. This means that the strikers value more than football players in other positions. It makes sense because football is a sport that focuses on scoring. Many strikers of the less unpopular football clubs are still more well known than the goalkeepers of the famous football clubs.

References

Metelski, Adam. (2021). Factors affecting the value of football players in the transfer market. *Journal of Physical Education and Sport. 21*. 1150-1155. 10.7752/jpes.2021.s2145.

Adiwiyana, H., & Harymawan, I. (2021). Factors that Determine the Market Value of Professional Football Players in Indonesia. *Jurnal Dinamika Akuntansi, 13*(1), 51-61. doi:https://doi.org/10.15294/jda.v13i1.26079