Málaga Economic Theory Research Center Working Papers

The distribution of soccer leagues scores that generates the minimum of competitive balance: Truncated-Cascade Distribution

Antonio Ávila-Cano and Francisco Triguero-Ruiz

WP 2018-4
June 2018

Departamento de Teoría e Historia Económica
Facultad de Ciencias Económicas y Empresariales
Universidad de Málaga
ISSN 1989-6908

The distribution of soccer leagues scores that generates the minimum of competitive balance: *Truncated-Cascade Distribution*

Antonio Avila-Cano
Department of Economics and Economic History
Programa de Doctorado en Economía y Empresa
Universidad de Málaga (Spain)

Francisco Triguero-Ruiz
Department of Languages and Computer Science
Universidad de Málaga (Spain)

Abstract

Competitive balance can be measured using standardized indexes, such as the Herfindahl-Hirschman index (HHI). The standardized HHI calculus is based on knowing the maximum value of HHI. This value corresponds to the minimum value of the competitive balance.
Measuring competitive balance is affected by the scoring system used. There are competitions that have scoring systems that do not award twice as many points for winning as they award for ties. In this case, the scores distribution representative of the minimum competitive balance is unstable because the total points at the end of the championship can vary. This issue has been addressed by reconstructing the results obtained in leagues. Nevertheless, this solution generates cardinal and ordinal negative effects, that we verify for the major European soccer leagues over twenty seasons.
The aim of this article is to redefine the perfectly unbalanced distribution in order to construct a new one that generates the maximum level of concentration: we call this *truncated-cascade distribution*. Thus, we show that the instability problem does not involve recalculating the scoring based on the results.
The minimum value of competitive balance is generated by a *truncated-cascade distribution* of results at a level that can be previously calculated. Thus, we calculate the cut-off point of the *truncated-cascade distribution* using a 5-grade polynomial equation obtained by recurrent calculation. Besides, we provide the cut-off points and maximum HHI values for leagues with different number of teams.

Keywords
Competitive balance, Herfindahl-Hirschman index, Perfectly unbalanced distribution, major European soccer leagues

JEL classification: Z20, Z21

The distribution of soccer leagues scores that generates the minimum of competitive balance: *Truncated-Cascade Distribution*

1. Introduction

Competitive balance (CB) is a basic and current concept in sports economics that indicates the degree of control exerted by certain teams owing to their performance quality. Before a championship begins, the CB is based on the available information on the strengths of the different teams, whereas at the end the CB is based on the results. Although the theory differentiates these approaches, the CB is calculated using ex-post analysis, where the distribution of the results is represented by a number.

The CB is conceptually related to concentration. This concept is used in industrial economics to denote a variable that represents the structure of the markets based on the distribution of the size of the firms. This variable indicates the degree of control over economic activities exerted by large firms. The degree of control reaches its minimum value if all the firms are the same size, whereas it reaches its highest values if only a few firms form the bulk of the market. Monopoly would be the extreme case.

Similarly, in the setting of sports, the CB reaches its maximum value if all the teams that participate in a given tournament achieve the same result, whereas it is at its minimum value if only few teams dominate the league. Therefore, there is an inverse relationship between CB and concentration.

Andreff (2015); Larsen, Fenn & Spenner (2006); and Zimbalist (2002 among others, provided a broad panorama of the literature on the concept of CB and its empirical application. They proposed diverse indexes for measuring the CB. Particularly, we use the Herfindahl-Hirschman index that is a well-know concentration index and widely used in practice.

Most of these indexes incorporate characteristics that are typical of sporting competitions. In a competition, there are aspects that influence the measurements of the CB and restrain the theoretical range of the indexes, leading to their standardization. Among such aspects, we emphasise the following: (i) the number of teams in the competition; (ii) the inability of any team to reach a monopoly position; and (iii) the scoring system used.

Competitions such as the major European soccer leagues have scoring systems that do not award twice as many points for winning as they award for ties. In this case, the scores distribution representative of the minimum CB is unstable because the total points at the end of the championship can vary. This issue has been addressed by reconstructing the results obtained in leagues with a pattern that complies the stability condition, which is that the sum of points at the end of the championship does not vary. This solution generates cardinal and ordinal negative effects as we show below.

In this article, we have made a contribution on the minimum value of CB, generated by a distribution which is known as perfectly unbalanced (PUD). This distribution, that represents the most imperfect competitive balance, has been called by the literature as the least balanced, the most unequal or the imperfect competitive balance distribution. This is the distribution that generates the highest value that the Herfindahl-Hirschman index can reach.

We advocate the calculation of the minimum value of the CB in advance of the start of competitions. This novel approach avoids recalculating the results whenever the scoring system does not fulfil the stability condition (Pawlowski, Breuer & Hovemann, 2010; Gayant & Le Pape, 2015). Therefore, this approach prevents the results, both cardinal and ordinal from being affected.

The rest of the article is structured as follows: In the second Section, we present the main aspects specific to the calculation of CB and we discuss the impact of the scoring pattern on competition. In the third Section, we justify the redefinition of PUD, and propose the distribution that generates the minimum level of CB, that we called *truncated-cascade distribution*. The fourth Section presents a calculus and practical application of our proposal to the major European soccer leagues over the last 20 years. The fifth Section provides a discussion of the results and describes the advantages of our proposal compared to those presented in the current literature. Two appendixes provide a summary of the mathematical demonstrations of the calculus proposed in the third Section.

2. Main Aspects of the Measurement of Competitive Balance

The close relationship between the concept of CB and the concepts of concentration and inequality led to the use of diverse indexes that have undergone modifications owing to the specific characteristics of sports competitions. Hence, most currently used measures have been relativized regarding the maximum or minimum values that they can reach. Thus, they are standardized and have the unit interval as the theoretical range. Specifically, the index takes a value of zero for the minimum concentration (i.e., the maximum value of CB corresponding to a distribution in which all teams have the same number of points) and it takes a value of 1 for the maximum concentration (i.e., minimum value of CB which is reached when the results correspond to what can be called the PUD).

The measures of CB used originated from typical measures of dispersion and concentration (Borooah & Mangan, 2012; Eckard, 2001; Fort & Quirck, 1995; Gayant & Le Pape, 2012; Gayant & Le Pape, 2015; Horowitz, 1997; Humphreys, 2002; Pawlowski, Breuer & Hovemann, 2010; Schmidt, 2001; Schmidt & Berri, 2001; Utt & Fort, 2002).

One of the most important measures is the Herfindahl-Hirschman index (HHI) proposed by Herfindahl (1950) and Hirschman (1945; 1964) (Depken, 1999; Larsen, Fenn & Spenner, 2006; Owen, Ryan & Weatherston, 2007).

In this article, we use the HHI defined as the sum of the squares of the shares that are also built from the points obtained by the teams at the end of the championship. This is its standardized version (i.e., corrected according to the minimum value that can be obtained, which would be analogous to the modification proposed by Depken, 1999) and relativized to the maximum range that it can reach:

$$HHI_{NORM} = \frac{HHI - HHI_{min}}{HHI_{max} - HHI_{min}}$$

where $HHI_{NORM} \in [0,1]$.

Owen, Ryan, & Weatherston (2007) proposed the HHI_{NORM} index as a measure of CB. Furthermore, relationships between some of the indexes have been analysed and verified (Borooah & Mangan, 2012; Gayant & Le Pape, 2015).

The mechanism for generating results in a competition entails three aspects: the scoring system in force; the number of participants; and the bilaterality of the competitive relationship. These aspects affect the maximum and minimum limits of the concentration of the results.

2.1. Implications of the Scoring System

The scoring system is part of the mechanism that generates results in a competition and entails rewards or incentives for winning. The possibility of ties means that the number of total points at the end of the championship can vary depending on the number of teams and the matches tied. This is a core aspect, since the distribution that reflects the minimum value of CB can be unstable.

