The analysis of factors influencing CSL teams' market value

Qilin Dai, Yuhang Jiang, Sijie Yang, Yanming Zhang

May 2023

1 Introduction

1.1 Problem restatement

This essay is mainly focused on the effects of different factors on Chinese Super League(CSL) teams' market value. After the opening of CSL in the new season, as loyal fans, we have a keen interest in the relationship between the market value of the Chinese Super League teams and the strength of the teams, the development degree of the city where the teams are located and the popularity of different teams.

1.2 Hypothesis expound

Hypothesis for the inference for slope:

The null hypothesis H_0: $\beta = 0$

The alternative hypothesis H_a: $\beta \neq 0$

Where β represents the real population slope for the LSRL of market value versus a certain explanatory variable.

Other than the hypothesis for the inference of slope, we also suppose that the strength of the teams is the most influential factor among the three different aspects we chose.

1.3 Expected result

After conducting the Linear Regression Analysis, the residual plots for market value versus different explanatory variables show no obvious certain pattern, which provides convincing evidence that the linear is suitable for all factors we chose. The p-value for each inference for slope is smaller than the significance level(0.05), and the p-value for the strength of the teams is the smallest.

2 Background research

According to The Economics Theory of Professional Team Sports written by Stefan Késenne, the market value for a sports team is often related to various factors, such as the local economic level and popularity of the team[2]. Two relevant studies discuss the factors that relate to the market value of a team. A recent study conducted by Carol Pina examined the muti-factors that influence the professional sports franchise amongst teams in different sports leagues, including the Major League Soccer, the National Basketball Association, etc .[3]. The study figured out several factors for the franchise values of teams in different professional sports leagues through linear regression analysis. Besides, an article published by David Butler mainly discusses the relationship between the salary caps and the points of the teams in La Liga[1]. The result reveals the log of salary caps has a moderately strong and linear correlation with the points of La Liga teams.

3 Methods and procedures

3.1 Data collection

We selected the points to represent the strength of the teams, the GDP of the corresponding city to represent the development degree of the city where the teams are located, and attendance to represent

the popularity of a team. Data related to the market value, attendance, and points are collected from the website transfermarkt; the GDP of the corresponding city is collected from the CEIdata.

3.2 Inference

3.2.1 Check conditions

To conduct the inference for slope, we have to check the following conditions: Linear, independence, normally distributed error, and equal variance. Since we are using the population data, we do not have to check the random selection condition. Also, because we collected data from different teams, it is reasonable to assume that all data points are independent of each other. In terms of other conditions, we applied Linear Regression to all of the three factors and draw the residual plots, as shown in Figure 1, for each of them.

