Competitive Balance and Home Team Advantage

Aram Sargsyan, Ashot Janibekyan, Barsegh Atanyan, Sarin Sulahian

Objective

Our objective for this group project is to see the relationship between competitive balance and home team advantage in football. What is the correlation between these two? To answer this question, we decided to apply several used methods both for competitive balance and home team advantage to suffice our question.

Let us start off by briefly introducing the main idea behind competitive balance and home team advantage. In a perfectly balanced league or competition each team in any given sport enjoys the same probability of outcome to win in each game. However, the perfectly balanced situation is so rare that we can state it never happens. Thus, various competitive balance measures are introduced to describe the competitiveness of a given competition. Home team advantage is the idea that describes the benefits of home teams over visiting teams. Scientists have long been arguing about the physical and psychological reasons behind this phenomenon. Such factors might be the fans supporting the home team more, the bias of referees, psychological or physiological advantages of playing in a familiar setting, and so on. The fact is that it exists. We can imply, basically, that if two competitors have the same abilities(or not), those who play in their field have some sort of advantage over the other.

In short, if we were to introduce our idea to a sport fan who has not much to do with sports analytics, we would tell if playing at home matters more or less in competitive leagues. Intersting, yes?

Data

The data contains seasons 1994-2019, the respective countries are: England, France, Germany, Italy, Netherlands, Spain, Greece, Portugal, Scotland, Turkey. Top 5 of the countries contain also the data of several leagues. The extracted features include Winning Percentages, Number of Conceded and Scored Goals, total points, etc. Based on these features different metrics for home team advantage and competitive balance are calculated. Some metrics require computation of additional features as well.

Different Metrics

Home Team Advantage

Combined Measure of Home Team Advantage based based on Beta Distribution

Combined measure of home advantage is random variable C that can take values -1, 0, and 1. $C = -1$ for team T1 if two matches between teams T1 and T2 in a season ended with a better result – measured by goal differences in matches – for team T1 away from home (at T2's ground). $C = 0$ for team T1 if goal differences in both matches were exactly the same from T1's point of view, and $C = 1$ for team T1 if this team recorded better result – measured by goal differences in matches – at its own ground. With results hT1: aT2 at the home ground of team T1 and hT2: aT1 at the home ground of team T2 the value of random variable C for team T1 is determined as $C = sgn((hT1 - aT2) - (aT1 - hT2))$.

Let $Z_r, r = -1, 0, 1$ are random variables which describe number of cases (e.g., for a team in a season) where it is possible to identify home advantage ($r = 1$), away advantage ($r = -1$), and no advantage ($r = 0$), i.e., Zr sums number of cases where $C = r$. Vector (Z_1, Z_0, Z_1) follows trinomial distribution with parameters p_1, p_0, p_1, and K with probability function

$p(k_1, k_0, k_1) = \frac{k!}{k_1! \cdot k_0! \cdot k_1!} \cdot p_1^{k_1} \cdot p_0^{k_0} \cdot p_1^{k_1}$ where K is total number of observations of combined measure in a season, p_1, p_0, p_1 are probabilities of occurring home advantage (r = 1), away advantage (r = -1), and no advantage ($r = 0$). k_1, k_0, k_1 ($k_1 + k_0 + k_1 = K$) are observations of appropriate advantage. Maximum likelihood estimator of parameters $p_r, r = -1, 0, 1$ is $p^r = \frac{k_r}{K}, r = -1, 0, 1$.

Poisson Model based Home Team Advantage

The Poisson model is discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time.We use the Poisson regression which is part of the Generalized Linear Regression models family. Lambda is determined by a set of k regressor variable (the x's),the explanatory variables, $x = (x_1, x_2 \ldots x_k)$, can be continuous and categorical, which redirects to the poisson regression. These counts are to be modeled in a contingency table where the convention is called a log-linear model. log link: $\log(\lambda) = \beta_0 + \beta_1 \cdot x$ or $\lambda = e^{\beta_0 + \beta_1 \cdot x}$, where λ is never negative.

note: We assume that variables are independent based past research using the chi^2 test.

Competitive balance

Entropy

Horowitz (1997) chose to use the relative-entropy measure of information theory to measure seasonal competitive balance

Similarly, the minimum entropy or uncertainty about which team might have won a game selected at random from the schedule, is achieved when the best team wins all games. Relative entropy, $R = H = HM$, measures the degree of uncertainty about which team might have won a randomly-selected game, relative to the maximum uncertainty possible.