We will use a league with $N \in \mathbb{N} - \{0, 1\}$ teams in a single round-robin system. Points are awarded following a pattern $P = \{p_w, p_t, p_l\}$ (win, tie, or loss). We assume that $p_w, p_t, p_l \in \mathbb{N}$ and $p_w > p_t > p_l$. Let us assume $p_l = 0$ and $p_w \geq 2\,p_t$. Let us denote by $p_i, i \in \{1, 2, \dots, N\}$, the number of points obtained at the end of the championship by the team i, whereas the vector $p = (p_1, p_2, \dots, p_N)$ represents the final scoring of the championship. The subscript i indicates the position of each team in the final ranking. Given vector p, we define the vector of shares $s = (s_1, s_2, \ldots, s_N)$, and for $s_i = \frac{p_i}{\sum_1^N p_i}$. An index that measures the CB will be a function defined on the vector of shares, which assigns a real number belonging to the unit interval.

Each team play $(N-1)$ matches and the number of matches played is $N(N-1)/2$. The tournament can be either a single or a double round-robin. This aspect is unimportant in analytical terms (i.e., it does not entail a loss of generality), because the formulation is maintained by multiplying the results by 2. Specifically, $N(N-1)$ matches would be played and each team would play 2(N - 1) matches.

2.1.1. Stability condition

We say that a scoring system complies the stability condition if the system gives zero points to the loser of each match and gives the winner twice as many points as it gives for a tie (Borooah & Mangan, 2012; Gayant & Le Pape, 2012). The stability condition implies that the number of points to be distributed remains constant (i.e., more generally, this is a system that, for a tie, gives half the number points as it gives for the sum of winning and losing (Gayant & Le Pape, 2015).

For example, a {3,1,0} pattern gives 3 points to the winner, 0 to the loser and 1 to each team in the case of a tie. The {3,1,0} pattern entails a higher incentive for winning than a {2,1,0} pattern. In both patterns, 2 points are shared by both teams (1 point each) if the match ends tied. With {2,1,0} pattern the total number of points is constant whereas with {3,1,0} may vary. This last pattern is usual in European soccer leagues.

Given a championship, w is the number of wins and t the number of ties at the end of the championship. We define $x = \frac{w}{N(N-1)/2}$ as the proportion of matches with a winner and a loser, and $1 - x = \frac{t}{N(N-1)/2}$ as the proportion of matches without a winner, and $w + t = N(N-1)/2$, the number of points at the end of the championship will be:

$$\sum_{i=1}^{N} p_i = p_w w + 2\, p_t t = p_w \cdot x \cdot N \cdot \frac{(N-1)}{2} + p_t \cdot (1-x) \cdot N \cdot (N-1) =$$

$$= N \cdot (N-1) \cdot \left[p_w \cdot \frac{x}{2} + \, p_t \cdot (1-x)\right] = N \cdot (N-1) \cdot \left[\left(\frac{p_w}{2} - p_t\right) \cdot x + p_t\right]$$

Then, if $p_w > 2\, p_t$, $\sum_{i=1}^{N} p_i$ depends on the number of teams, *N*, and on the number of wins and ties, *x* (Borooah & Mangan, 2012).

This fact affects the vector of shares s, which identifies the results of the competition, through the denominators of each component, and this dependency on the number of ties in this type of scoring pattern generates the structural instability that we are attempting to correct. This has been identified as one of the main problems when measuring CB. Thus, Gayant and Le Pape (2015) suggested that with a pattern that does not meet $p_w = 2\, p_t$, the configuration of the vector of results can fail to fulfil a rule that seems reasonable (i.e., if the CB is measured with a typical index, like the dispersion index or the HHI, then the value should lie between the maximum and minimum values that the index can reach in theory).

2.1.2. Reconstructing the results

For this reason, it has been suggested that the results should be reconstructed using a {2,1,0} pattern as a solution to measure CB in leagues using the {3,1,0} pattern. This is the case in the major European soccer leagues and in the annual UEFA Champions League (Pawlowski, Breuer & Hovemann, 2010; Gayant & Le Pape, 2015).

Nevertheless, this procedure is not neutral, because it generates a distribution of results that differs from the one obtained through the real mechanism of competition, in which the reward for winning is greater and the teams compete accordingly.

This aspect is especially relevant and deserves attention because reconstructing the results is not neutral, either in cardinal or ordinal terms. Thus, recalculation: (i) generates a different distribution in quantitative terms, which affects the value of the CB index used; and (ii) can also alter the ordinality of the ranking teams in the vector of result, given that the purpose of any measurement is comparison.

Such changes in ordinality can take place internally in the championship in such a way that a team can be classified into a ranking that differs from the real one. The only requirement is that the differential of points between two or more teams can change if there is a lower reward for wins.

We have recalculating with a {2,1,0} pattern the scores for the five major European Leagues from 1996/97 to 2016/17 season. Table 1 shows internal changes caused by recalculating the results of each league and season. There are changes in the ranking of position in 96% of the competitions and in 24% of the teams. Furthermore, there would be changes in classification for the UEFA Champions League or UEFA Europa League in 23% of cases, and relegations from category in 27% of cases.

This aspect can also affect comparisons between the CB of different leagues because a league can be more or less balanced than another depending on what system is used.

The Section entitled Empirical Application presents a practical example that shows, in conceptual and practical terms, that the reconstruction of results using different scoring systems is not advisable. This finding makes the identification of the core problem we are addressing even more interesting. The third Section is dedicated to solving this challenge.

2.2.Perfectly Balanced Distribution of Scores

The perfect competitive balance distribution of shares, that we called perfectly balanced distribution (PBD), is representative of the maximum CB (Borooah & Mangan, 2012; Gayant & Le Pape, 2015). In this distribution, the teams obtain the same number of points, and so the share is equal for all them: $s^{PBD} = \left(\frac{1}{N}, \dots, \frac{1}{N}\right)$. In s^{PBD}, the standard deviation of the shares is zero, and the HHI is equal to the inverse of the number of teams (Depken, 1999) and reaches its lowest level (Owen, Ryan & Weatherston, 2007).

Given a number of teams in a championship, the **maximum** CB is reached when none of them stand out from the rest (i.e., when there is an even distribution of results). Therefore, the maximum balance depends of the number of teams of the competition.

In this sense, when measuring CB between competitors over time, a typical problem that arises is changes in the number of competing teams (Fort & Quirck, 1995; Depken, 1999). For example, the German soccer league has 18 teams, whereas the Spanish, Italian, English, and French leagues have 20 teams. Furthermore, the number of teams has changed several times in all of the leagues over the last 50 years.

The PBD indicates that all teams have the same possibility of winning a match and ultimately the championship. This result can be obtained if all teams tie all matches or, in a double round-robin system with or without possibility of a tie, if each team wins and loses matches against the same rival.

All these issues are well known in sports economics. Our interest will henceforth focus on calculating the distribution of scores that generates the maximum value of HHI, that is HHI_{max}.

2.3.Perfectly Unbalanced Distribution of Scores

The determination of the **minimum** CB is more complicated. Although a market can be monopolised by a firm, which entails maximum concentration, this is not the case in a sports competition, in which CB can be measured over time or a season. When measured over time, if the same team won all the championships, there would be a *monopoly* during the period analysed and CB would reach its minimum value. When measured over a season, a team cannot win more matches than those played, and it is not possible a monopoly situation. That is, the distribution of the results cannot be such that the winner of the competition accumulates all the wins as well as all the points of the other teams (except in the unreal case of a league comprising two teams). The reason for this is simple. The vector of the sizes of the firms in a market is the result of a competition process in which all agents interact with each other in multiple ways. Nevertheless, the vector of the total wins or points gained in a league is the result of a competition process between all teams considered as pairs. This bilaterality prevents teams from accumulating all the points because they cannot play all the matches.