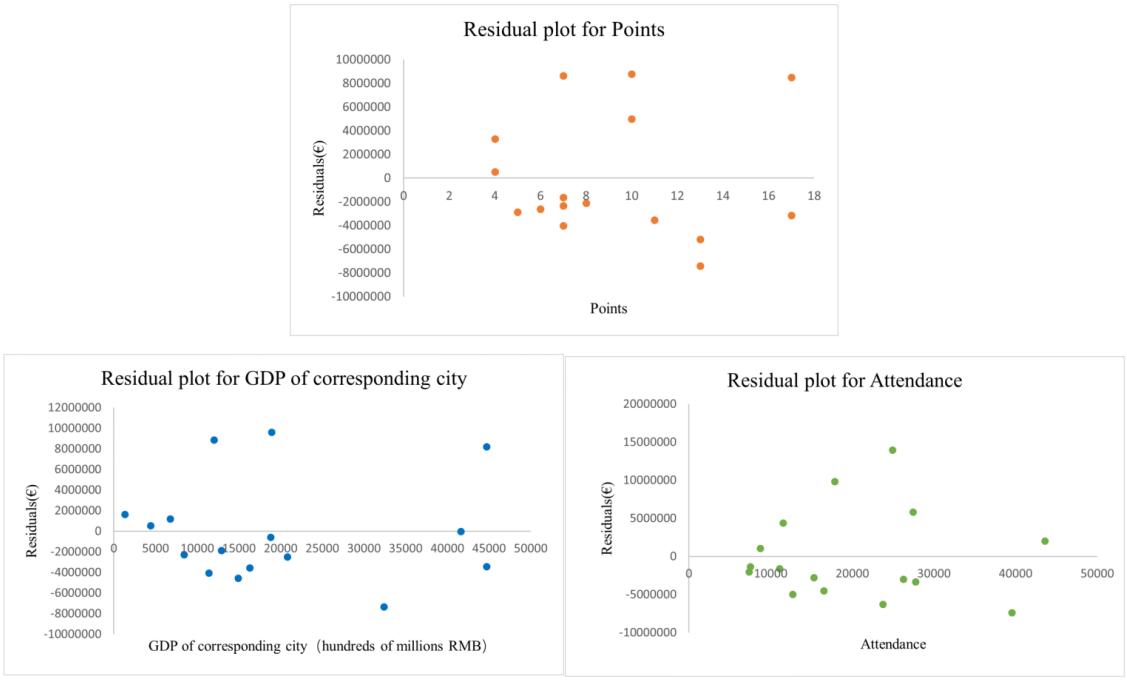

Figure 1: Residual plots for three explanatory variables

The residual plots show no obvious pattern, and the data points are randomly scattered around 0. It can be inferred that the linear model is fit for all these three factors. Besides, the plots do not indicate the heteroskedasticity of the data, which obeys the equal variance condition. For the normally distributed error condition, three factors are all suitable since the data points are dense around 0 and get sparse as the y-axis value increases. There are no obvious outliers in the plot.

3.2.2 Theoretical inference

For the GDP of the corresponding city, it is a decent representation of the development degree of the local city. We suppose the economic condition of a city directly impacts the market value. A prosperous city with a strong economy, high disposable income, and a large fan base is likely to attract more sponsors, investors, and fans, increasing the demand for the team and ultimately raising its market value. Conversely, a struggling city with a weak economy may face challenges in generating revenue, attracting sponsors, and maintaining fan support.

For the points, it is the most direct representation of the strength of a team. We speculate that the strength of the teams is strongly associated with the market value since a stronger team could attract more attention and often have better players in the team. This could lead to better advertising effects and create more commercial profits.

For the attendance, we take it as a good representation of the popularity of teams. Similarly, we suppose that popularity is a vital factor influencing the market value since it relates to the investments

and fans' support closely.

Nevertheless, the factors chosen might be weakly correlated to and have little impact on the market value even though the three aspects studied are vital. This is because the factors are not representative of the corresponding aspect. They might be determined by many potential factors and are not directly related to the market value.

3.2.3 Hypothesis test

We used Least Square Regression to construct a linear model for each of the factors. On top of that, we calculated r and R-square for each linear model to figure out the correlation between market value and the three factors.

With the linear models obtained, we conducted the inference for slope through the Student t-test. We calculated the p-value for each explanatory variable and checked the hypothesis.

By comparing the R-square, we can know which of the variables is most correlated to the market value. Besides, we can also know which of the variables is the most influential one by comparing the p-value obtained.