Surprise Indices

The Surprise Index was introduced by Groot and Groot (2003)- for reference- https://ideas.repec.org/p/ant/wpaper/2005034.html (https://ideas.repec.org/p/ant/wpaper/2005034.html), it is the ratio between P, the realized surprise points and M, the maximum number of surprise points that is possible when the teams are perfectly balanced. Two surprise points are given when a team looses from a lower ranked team and one point is awarded when the game ends in a tie. These points are weighted with the rank difference. $M = \frac{(N-1)N(N+1)}{3}$

$S = \frac{P}{M}$

Gini Competitive Balance

The Gini coefficient is originally developed to measure income inequalities. Schmidt (2001) and Schmidt & Berri (2001) use it to measure the inequality of winning percentages. The lower the Gini coefficient is, the more competitive the league is. $G = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left|x_i - x_j\right|}{2n^2\bar{x}}$

Calculating HHI and Noll-Scully

Herfindahl-Hirschman index comes from economics, and is a commonly accepted measure of market concentration. Increasing number of the teams affects lower bound (in some cases upper bound as well)

$$HHI = \sum_{i=1}^{n}\left(\frac{p_i}{\sum_{i=1}^{n} p_i}\right)^2$$

The Noll-Scully compares what a league actually looks like to what it would ideally look like. In a perfect world — i.e. one where actual balance matches the ideal — the metric would return a 1.0, meaning the league is balanced. If the number is above 1.0 it means the teams are further away in wins than we would expect given the ideal.If the number is below 1.0 — which is very rare case— it means the teams are closer in wins than we would expect.

Ideally the number will be equal to 1, and higher the number means that there is less competition, and if it is closer to 1, it means that the competitive balance is higher. $\sigma_{relative} = \frac{\sigma}{0.5/\sqrt{m}}$

Competitive Balance vs Home Team Advantage

Below we can see comparative analysis of different Home Team Advantage and Competitive Balance metrics. Looking at all the leagues together we see that there is such little positive correlation it is not significant. When looking into the plot league by league there are some cases such as Le Championat or Ligue 2 that correlation is noticeable.