The PUD has been characterized by Fort & Quirck (1997), Gayant & Le Pape (2012 and 2015), Horowitz (1997), Larsen, Fenn & Spenner (2006), Owen, Ryan & Weatherston (2007), and Utt & Fort (2002). In this distribution, each team has defeated all those below them and they have lost against all the teams above them.

If the number of teams does not vary and there are no ties (e.g. as in baseball or basketball), the number of total points distributed between the teams at the end of the championship does not vary. This is because the number of matches is already established and, given the scoring system used, there will be always a winner in each match.

Once the tournament is over and the positions have been ordered, the results of distribution can be visualized as being *in cascade* in which each team has fewer points than the team in the preceding position. Figure 1 shows it for N=20 with a {2,1,0} pattern. Really, this distribution has rarely been formalized (Gayant & Le Pape, 2015). We are going to do it.

Firstly, the points obtained by the teams would follow the form:

$$p_i = p_w \cdot (N - i) \; for \; i = 1 \ldots N.$$

Secondly, the reason for such result is that, given the number of points in the championship: $\sum_{i=1}^{N} p_i$, we have been assuming that the marginal value de HHI is:

$$\frac{\partial \sum_{i=1}^{N} s_i^2}{\partial s_i} = \frac{2}{\sum_{i=1}^{N} p_i} \cdot p_i$$

That is, when $\sum_{i=1}^{N} p_i$ is constant, the marginal value increases with the team's scores. The maximum number of points that any team can obtain will be restricted by maximum number of matches won. This is the *principle of saturation*.

Thus, having established the number of points at stake, the distribution that generates the highest concentration is built in such a way that a team accumulates points up to the maximum possible (limited by the aforementioned restriction, that is, the principle of saturation) and so on. The contribution to the concentration of each point increases in relation to the points won by the team that obtains this point, which is assigned to the team with the highest share up to the restriction imposed by the bilateral competition with the other teams. This result is applicable to the case described by the aforementioned authors, who addressed scoring patterns of the type $\{p_w = 2 \cdot p_t, p_t, p_l = 0\}$, where each team has $p_w(N-i)$ points in a single round-robin (and twice as many points in a double round-robin).

Thirdly, the distribution of results will be $p = \{p_w(N-1), p_w(N-2), p_w(N-3), \dots, p_w, 0\}$. Given the total points $\sum_{i=1}^{N} p_i = \sum_{i=1}^{N} p_w\,(N-i) = p_w\,N(N-1)/2$, the vector of shares will be:

$$s^{PUD} = \left(\frac{1}{N}, \frac{N-2}{N(N-1)}, \frac{N-3}{N(N-1)}, \dots, \frac{1}{N(N-1)}, 0\right)$$

The share of the winner will be precisely the one that any team in a PBD would have. Borooah & Mangan (2012) obtain this result in a double round-robin league.

On the other hand, if there are ties (e.g., as in soccer or rugby), the number of points obtained by all teams at the end of the championship, given N, will vary as a function of the number of ties, as seen above, and the distribution of points will be unstable, except if $p_t = \frac{p_w + p_l}{2}$, which would be the case for a pattern $\{2 \cdot p_t, p_t, 0\}$ (Borooah & Mangan, 2012; Gayant & Le Pape, 2012 and 2015).

Therefore, there is instability under the patterns $2 \cdot p_t \neq p_w + p_l$, and particularly under pattern $p_w > 2 \cdot p_t$ with $p_l = 0$, which implies that when computing the CB of a league, the use of defined relative indexes (e.g., the HHI or standard deviation) can lead to higher concentrations than those corresponding to the *complete-cascade distribution* (Gayant & Le Pape, 2015). In this case, what is the PUD? And, how can we apply the principle of saturation?

Given that for the same number of matches won, the total sum of points is constant, the principle of saturation implies that the matches won are accumulated at the top teams in the ranking. The remaining matches are tied. Besides, if a team wins a match, the highest HHI value results from winning all of its matches to the lowest ranked teams.

As mentioned above and discussed below, the reconstruction of results alters the mechanism of incentives of the championship, and can also have cardinal and ordinal impacts on the positions of the teams in each league and between leagues, which effect is not compatible with the aim of comparing levels of CB.

Nevertheless, there is no reason for this to be the case. The real issue is that the *complete-cascade distribution* is not the PUD in all cases. *Complete-cascade distribution* is the least balanced when the scoring system is constructed without the possibility of ties or with a pattern that complies the stability condition. However, with another pattern, the PUD is other than the *complete-cascade distribution*. Given N (which is known since the beginning), and the scoring pattern, this distribution can be known and predicted. Consequently, if we know the distribution, there will be no problem in using a pattern, such as {3,1,0}, in order to measure CB without the having to reconstruct the results. We dedicate the following Section to identify such distribution.

3. Characterisation of Distribution which Generates the Minimum Competitive Balance: Truncated-Cascade Distribution

In this Section, we formalize the HHI for N teams according to the scoring system. We demonstrate that, for the total number of points earned in the competition, the PUD can be obtained by accumulating the points of the winning match in the first q teams of the results table. We show that HHI has a unique maximum for each N denominated as HHI_{max}. The value of this maximum is obtained using a recurrent procedure. Finally, we characterize the value of q, which determines HHI_{max} using a fifth-degree polynomial equation. For each N, the q value is calculated by iterative methods for solving equations. To perform these calculations, MsExcel offers features such us "Goal Seek" within "Data" tab or the complement "Solver".

3.1. The Distribution of Scores

We define a highly asymmetric final distribution of points when the champion has beaten the other teams, the runner-up has beaten all the teams below them in the final ranking, and so on up to a position (q) from which the teams have tied all their remaining matches. Two groups of teams can emerge: teams that have won at least one match, and those that have tied all the matches that they have not lost to the teams in the preceding positions.

For every $0 \leq q \leq N-1$, with $q \in \mathbb{N}$, we obtain a generalized distribution that we will call the *truncated-cascade in* (q) *distribution*. If $q = 0$, all the teams tie all those matches and we obtain the PBD. If $q = N-1$, we obtain the PUD that we call *complete-cascade distribution,* which constitutes a particular case of *truncated-cascade in (q) distribution.*

Under these conditions, using a pattern $P = \{p_w, p_t, 0\}$, the points obtained by the teams would follow the form:

$$p_i = \begin{cases} p_w \cdot (N-i) \; for \; i = 1 \ldots q \\ p_t \cdot (N-q-1) \quad for \; i = q+1 \ldots . N \end{cases}$$

The teams that occupied the first (q) positions won all matches except for those played against teams preceding them in the table, whereas the teams occupying $(N-q)$ positions tied all matches.

3.2.The Corresponding Herfindahl-Hirschman Index

Therefore, for a *truncated-cascade in (q) distribution* and {3,1,0} pattern we can obtain the following expression of HHI (see demonstration in Appendix A):

$$HHI(q)=\frac{2\cdot q^3+\left(-6\cdot N+\frac{5}{2}\right)\cdot q^2+\left(6\cdot N^2-5\cdot N+\frac{1}{2}\right)\cdot q+N\cdot(N-1)^2}{\left(\frac{-q^2-(1-2\cdot N)\cdot q+2\cdot N\cdot(N-1)}{2}\right)^2}$$

Based on this, we wish to calculate the critical value (q^*) that forces the HHI reach its maximum value with a given N and scoring pattern.

