

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 factors among 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 provide 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 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 (2014). Two relevant studies discuss the factors that relate to the market value of a team. A recent study conducted by Carol Pina examined the multi-factors that influence the professional sports franchise amongst teams in different sports leagues, including the Major League Soccer, the National Basketball Association, etc (2018). 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 (2021). 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 select the points to represent the strength of the teams, GDP of 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 that 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

In order to conduct the inference for slope, we have to check following conditions: Linear, independence, normal 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 from 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.

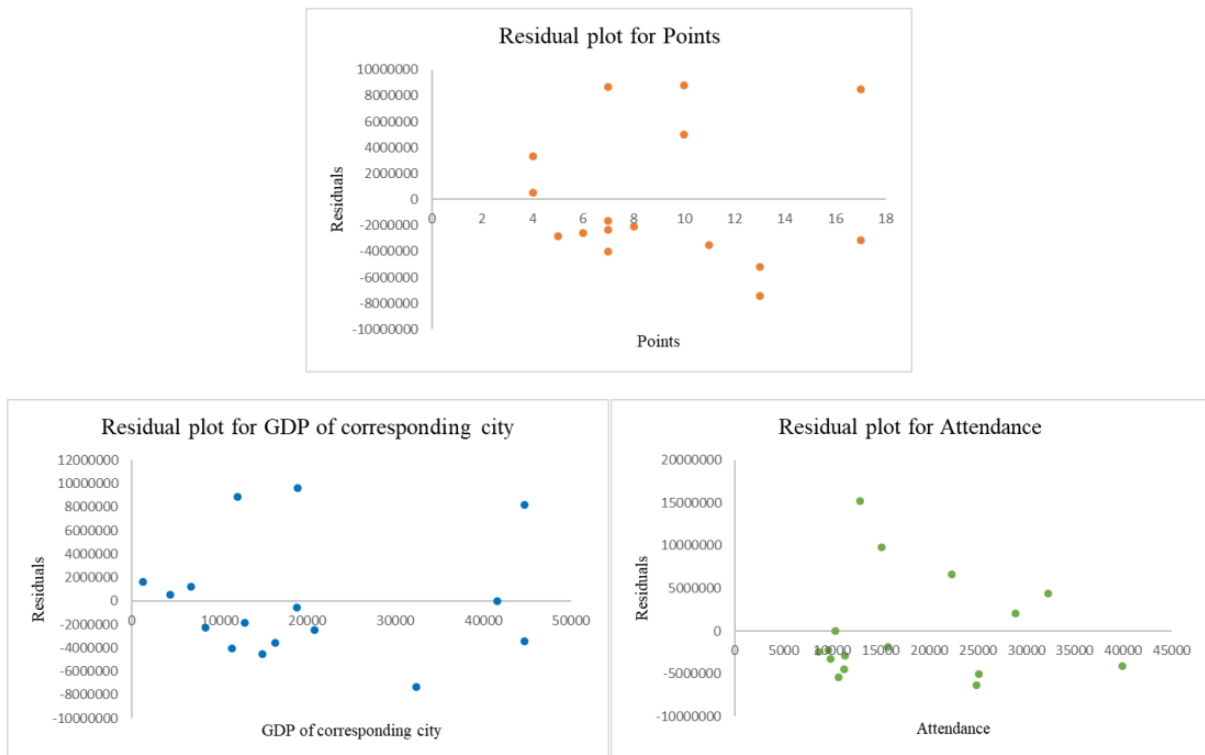


Figure 1: Residual Plots for three explanatory variables

The residual plots show no obvious pattern, and the data points 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 normal distributed error condition, three factors are all suitable since the data points are dense around 0 and getting sparse as y-axis value increasing. There is no obvious outliers in the plot.

3.2.2 Theoretical inference

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 a better advertising effects and creating more commercial profits.

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 attendance, we take it as a good representation of the popularity of teams. Similarly, we suppose that the 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 having little impact on the market value even though three aspects studied is vital. This is because the factors are not representative for the corresponding aspect. They might be determined by many potential factors and is not directly related to the market value.

3.2.3 Hypothesis test

3.2.4 Confidence interval

4 Analysis of data

5 Improvement

5.1 Existing Problem

After conducting the Linear Models, we found out that the attendance is not very strongly associated with the market value by examining the p-value and the R^2 . 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 Method and procedures

5.2.1 Data collection and Inference

We select the number of followers on the teams' Weibo account to represent the popularity. Along with the development of the Internet, the social media has become a very important part in everyone's daily life. It is a channel for people to express their idea 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.

Before conducting the inference for slope, we need to check the linear condition first. As shown in Figure 2 below, the data points randomly scatters around 0 and show no obvious pattern in the residual plot. It is reasonable to conduct the inference for slope.

5.2.2 Hypothesis test and confidence interval

5.3 Analysis of data

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