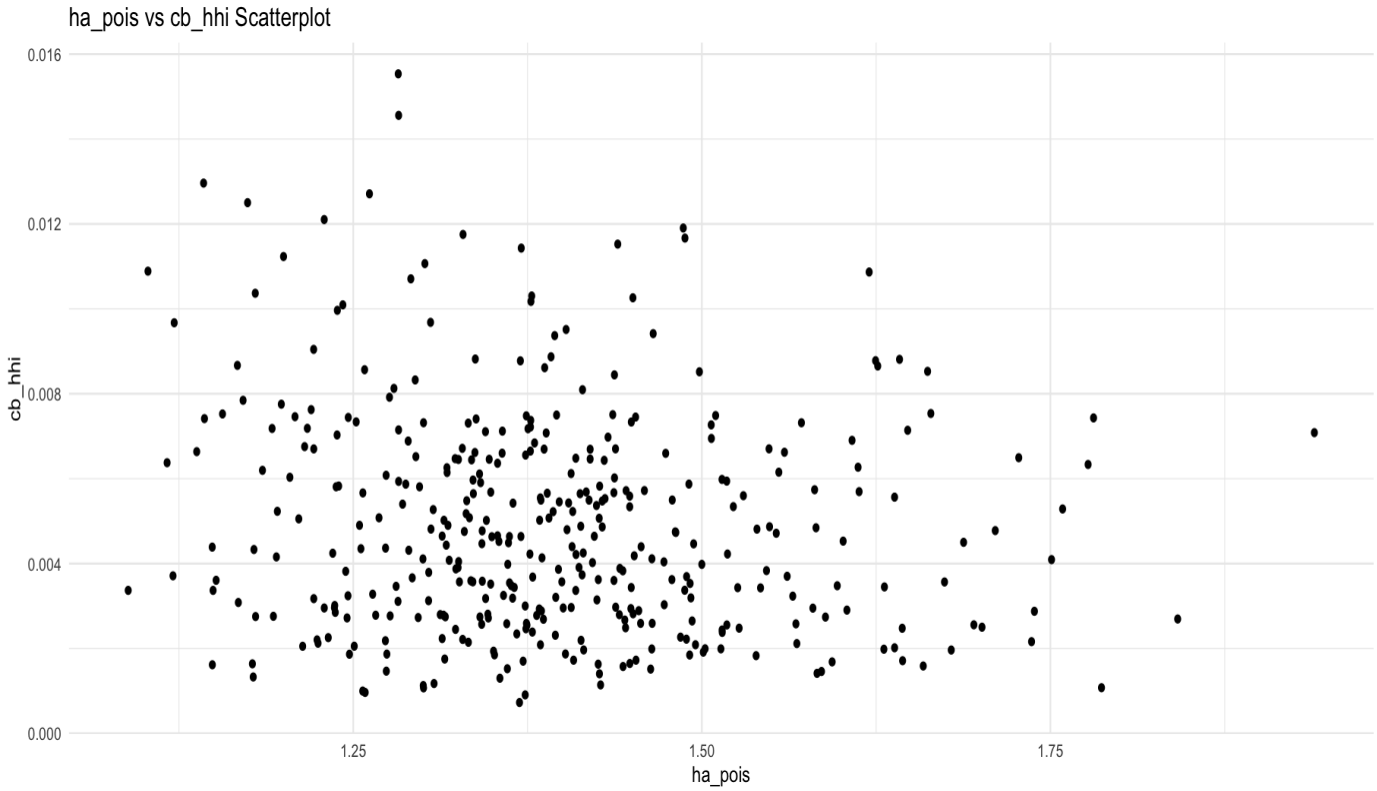
Method for Home Team Advantage

ha_pois

Method for Competitive Balance

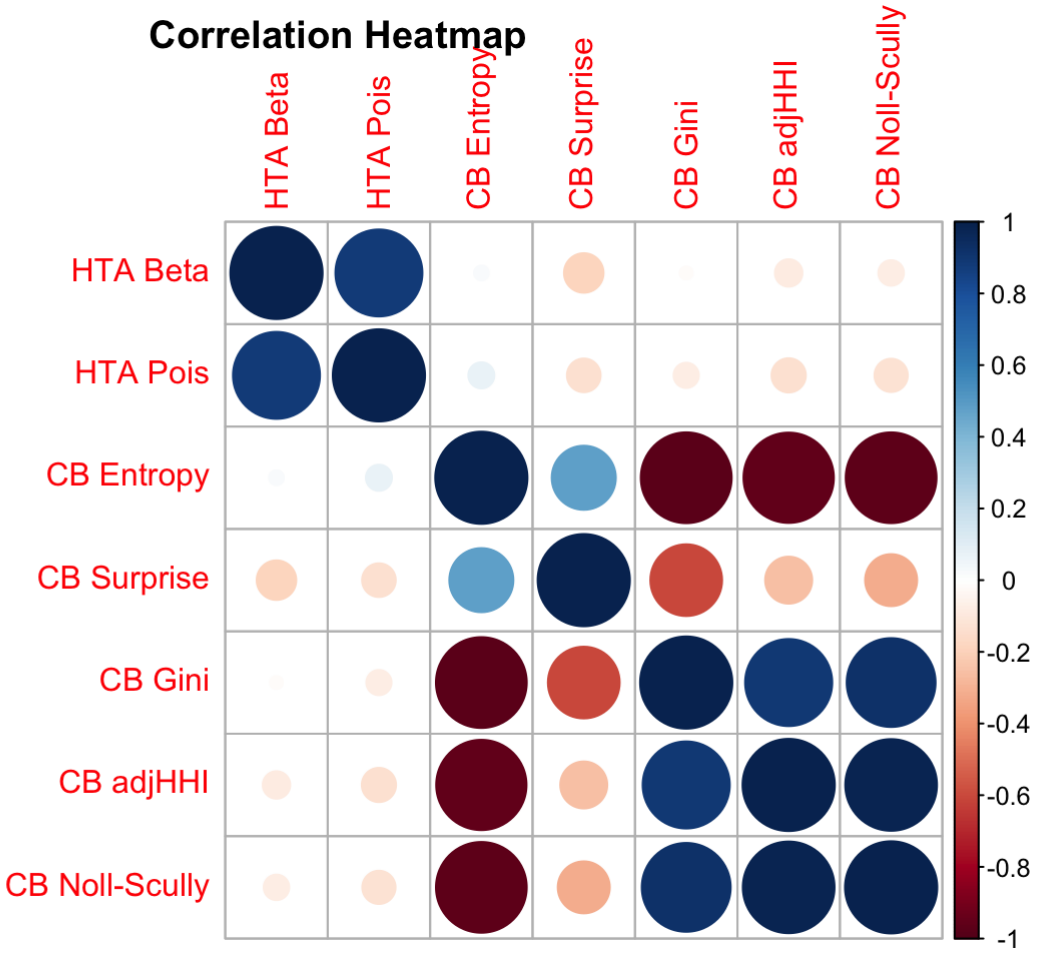
cb_hhi

☐ Country-wise



Correlation matrix

We can see that we have a strong linear correlation among several methods of calculating the same metric. On the other hand, we see that very small correlation exists among two methods of calculating two metrics. Hence, one can conlude that there is not a strong linear relationship between home team advantage and competitive balance.