3.3.The Critical Value of *q*

We seek the unique critical value (q^*) that makes the value of $HHI=\sum_{i=1}^{N}s_i^2$ reach its maximum. To this end, we build a recurrent formulation of the above expression to determine the evolution of HHI for two consecutive states of q, with $0 \leq q \leq N-1$ (see Appendix B). Therefore, for each N, the value of q^* results from solving the following inequality:

$$0>\left[-\frac{1}{2}\cdot q^5+\left(\frac{5}{2}\cdot N-\frac{9}{4}\right)\cdot q^4+\left(-8\cdot N^2+12\cdot N-\frac{7}{2}\right)\cdot q^3+\left(11\cdot N^3-27\cdot N^2+18\cdot N-\frac{9}{4}\right)\cdot q^2+\left(8\cdot N^4-6\cdot N^3-9\cdot N^2+\frac{15}{2}\cdot N-\frac{1}{2}\right)\cdot q-4\cdot N^5+12\cdot N^4-13\cdot N^3+6\cdot N^2-N\right]$$

So far, we developed a formalization of the index HHI for N teams according to the scoring system. In summary: (i) for the number of total points earned in the competition, the PUD can be obtained by accumulating the points of the winning match in the first q teams of the results table; (ii) HHI has a unique maximum for each N denominated as HHI_{max}. The value of this maximum is obtained using a recurrent procedure. And (iii) q^* is value used to determine HHI_{max} from a fifth-degree polynomial equation.

3.4.Calculus and Results

Using MsExcel, we developed a complementary tool in the form of a group of automated instruments to confirm the theoretical results already formulated and obtain the value of q^* for any N. We also designed synopsis tables with the information needed to conduct the appropriate comparisons. Likewise, we developed macros that accelerate the calculations.

These macros are based on the numerical optimisation and approximation techniques provided in the MsExcel toolset. The macros were parameterized for any scoring system.

Table 2 shows the results of this calculation for a {3,1,0} pattern for each N up to a total 50 teams. The value of (q^*) increases as the number of teams increases and, for each N, the value of HHI increases as a function of the increasing truncation of the *cascade* up to an identifiable critical value. The maximum values of the index for each *N* are shown in bold in Table 2, which shows the values representing the truncation of the PUD.

Figure 2 shows *truncated-cascade in (q=7) distribution* with {3,1,0} pattern for N=20. The value of the HHI in $q^*=7$ is 0.075402 under the *truncated-cascade distribution*, whereas the value of the HHI is 0.068421 under the *complete-cascade distribution*. We can visualize the differences with respect to the *complete-cascade distribution* in Figure 1.

For both scoring patterns, {2,1,0} and {3,1,0}, Table 3 shows the maximum values of the HHI and the critical value q^* corresponding to with the *complete-cascade distribution* and *truncated- cascade distribution*. Using these values, we can standardize the HHI corresponding to the distribution of points obtained.

Thus, with a {2,1,0} pattern, the PUD is always the one that we have called *in complete-cascade*. The critical value is q^*=N-1. It should be noted that, in this case, the total number of points does not vary.

Nevertheless, this is not the case under a scoring pattern {3,1,0}, such as the one used in the major European leagues and in the UEFA Champions League (UCL). Although it is true that the PUD changes according to the number of teams and scoring system, this distribution can be predicted before the beginning of the championship. For instance, in the qualifying round of the UCL, in which there are four teams per group, the PUD corresponds to the case in which one team wins all the matches and the other teams tie all the matches except for the one they lost to the winning team ($q^*=1$). So HHI_{max}=0.413333.

4. Empirical Application to the Major European Soccer Leagues: The Negative Effects of Recalculating the Scores

The five major European soccer leagues are Premier League in England, Primera División in Spain, Serie A in Italy, Bundesliga in Germany and Ligue 1 in France. These leagues constitute the main nucleus of soccer in Europe and form the basis of the classification mechanism for the UCL. During the study period (1997-2017), the number of teams increased from 18 to 20 in the 2004-2005 season in Italy and in the 2002-2003 season in France. All the leagues have adopted the {3,1,0} scoring system.

Notice that in any of these leagues, the PUD is the one characterized by q^*=7, which represents a *cascade* distribution up to the team in the seventh position and all ties for the other teams. This distribution is the case in the English, Italian, Spanish, and French leagues, which have 20 teams, and in the German league, which has 18 teams. Nevertheless, the change in N affects the maximum value that HHI can reach: in the first four leagues HHI_{max} is 0.0754015 and in the German league it is 0.0839378 (Table 3). Therefore, the results make it possible to use the measurements of the CB when it is defined on the basis of the HHI_{NORM} with {3,1,0}.

Notice that under the pattern {2,1,0}, the maximum values of HHI are always less than those under the pattern {3,1,0}. Therefore, if any league developed under the pattern {3,1,0} is rebuilt under the pattern {2,1,0}, then the value of the CB will always be less than its real value.

For each of these major leagues, we calculated the standardized HHI on the basis of the shares defined according to the current scoring pattern, {3,1,0}, and the number of teams in each season. Besides, we have recalculating the results for each league and season according to the {2,1,0} pattern and, once again, se have calculated the HHI_{NORM}.

Table 4 shows the HHI_{NORM} values calculated for each season and league. The scoring patterns {3,1,0} and {2,1,0} are shown in separate columns. The table shows the five different leagues referred to and 20 seasons (1997/98 - 2016/2017). For each season (rows), competition (columns), and scoring pattern two values are shown: HHI_{NORM} values and the ranking between leagues achieved in terms of concentration (shown in parentheses).

Changes of ranking between competitions are relevant and are shaded in the table. These changes affect all the leagues and half of the seasons. We quantified 23 changes of positions. Therefore, it would not be neutral to rebuild the results by changing the scoring pattern. Furthermore, recalculating the scores may suppose changes in final ordering of teams in a league, as we have seen in Table 1.

We have provided a general expression for the HHI as a function of the truncating value *q*, thus obtaining the *q** that defines the maximum value of HHI, and therefore the PUD. Furthermore, given N, for the pattern {3,1,0} is possible to obtain the maximum values of the HHI, thus avoiding having to recalculate the final results on the basis of another scoring pattern. We have proven that this recalculation generates negative effects.

5. Conclusions

Competitive balance in a sports competition is usually obtained from the distribution of points (or wins) achieved by each team at the end of the championship. From this distribution, indexes such as the HHI are calculated, which require standardization in order to take account of the special characteristics of sports competitions. Standardization requires know the maximum and minimum theoretical values of the index.

The distribution that has usually been considered to generate the minimum competitive balance, which we have called Complete-cascade distribution, is not valid for any points award pattern. So, measuring competitive balance is affected by the scoring system used.

This distribution is valid if there are no ties in the competition or if the winner's remuneration is equal to twice the remuneration of the tied team. Therefore, this is the condition of stability.

If the condition of stability is not fulfilled, as is often the case in football leagues, *Complete-cascade distribution* does not always generate the minimum competitive balance. In this case, the literature has proposed reconstructing the scores, based on the results achieved, with a pattern that complies the stability condition.

This article allows us to conclude that the reconstruction of the results generates negative effects in ordinal and cardinal terms. Cardinal effects would affect possible modelling in comparative studies, as the value of the indices changes. The ordinal effects are even more severe as they can affect to (a) the leagues ranking for a season, and (b) to the ranking between teams in the same league in a season.

These effects has been showed. We verified for the major European soccer leagues over 20 seasons that the recalculation of HHI using a scoring pattern other than the real one causes changes, both in cardinal and ordinal terms. Ordinal changes potentially affect both the selection of teams to play in the UCL or the UEL, as well as relegation from category.

Reconstructing the results, with a pattern fulfilling the stability condition, implies that the resulting HHI_{max} is always less than the value of the HHI_{max} corresponding to the pattern actually used. Therefore, the value of the competitive balance will always be less than the actual value, since the value of HHI_{NORM} will be greater.

We have characterized a new distribution, which we call *Truncated-cascade distribution*, that allows obtaining the HHI_{max} for patterns that do not fulfill the stability condition as is the case with {3,1,0} pattern. This distribution generalizes the *complete-cascade distribution*.

Truncated-cascade distribution can be defined before the competition, which avoids the reconstruction of results. So, we can calculate competitive balance with the actual pattern.

We demonstrated that the measurement of CB does not have to involve recalculating the scoring pattern on the basis of results that, ultimately, would generate cardinal and especially ordinal negative effects.

References

Andreff, W. (2015). *Disequilibrium sports economics. Competitive imbalance and budget constraints*. Cheltenham, UK: Edward Elgar Publishing Ltd.