4 Results

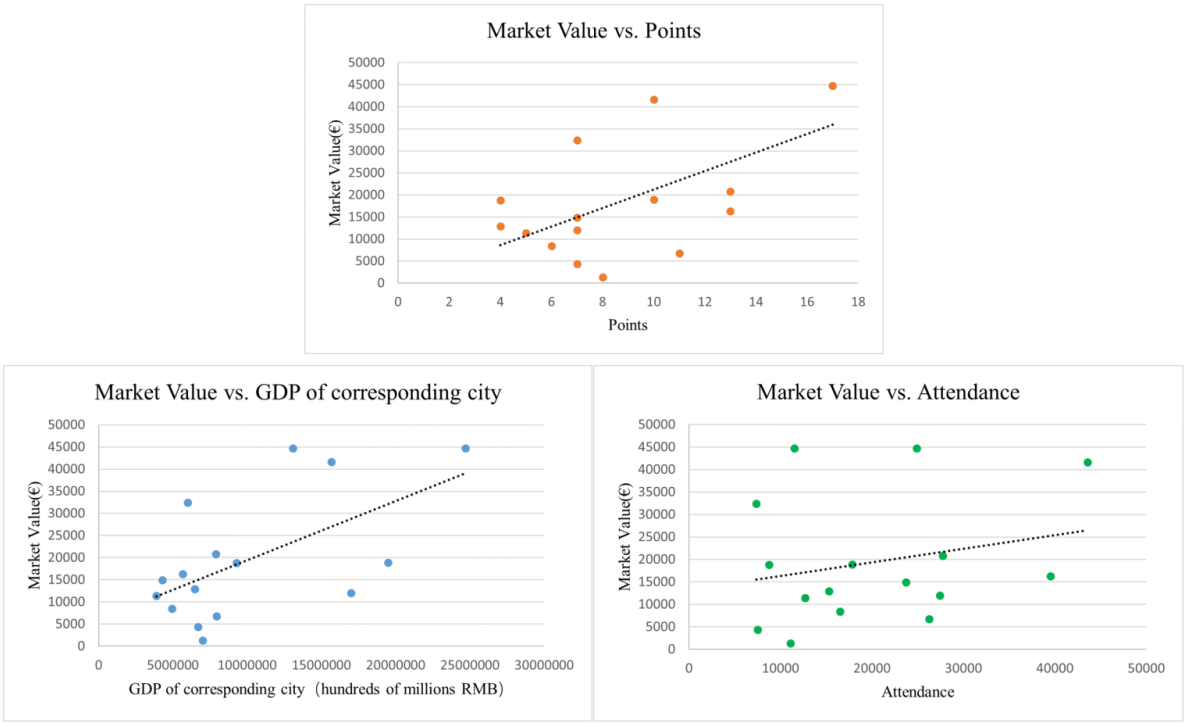

Figure 2: Scatter plot of the three explanatory variables with LSRL

We did the regression analysis between the GDP and the market value. We obtained a p-value of 0.017. Because the p-value 0.017 is less than the alpha value 0.05, we reject the null hypothesis. We do have sufficient evidence to conclude that there is a relationship between the GDP and the market value. Since the r is 0.586, the linear association between GDP and market value is moderate and positive. The r-squared is 0.344, which means about 34.4 percent of the variation in market value is explained by the linear relationship with GDP.

Then we did the regression analysis between the Pts and the market value. We obtained a p-value of 0.004. Because the p-value 0.004 is less than the alpha value 0.05, we reject the null hypothesis. We do have sufficient evidence to conclude that there is a relationship between the Pts and the market value. Since the r is 0.672, the linear association between Pts and market value is moderately strong and positive. The r-squared is 0.451, which means about 45.1 percent of the variation in market value is explained by the linear relationship with Pts.

After that, we did the regression analysis between attendance and the market value. We obtained a p-value of 0.356. Because the p-value 0.356 is greater than the alpha value 0.05, we fail to reject the null hypothesis. We do not have sufficient evidence to conclude that there is a relationship between attendance and market value. Since the r is 0.247, the linear association between attendance and market value is weak and positive. The r-squared is 0.061, which means about 6.1 percent of the variation in market value is explained by the linear relationship with attendance.

Table 1: Results of the three explanatory variables

Factors	R-square	r	p-value
Pts	0.344	0.586	0.017
GDP	0.451	0.672	0.004
Attendance	0.061	0.247	0.356

In short, we can conclude that the market value of a CSL team is correlated to the local economic development level(GDP) and the performance of the team(Pts). However, we fail to demonstrate that the popularity of a team(attendance) is correlated to its market value. We decided to find an alternative factor that might influence a team's market value.

5 Improvement

5.1 Existing Problem

After conducting the Linear Models, we found out that attendance is not very strongly associated with the market value by examining the p-value and the R2. This may refer to the complexity of the attendance. We have to find another factor to represent the popularity of the teams and have a strong correlation with the market value.

5.2 Data collection and Inference

We selected the number of followers on the teams' Weibo accounts to represent their popularity. Along with the development of the Internet, social media has become a very important part of everyone's daily life. It is a channel for people to express their ideas and acquire different types of information. Therefore, we suppose that the number of followers is a good representation of popularity and is strongly correlated with the market value.

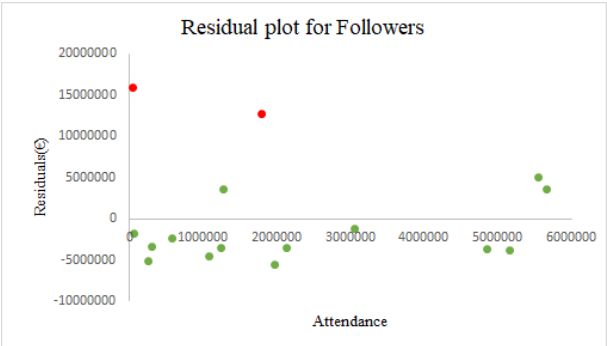

Figure 3: Residual Plot of the followers

Before conducting the inference for slope, we need to check the linear condition first. As shown in Figure 3 below, the data points randomly scatters around 0 and show no obvious pattern in the residual plot. However, there are 2 potential outliers shown in the residual plot(marked as red); we have to remove them before we conduct the inference for slope.