2 Different Leagues

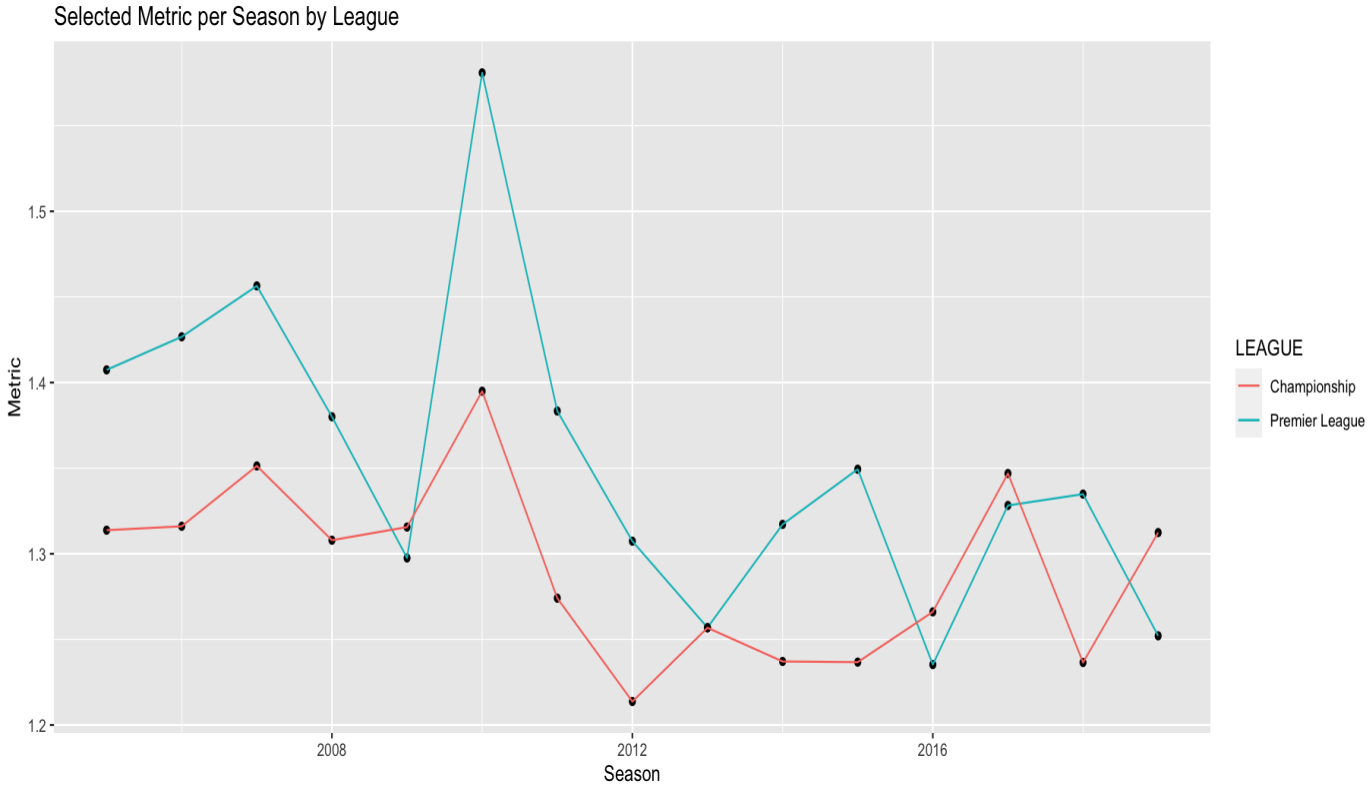
In the below case, we are looking at the metric for 2 different divisions within the same country. As we can see, generally we get similar trends and different metrics give reasonably similar results, which are not exactly the same, but are not significantly different as well. This means, that the metrics developed by us are consistent and depict the situation realistically. So, the above mentioned physiological and psychological factors may be specific to a country not only a league.