Borooah, V.K., & Mangan, J. (2012). Measuring competitive balance in sports using generalized entropy with an application to English premier league football. *Applied Economics*, 44(9), 1093-1102, DOI: 10.1080.00036846.2010.537638.

Depken, C.A. (1999). Free-agency and the competitiveness of major league baseball. *Review of Industrial Organization*, 14, 205-217.

Eckard, E.W. (2001). Free agency, competitive balance and diminishing returns to pennant competition. *Economic Inquiry*, 39(3), 430-443.

Fort, R., & Quirck, J. (1995). Cross-subsidization, incentives and outcomes in professional team sports leagues. *Journal of Economic Literature*, 33, 1265-1299.

Herfindahl, O.C. (1950). *Concentration in the US Steel Industry* (Unpublished PhD dissertation). Columbia University.

Hirschman, A.O. (1945). *National power and the structure of foreign trade*. Berkeley: University of California Press.

Hirschman, A.O. (1964). The paternity of an index. *American Economic Review*, 54, 761-762.

Gayant, J.P., & Le Pape, N. (2012). How to account for changes in the size of Sports Leagues: The Iso Competitive Balanced Curves. *HAL archives-ouvertes*, halshs-00856122.

Gayant, J.P., & Le Pape, N. (2015). The metrics of competitive imbalance. In W. Andreff (Ed.), *Disequilibrium sports economics. Competitive imbalance and budget constraints*. (pp. 104-130). Cheltenham, UK: Edward Elgar Publishing Ltd.

Horowitz, I. (1997). The increasing competitive balance in Major League Baseball. *Review of Industrial Organization*, 12, 373-387.

Humphreys, B.R. (2002). Alternative measures of competitive balance in sports leagues. *Journal of Sports Economics*, 9, 191-210.

Larsen, A., Fenn, A.J., & Spenner, E.L. (2006). The impact of free agency and the salary cap on competitive balance in the National Football League. *Journal of Sports Economics*, 7(4), 374-390, DOI: 10.1177/1527002505279345.

Owen, P.D., Ryan, M., & Weatherston, C.R. (2007). Measuring competitive balance in professional team sports using Herfindahl-Hirschman Index. *Review of Industrial Organization*, 31, 289-302, DOI: 10.1007/s11151-008-9157-0.

Pawlowski, T., Breuer, C., & Hovemann, A. (2010). Top clubs´ performance and the competitive situation in European domestic football competitions. *Journal of Sports Economics*, 11(2), 186-202, DOI: 10.1177/1527002510363100.

Schmidt, M.B. (2001). Competition in Major League Baseball: The impact of expansion. *Applied Economic Letters*, 8, 21-26.

Schmidt, M.B., & Berri, D.J. (2001). Competitive balance: The case of Major League Baseball. *Journal of Sports Economics*, 2(2), 145-167.

Utt, J., & Fort, R. (2002). Pitfalls to measuring competitive balance with Gini coefficients. *Journal of Sports Economics*, 3(4), 367-373, DOI: 101177/152700202237502.

Zimbalist, A. (2002). Competitive balance in sport leagues: an introduction. *Journal of Sports Economics*, 32(2), 111-121.

Tables

Table 1. Summary of results due to change the scoring system {3, 1, 0} to {2, 1, 0} in the major European soccer leagues.

	Primera División (Spain)	Premier League (England)	Serie A (Italy)	Ligue 1 (France)	Bundesliga (Germany)	Total
Affected competitions (of total)	19/20	20/20	19/20	19/20	19/20	96/100
Affected teams (of total)	110/400	90/400	90/386	129/390	43/360	462/1936
Potencial classification for UCL and UEL (of total)	12/40	7/40	5/40	13/40	9/40	46/200
Potencial relegations from category (of total)	9/20	2/20	7/20	4/20	5/20	27/100

Table 2. Values of the HHI as a function of the number of teams (N) and of matches won in cascade (q). Scoring pattern {3, 1, 0}

(Maximum of HHI in bold)

	Matches won in cascade (q)									
Numbers of teams (N)	0	1	2	3	4	5	6	7	8	9
2	0.500000	**1.000000**								
3	0.333333	**0.593750**	0.555556							
4	0.250000	**0.413333**	0.411765	0.388889						
5	0.200000	0.312500	**0.325103**	0.312723	0.300000					
6	0.166667	0.248980	**0.266272**	0.261905	0.252066	0.244444				
7	0.142857	0.205729	0.223923	**0.224377**	0.218333	0.211238	0.206349			
8	0.125000	0.174603	0.192187	**0.195398**	0.192308	0.187014	0.181884	0.178571		
9	0.111111	0.151250	0.167658	**0.172390**	0.171387	0.167820	0.163537	0.159752	0.157407	
10	0.100000	0.133150	0.148223	0.153740	**0.154167**	0.152000	0.148729	0.145317	0.142459	0.140741
11	0.090909	0.118750	0.132504	0.138366	**0.139757**	0.138667	0.136316	0.133499	0.130773	0.128569
12	0.083333	0.107047	0.119569	0.125514	**0.127543**	0.127262	0.125683	0.123472	0.121094	0.118895
13	0.076923	0.097364	0.108767	0.114638	0.117080	**0.117400**	0.116445	0.114781	0.112803	0.110803
14	0.071429	0.089231	0.099629	0.105336	0.108033	**0.108797**	0.108338	0.107143	0.105553	0.103818
15	0.066667	0.082310	0.091812	0.097305	0.100148	**0.101235**	0.101168	0.100365	0.099125	0.097667
16	0.062500	0.076355	0.085060	0.090312	0.093225	0.094544	**0.094785**	0.094307	0.093373	0.092176
17	0.058824	0.071181	0.079175	0.084178	0.087107	0.088591	**0.089071**	0.088860	0.088189	0.087230
18	0.055556	0.066645	0.074007	0.078761	0.081669	0.083266	0.083931	**0.083938**	0.083491	0.082742
19	0.052632	0.062639	0.069437	0.073947	0.076809	0.078480	0.079287	**0.079471**	0.079213	0.078649
20	0.050000	0.059076	0.065369	0.069645	0.072444	0.074159	0.075074	**0.075402**	0.075302	0.074899
21	0.047619	0.055888	0.061728	0.065781	0.068506	0.070242	0.071238	0.071682	**0.071716**	0.071450
22	0.045455	0.053020	0.058452	0.062294	0.064938	0.066679	0.067734	0.068270	**0.068415**	0.068269
23	0.043478	0.050426	0.055491	0.059133	0.061693	0.063425	0.064522	0.065133	**0.065370**	0.065326
24	0.041667	0.048070	0.052802	0.056257	0.058731	0.060445	0.061570	0.062239	0.062554	**0.062596**
25	0.040000	0.045920	0.050350	0.053631	0.056018	0.057707	0.058849	0.059564	0.059942	**0.060059**