5.3 Results

We did a regression analysis using the data of Internet followers and market value. Based on the line fit plot and the residual plot, we recognized two outliers that are significantly deviated from the expected

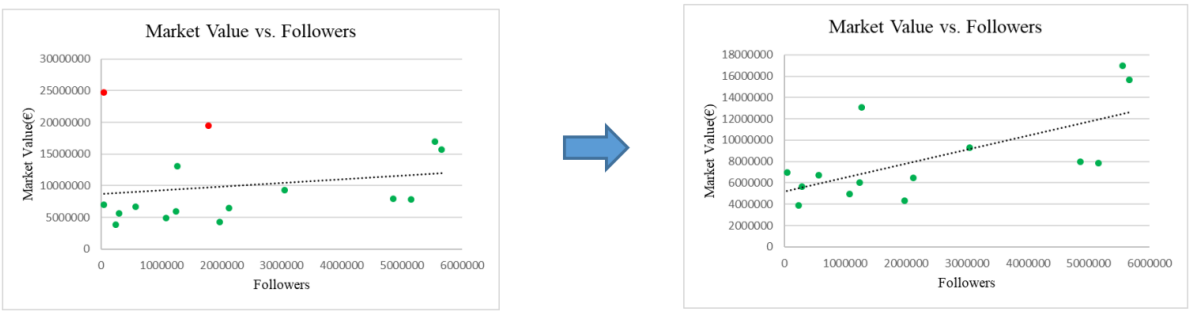

Figure 4: Scatter plot of the followers with LSRL

value, which are (51000,24710000) and (1788000,19480000) (marked as red in Figure 4). After excluding these two outliers, we redid the regression analysis and obtained a p-value of 0.0016. Because the p-value 0.0016 is less than the alpha value 0.05, we reject the null hypothesis. We do have sufficient evidence to conclude that there is a relationship between the Internet followers and the market value. Since the r is 0.760, the linear association between Internet followers and market value is moderately strong and positive. The r-squared is 0.577, which means about 57.7 percent of variation in market value is explained by the linear relationship with Internet followers.

Table 2: Results of the followers

Factors	R-square	r	p-value
Followers	0.577	0.760	0.0016

In short, we successfully demonstrate that the market value of a team is correlated with its Internet followers. By comparing it with the result above which we fail to demonstrate the correlation between the market value and attendance, we believe that the Internet followers is a better indicator of a team's market value comparing to on-site followers(attendance).

6 Further Exploration

The sample size we attained was considered small/insufficient. The major cause of a narrow time period found in our field of study. We've come to a problem-driven solution. The main problem we've faced is the inconsistency of context between datasets in different years. Our intention is to reduce the chronological variance by considering factors that vary quantitatively through time.

The first element is inflation in the economy, which is proportionally related to purchasing power of the population, which relates consensually with our independent variables such as attendance rate and market value.

We devise to derive a model for each variable about time, and with that, we can normalize different datasets over time hence expanding our sample size.

The second element is the fitting model applied. We currently use a basic linear regression method. It is a primitive model and lacks the power to adapt to different independent-dependent relations involving potential complexity that normally happens in real life.

We choose to inherit from the basic linear regression model but by shaping it more generically, it is possible to capture more patterns. The method we prospect and found is called the Generalized Linear Model(GLM), which allows independent variables to have other forms of distribution other than a normal one.

7 Conclusion

In conclusion, among the four factors we introduced to this analysis, GDP, Points, and Internet followers are correlated with the market value of Chinese Super League teams, while attendance is not correlated with the market value of teams. In other words, the market value of a CSL team is correlated with the local economic development level, the performance of the team, and the amount of supporters.

References

[1] David Butler (2021) *"La Liga Salary Caps."*, The Economics of Sport, www.sportseconomics.org/sports-economics/la-liga-salary-caps. Accessed 17 May 2023.

[2] Stefan Kesenne (2015) *Economic Theory of Professional Team Sports: An Analytical Treatment*, Edward Elgar.

[3] Carlos Pina (2018) *"DETERMINANTS OF SPORTS FRANCHISE VALUES: HOW DOES MLS HOLD UP AGAINST THE BIG FOUR?"*, California State Polytechnic University, Pomona.

[4] Transfermarkt (2023) *OVERALL MARKET VALUE CHANGE IN ALL CLUBS IN CHINESE SUPER LEAGUE*, https://www.transfermarkt.com/chinese-super league/marktwerteverein/wettbewerb/CSL

[5] Chinese economic Data checkpoint (2023) *China GDP data by region*, https://ceidata.cei.cn/new/ShowTabsByInterface?from=ceidatakey=GDP

[6] Transfermarkt (2023) *Chinese Super League - Attendance Figures.*, https://www.transfermarkt.com/chinese-super-league/besucherzahlen/wettbewerb/CSL

[7] Transfermarkt (2022) *Chinese Super League - Attendance Figures (Detailed View).*, https://www.transfermarkt.com/chinese-super-league/besucherzahlen/wettbewerb/CSL/saisonid/2022/plus/1

[8] Weibo.com (2023) *Sina Visitor System.*, https://weibo.com/csleague