League

England

Metric

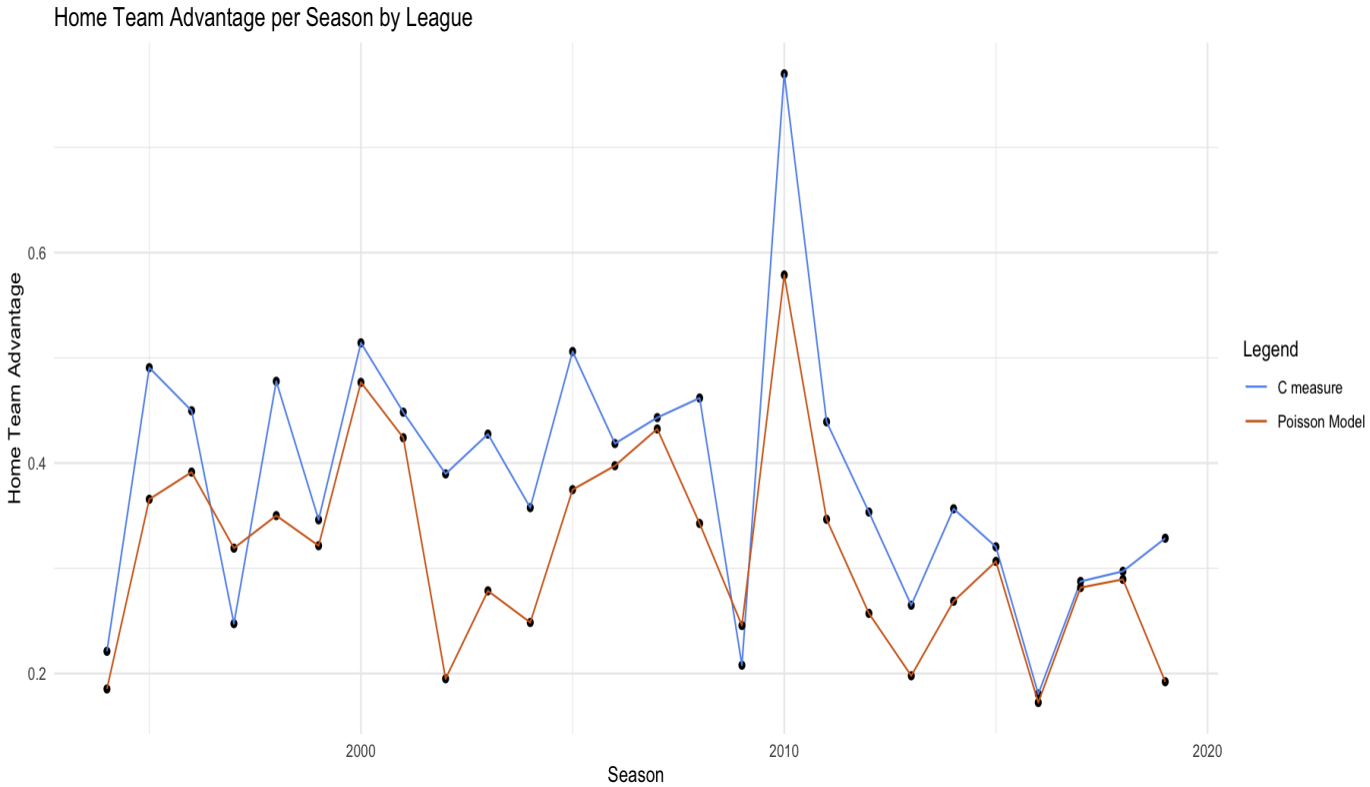
ha_pois



Home Team advantage per season per League

League

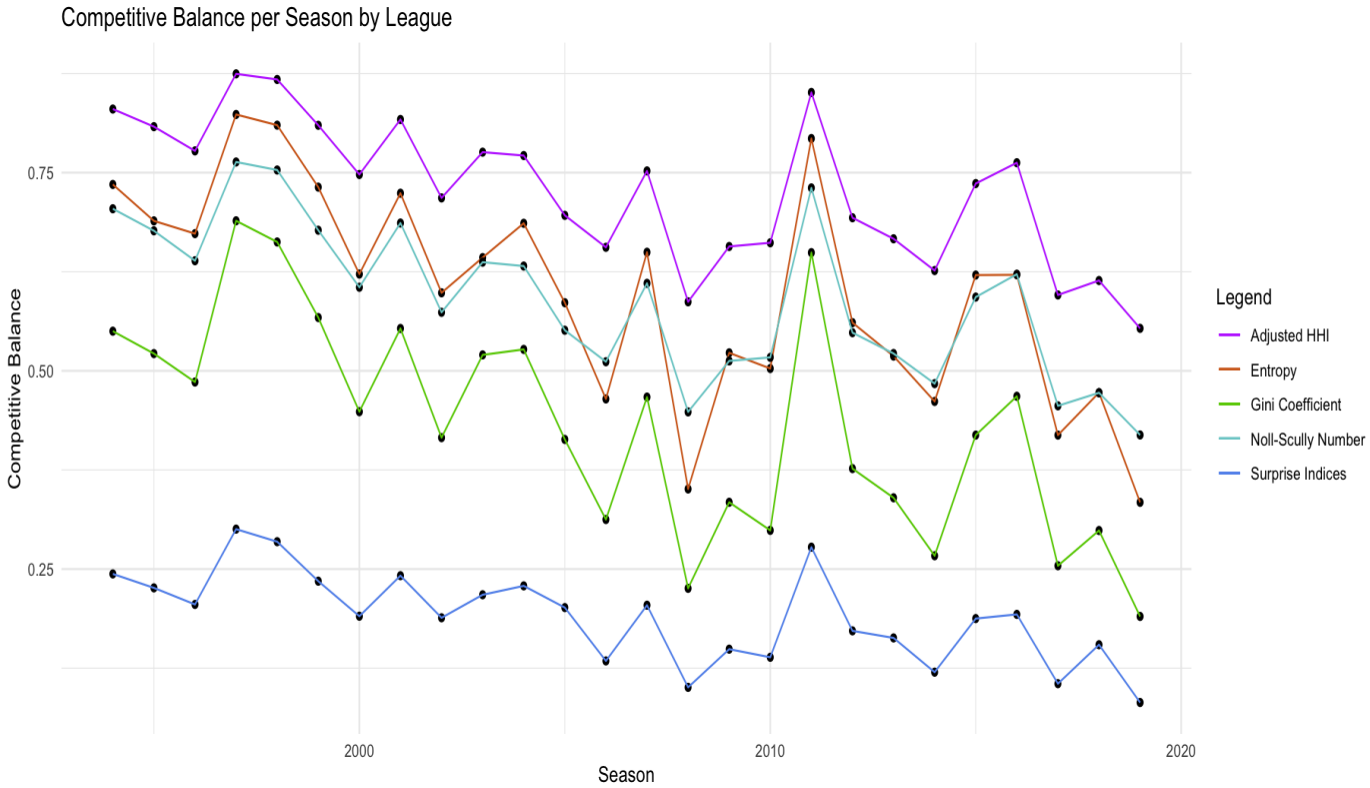
Premier League



On this plot we can see the different competitive balance measures for the selected league. As the linear correlation among each of them is very strong, they follow the same trends in general. Note that 3 of them- Gini Coefficient, Noll-Scully Number and Adjusted HHI are inversely proportional to competitive balance, that is why we have subtracted them from 1 to have the same trend for all measures.

League

Premier League


Competitive Balance per Season by League
Competitive Balance
0.75
0.50
0.25
2000
2010
2020
Season
Legend
Adjusted HHI
Entropy
Gini Coefficient
Noll-Scully Number
Surprise Indices

Conclusion

Let us conclude our analysis on the home team advantage and the competitive balance in a few points and revise what we have come up with. We did a various analysis using different metrics such as Bradley-Terry Model, Poisson Distribution, Surprise Index, Gini competitive balance, and so on. We compared each metric with each-other and on a whole level. We tested these metrics based on the leagues; we have concluded that there is more correlation when comparing each league by itself than if we compare it on a whole level. Also, comparing the different measures leads us to conclude that overall, the models between home team advantage and competitive balance do have some correlation, but it is not significant. There are also some extraordinary cases where some leagues are outliers, such as in the Premier League and Scotland. The best correlations were seen in leagues Le Championnat and League 2. Our project's overall outcome shows that there is some correlation between the competitive balance and home team advantage. Although it differs when calculating with different metrics, there might always be some outliers and extraordinary football cases.

Sources

https://www.mdpi.com/2227-9091/8/3/87/pdf (https://www.mdpi.com/2227-9091/8/3/87/pdf)
http://www.investopedia.com/terms/h/hhi.asp#ixzz4t8sTjDns
(http://www.investopedia.com/terms/h/hhi.asp#ixzz4t8sTjDns)
https://link.springer.com/content/pdf/10.1023/A:1007799730191.pdf
(https://link.springer.com/content/pdf/10.1023/A:1007799730191.pdf)