	9	10	11	12	13	14	15	16	17	18
26	**0.057695**	0.057679	0.057511	0.057227	0.056856	0.056425	0.055955	0.055463	0.054966	0.054479
27	0.055489	**0.055533**	0.055427	0.055206	0.054897	0.054523	0.054105	0.053658	0.053199	0.052741
28	0.053425	**0.053521**	0.053472	0.053307	0.053053	0.052733	0.052363	0.051961	0.051540	0.051112
29	0.051492	**0.051633**	0.051633	0.051519	0.051315	0.051043	0.050720	0.050360	0.049976	0.049580
30	0.049678	0.049858	**0.049901**	0.049832	0.049674	0.049447	0.049167	0.048847	0.048499	0.048135
31	0.047973	0.048187	**0.048267**	0.048239	0.048122	0.047936	0.047696	0.047413	0.047100	0.046767
32	0.046368	0.046611	0.046724	**0.046732**	0.046653	0.046504	0.046300	0.046053	0.045773	0.045470
33	0.044855	0.045122	0.045265	**0.045305**	0.045260	0.045145	0.044974	0.044760	0.044511	0.044237
34	0.043426	0.043715	0.043884	**0.043952**	0.043937	0.043853	0.043714	0.043529	0.043310	0.043063
35	0.042076	0.042383	0.042574	0.042668	**0.042680**	0.042624	0.042513	0.042357	0.042165	0.041944
36	0.040799	0.041121	0.041331	0.041447	**0.041484**	0.041454	0.041369	0.041239	0.041072	0.040875
37	0.039588	0.039924	0.040151	0.040286	**0.040345**	0.040338	0.040277	0.040171	0.040027	0.039854
38	0.038440	0.038786	0.039028	0.039181	0.039259	**0.039274**	0.039234	0.039150	0.039029	0.038876
39	0.037350	0.037705	0.037959	0.038128	0.038223	**0.038257**	0.038237	0.038174	0.038072	0.037940
40	0.036314	0.036675	0.036940	0.037123	0.037234	**0.037285**	0.037284	0.037239	0.037156	0.037043
41	0.035327	0.035694	0.035969	0.036163	0.036289	0.036355	**0.036371**	0.036343	0.036278	0.036182
42	0.034388	0.034759	0.035042	0.035247	0.035385	0.035465	**0.035495**	0.035483	0.035435	0.035355
43	0.033492	0.033867	0.034156	0.034370	0.034519	0.034612	0.034656	**0.034659**	0.034625	0.034561
44	0.032638	0.033014	0.033309	0.033531	0.033690	0.033794	0.033851	**0.033867**	0.033847	0.033797
45	0.031821	0.032199	0.032498	0.032727	0.032895	0.033009	0.033078	**0.033106**	0.033099	0.033061
46	0.031041	0.031420	0.031722	0.031956	0.032132	0.032256	0.032335	0.032374	**0.032379**	0.032353
47	0.030294	0.030673	0.030978	0.031217	0.031400	0.031532	0.031620	0.031670	**0.031685**	0.031671
48	0.029579	0.029957	0.030264	0.030508	0.030697	0.030836	0.030933	0.030992	**0.031017**	0.031014
49	0.028894	0.029271	0.029580	0.029827	0.030021	0.030167	0.030271	0.030339	0.030373	**0.030380**
50	0.028238	0.028613	0.028922	0.029172	0.029371	0.029523	0.029634	0.029709	0.029752	**0.029768**

Table 3. Perfectly Unbalanced distributions for different numbers of teams and scoring patterns

	Scoring pattern P={2, 1, 0}		Scoring pattern P={3, 1, 0}	
Number of teams (N)	Cascade (q=N-1)	Maximum of HHI	Truncated cascade (q*)	Maximum of HHI
2	1	1.0000000	1	1.0000000
3	2	0.5555556	1	0.5937500
4	3	0.3888889	1	0.4133333
5	4	0.3000000	2	0.3251029
6	5	0.2444444	2	0.2662722
7	6	0.2063492	3	0.2243767
8	7	0.1785714	3	0.1953981
9	8	0.1574074	3	0.1723899
10	9	0.1407407	4	0.1541667
11	10	0.1272727	4	0.1397569
12	11	0.1161616	4	0.1275433
13	12	0.1068376	5	0.1174003
14	13	0.0989011	5	0.1087967
15	14	0.0920635	5	0.1012346
16	15	0.0861111	6	0.0947846
17	16	0.0808824	6	0.0890706
18	17	0.0762527	7	0.0839378
19	18	0.0721248	7	0.0794709
20	19	0.0684211	7	0.0754015

Table 4. CB in main major European leagues and league ranking by season
Standardized HHI (League ranking by season)

	Primera División (Spain)		Premier League (England)		Serie A (Italy)		Ligue 1 (France)		Bundesliga (Germany)	
Season	(2-1-0)	(3-1-0)	(2-1-0)	(3-1-0)	(2-1-0)	(3-1-0)	(2-1-0)	(3-1-0)	(2-1-0)	(3-1-0)
1997/98	0.1387 (2)	0.1193 (3)	0.1179 (4)	0.1058 (4)	0.2456 (1)	0.2156 (1)	0.1362 (3)	0.1225 (2)	0.1024 (5)	0.0825 (5)
1998/99	0.1415 (5)	0.1273 (5)	0.1654 (2)	0.1395 (3)	0.1527 (4)	0.1314 (4)	0.1584 (3)	0.1586 (2)	0.1793 (1)	0.1668 (1)
1999/00	0.0754 (4)	0.0727 (4)	0.2034 (2)	0.1757 (2)	0.2144 (1)	0.1972 (1)	0.0614 (5)	0.0559 (5)	0.1664 (3)	0.1493 (3)
2000/01	0.1184 (3)	0.1079 (3)	0.1464 (2)	0.1353 (2)	0.2015 (1)	0.1768 (1)	0.1060 (4)	0.0906 (4)	0.1022 (5)	0.0873 (5)
2001/02	0.0923 (5)	0.0783 (5)	0.2120 (2)	0.1929 (2)	0.2325 (1)	0.1936 (1)	0.1099 (4)	0.0949 (4)	0.2113 (3)	0.1882 (3)
2002/03	0.1246 (3)	0.1135 (3)	0.1880 (2)	0.1593 (2)	0.1927 (1)	0.1769 (1)	0.1164 (5)	0.1016 (5)	0.1195 (4)	0.1041 (4)
2003/04	0.1194 (5)	0.1080 (5)	0.1769 (3)	0.1618 (3)	0.2735 (1)	0.2610 (1)	0.1521 (4)	0.1351 (4)	0.2088 (2)	0.1701 (2)
2004/05	0.1618 (3)	0.1426 (3)	0.2128 (1)	0.2058 (1)	0.1494 (4)	0.1405 (4)	0.0868 (5)	0.0781 (5)	0.1780 (2)	0.1593 (2)
2005/06	0.1581 (4)	0.1503 (4)	0.2692 (2)	0.2293 (2)	0.3225 (1)	0.2754 (1)	0.1476 (5)	0.1327 (5)	0.1847 (3)	0.1790 (3)
2006/07	0.1370 (3)	0.1249 (4)	0.1929 (2)	0.1731 (2)	0.2112 (1)	0.2046 (1)	0.0799 (5)	0.0793 (5)	0.1331 (4)	0.1270 (3)
2007/08	0.1519 (4)	0.1369 (4)	0.3133 (1)	0.2693 (1)	0.1838 (2)	0.1732 (2)	0.1314 (5)	0.1170 (5)	0.1703 (3)	0.1449 (3)
2008/09	0.1598 (5)	0.1411 (5)	0.2600 (1)	0.2287 (1)	0.1799 (3)	0.1659 (3)	0.1780 (4)	0.1538 (4)	0.1987 (2)	0.1787 (2)
2009/10	0.2406 (2)	0.2369 (1)	0.2491 (1)	0.2259 (2)	0.1729 (5)	0.1507 (5)	0.1825 (4)	0.1568 (3)	0.1842 (3)	0.1546 (4)
2010/11	0.2090 (1)	0.1868 (1)	0.1259 (4)	0.1154 (4)	0.1678 (2)	0.1553 (2)	0.1113 (5)	0.1011 (5)	0.1499 (3)	0.1291 (3)
2011/12	0.2068 (2)	0.1916 (3)	0.2258 (1)	0.2075 (1)	0.1755 (4)	0.1567 (4)	0.1555 (5)	0.1470 (5)	0.2015 (3)	0.1954 (2)
2012/13	0.2376 (1)	0.2113 (3)	0.2376 (2)	0.2230 (1)	0.2293 (4)	0.2047 (4)	0.1468 (5)	0.1275 (5)	0.2368 (3)	0.2183 (2)
2013/14	0.2553 (4)	0.2252 (4)	0.2840 (2)	0.2464 (3)	0.2902 (1)	0.2566 (1)	0.2045 (5)	0.1826 (5)	0.2699 (3)	0.2477 (2)
2014/15	0.3263 (1)	0.3009 (1)	0.2030 (3)	0.1824 (3)	0.2034 (2)	0.1888 (2)	0.1872 (4)	0.1493 (5)	0.1662 (5)	0.1582 (4)
2015/16	0.2361 (1)	0.2232 (1)	0.1992 (4)	0.1671 (4)	0.2254 (2)	0.2037 (3)	0.1737 (5)	0.1605 (5)	0.2237 (3)	0.2046 (2)
2016/17	0.3246 (2)	0.2881 (1)	0.3117 (3)	0.2643 (3)	0.3305 (1)	0.2879 (2)	0.2321 (4)	0.2073 (4)	0.1992 (5)	0.1634 (5)

Figures

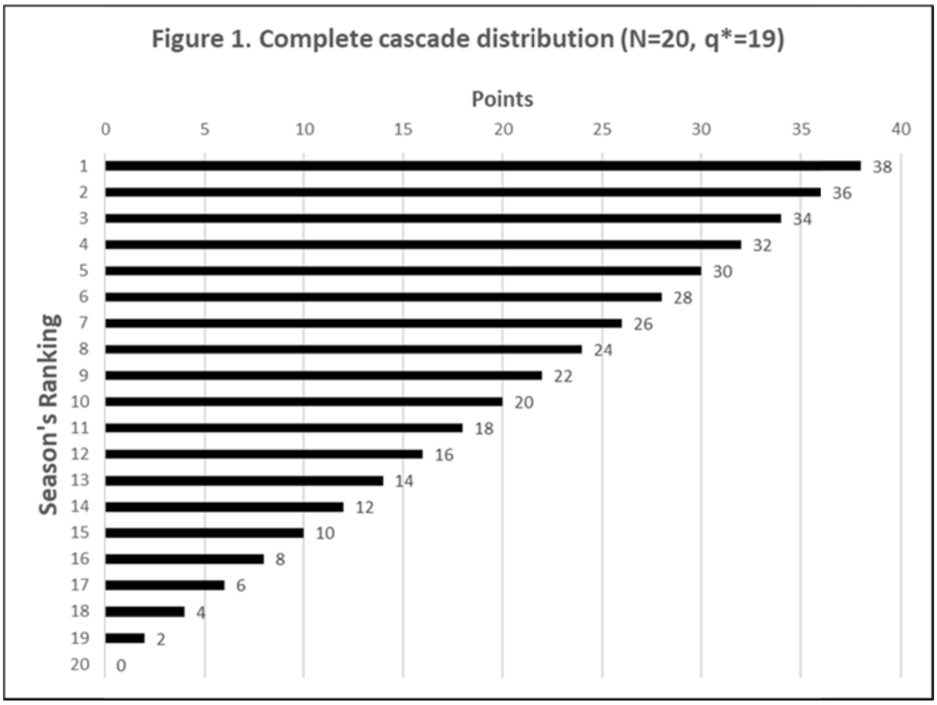
Figure 1. Complete cascade distribution (N=20, q*=19)
Points
0 5 10 15 20 25 30 35 40
Season's Ranking
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
38 36 34 32 30 28 26 24 22 20 18 16 14 12 10 8 6 4 2 0

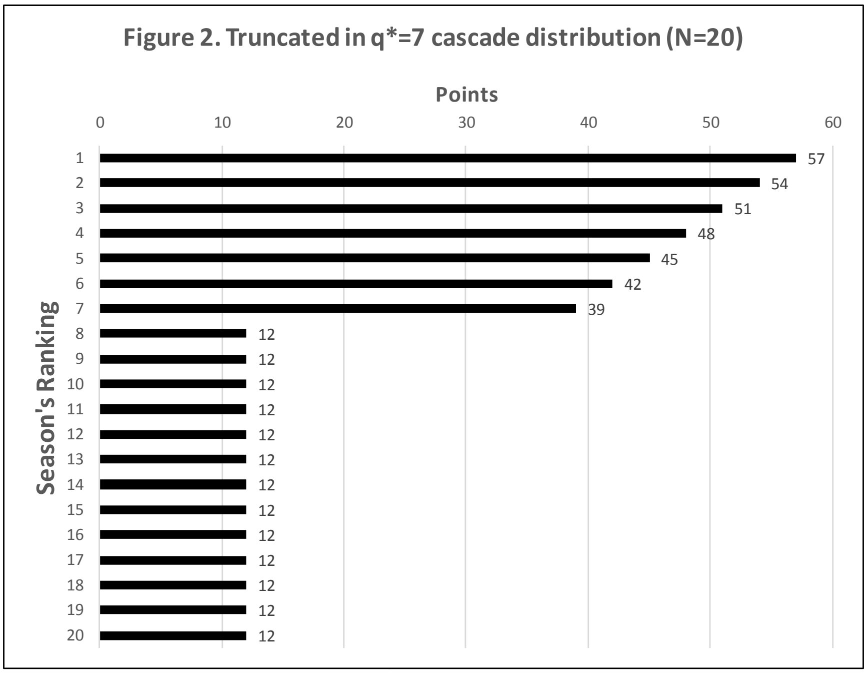
Figure 2. Truncated in q*=7 cascade distribution (N=20)
Points
0 10 20 30 40 50 60
Season's Ranking
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
57 54 51 48 45 42 39 12 12 12 12 12 12 12 12 12 12 12 12 12

Appendix A. Characterization of the HHI as a function of the number of teams (N) and the value defining the frontier between winning and tying teams (q).

We can obtain HHI = $\sum_{i=1}^{n} s_i^2 = \frac{\sum_1^n p_i^2}{(\sum_1^n p_i)^2}$ from the following expressions. The sum of points, whose square appears in the **denominator**, would be:

$$\sum_{i=1}^{N} p_i = \sum_{i=1}^{q} p_w \cdot (N-i) + \sum_{i=q+1}^{N} p_t \cdot (N-q-1) = p_w \cdot N \sum_{i=1}^{q} 1 - \sum_{i=1}^{q} p_w \cdot i + p_t \cdot (N-q-1) \cdot \sum_{i=q+1}^{N} 1 = \frac{p_w \cdot q \cdot (2 \cdot N - q - 1)}{2} + p_t \cdot (N-q-1) \cdot (N-q)$$

$$= \frac{(2 \cdot P_t - P_w) \cdot q^2 + (2 \cdot P_t - P_w) \cdot (1 - 2 \cdot N) \cdot q + 2 \cdot P_t \cdot N \cdot (N-1)}{2}$$

where we have required the general expression of the sum of an arithmetic progression from 1 to q. The first addend is the number of points obtained by the teams that have won *in cascade*. This addend is also the number of wins multiplied by the points awarded. The second addend is the number of points accumulated by the teams that have tied. This addend is also the number of ties multiplied by the award.

The **numerator** is:

$$\sum_{i=1}^{N} p_i^2 = \sum_{i=1}^{q} \left(p_w \cdot (N-i)\right)^2 + \sum_{i=q+1}^{N} \left(p_t (N-q-1)\right)^2 = p_w^2 \cdot \left[\sum_{I=1}^{q} (N^2 - 2 \cdot N \cdot i + i^2)\right] + p_t^2 (N-q-1)^2 \cdot \sum_{i=q+1}^{N} 1$$

$$= p_w^2 \cdot \left[\sum_{i=1}^{q} N^2 - 2 \cdot N \cdot \sum_{i=1}^{q} i + \sum_{i=1}^{q} i^2\right] + p_t^2 (N-q-1)^2 \cdot (N-q)$$

Given that:

$\sum_{i=1}^{q} i = \frac{q \cdot (1+q)}{2}$ and $\sum_{i=1}^{q} i^2 = \frac{q \cdot (q+1) \cdot (2 \cdot q + 1)}{6}$, we have that:

$$\sum_{i=1}^{q} N^2 - 2\cdot N\cdot\sum_{i=1}^{q} i + \sum_{i=1}^{q} i^2 = N^2\cdot q - 2\cdot N\cdot\frac{q\cdot(1+q)}{2} + \frac{q\cdot(q+1)\cdot(2\cdot q+1)}{6}$$

Then:

$$\sum_{i=1}^{N} p_i^2 = p_w^2\left(N^2\cdot q - N\cdot q\cdot(1+q) + \frac{q\cdot(q+1)\cdot(2\cdot q+1)}{6}\right) + p_t^2\cdot(N-q-1)^2\cdot(N-q)$$

And, **therefore**:

$$\sum_{i=1}^{N} s_i^2 = \frac{p_w^2\cdot\left(N^2\cdot q - N\cdot q\cdot(1+q) + \frac{q\cdot(q+1)\cdot(2\cdot q+1)}{6}\right) + p_t^2\cdot(N-q-1)^2\cdot(N-q)}{\left(\frac{p_w\cdot q\cdot(2\cdot N-q-1)}{2} + p_t\cdot(N-q-1)\cdot(N-q)\right)^2}$$

The above expressions can be reformulated as a function of (q), obtaining an expression of the HHI as a function of a quotient of polynomials in (q), whose terms depend on the scoring pattern and on the number of teams:

$$\sum_{i=1}^{N} s_i^2 = \frac{\sum_1^N p_i^2}{(\sum_1^N p_i)^2} =$$

$$= \frac{\left(\frac{p_w^2}{3} - p_t^2\right)\cdot q^3 + \left((3\cdot p_t^2 - p_w^2)\cdot N + \frac{1}{2}p_w^2 - 2\cdot p_t^2\right)\cdot q^2}{\left(\frac{(2\cdot p_t - p_w)\cdot q^2 + (2\cdot p_t - p_w)\cdot(1-2\cdot N)\cdot q + 2\cdot p_t\cdot N\cdot(N-1)}{2}\right)^2}$$

$$+ \frac{\left((p_w^2 - 3\cdot p_t^2)\cdot N^2 + (4\cdot p_t^2 - p_w^2)\cdot N + \frac{1}{6}\cdot p_w^2 - p_t^2\right)\cdot q + p_t^2\cdot N\cdot(N-1)^2}{\left(\frac{(2\cdot p_t - p_w)\cdot q^2 + (2\cdot p_t - p_w)\cdot(1-2\cdot N)\cdot q + 2\cdot p_t\cdot N\cdot(N-1)}{2}\right)^2}$$

For the case of a {3,1,0} scoring system, the former expression is reduced to:

$$HHI(q) = \frac{2\cdot q^3 + \left(-6\cdot N + \frac{5}{2}\right)\cdot q^2 + \left(6\cdot N^2 - 5\cdot N + \frac{1}{2}\right)\cdot q + N\cdot(N-1)^2}{\left(\frac{-q^2 - (1-2\cdot N)\cdot q + 2\cdot N\cdot(N-1)}{2}\right)^2}$$

■

Based on this, we wish to calculate the critical value (q^*) that forces the HHI reach its maximum value with a given N and scoring pattern. Because the HHI function is continuous and defined for $q \in [0, N-1]$ and $HHI(q) \in [{}^{1}/_{N}, 1]$ it reaches its maximum for each N. In addition, this maximum is unique.

For the purposes of demonstration, let us assume that it is not unique. If there are two different values, $q_1^* < q_2^*$, where the maximum of HHI is reached, then HHI(q_1^*) = HHI(q_2^*). The total points of the competition in q_1^* will be less than the corresponding total of points of q_2^* (this is because for each win in the pattern {3,1,0}, 1 point is added to the total). Therefore, the numerator and denominator of HHI increase with the number of points in a non-proportional way (the scoring system is fixed) and henceforth, two different values q_1^* and q_2^* cannot reach the same HHI value. Consequently, HHI for $q \in [0, N-1]$ has a unique maximum value and q^* represent the number of teams of the truncated cascade of winners. Therefore HHI increases from $HHI_{min}=1/N$ to HHI_{max}=HHI(q*), decreasing from HHI_{max} to 1.

Appendix B. Critical value q* to characterize the *truncated-cascade distribution* used to generate HHI_{max}.

From Appendix A:

$$\left[\sum_{i=1}^{N} p_i\right]_q = \frac{\left(2\cdot p_t - p_w\right)\cdot q^2 + \left(2\cdot p_t - p_w\right)\cdot(1-2\cdot N)\cdot q + 2\cdot p_t\cdot N\cdot(N-1)}{2}$$

$$\left(\left[\sum_{i=1}^{N} p_i\right]_q\right)^2 = \left(\frac{(2\cdot p_t - p_w)\cdot q^2 + (2\cdot p_t - p_w)\cdot(1-2\cdot N)\cdot q + 2\cdot p_t\cdot N\cdot(N-1)}{2}\right)^2$$

$$= \frac{1}{4}\{(2\cdot p_t - p_w)^2\cdot q^4 + 2\cdot(2\cdot p_t - p_w)^2\cdot(1-2\cdot N)\cdot q^3$$

$$+ \ [(2\cdot p_t - p_t)^2\cdot(1-2\cdot N)^2 + 4\cdot(2\cdot p_t - p_w)\cdot p_t\cdot N\cdot(N-1)]\cdot q^2 + 2$$

$$\cdot\,(2\cdot p_t - p_w)\cdot(1-2\cdot N)\cdot 2\cdot p_t\cdot N\cdot(N-1)\cdot q + 4\cdot {p_t}^2\cdot N^2\cdot(N-1)^2\}$$

$$\left[\sum_{i=1}^{N} p_i^2\right]_q = \left(\frac{p_w^2}{3} - p_t^2\right) \cdot q^3 + \left((3 \cdot p_t^2 - p_w^2) \cdot N + \frac{1}{2} p_w^2 - 2 \cdot p_t^2\right) \cdot q^2$$

$$+ \left((p_w^2 - 3 \cdot p_t^2) \cdot N^2 + (4 \cdot p_t^2 - p_w^2) \cdot N + \frac{1}{6} \cdot p_w^2 - p_t^2\right) \cdot q + p_t^2 \cdot N \cdot (N-1)^2$$

First step. We calculate, now, the expressions for $0 \le q \le N-1,\ with\, q \in \mathbb{N}$ of $[\sum_{i=1}^{N} p_i]_{q+1}$, $\left([\sum_{i=1}^{N} p_i]_{q+1}\right)^2$, and $[\sum_{i=1}^{N} p_i^2]_{q+1}$.

$$\left[\sum_{i=1}^{N} p_i\right]_{q+1} = \left[\sum_{i=1}^{N} p_i\right]_q + (P_w - 2 \cdot P_t) \cdot (N - q - 1)$$

$$\left(\left[\sum_{i=1}^{N} p_i\right]_{q+1}\right)^2 = \left(\left[\sum_{i=1}^{N} p_i\right]_q\right)^2 + 2 \cdot (P_w - 2 \cdot P_t) \cdot (N - q - 1) \cdot \left[\sum_{i=1}^{N} p_i\right]_q + (P_w - 2 \cdot P_t)^2 \cdot (N - q - 1)^2$$

$$\left[\sum_{i=1}^{N} p_i^2\right]_{q+1} = \left[\sum_{i=1}^{N} p_i^2\right]_q + \left[3 \cdot \left({P_w^2}/{3} - P_t^2\right)\right] \cdot q^2$$

$$+ \left[3 \cdot \left({P_w^2}/{3} - P_t^2\right) + 2 \cdot \left(P_w^2 \cdot \left(-N + {1}/{2}\right) + P_t^2 \cdot (3 \cdot N - 2)\right)\right] \cdot q$$

$$+ \left({P_w^2}/{3} - P_t^2\right) + \left(P_w^2 \cdot \left(-N + {1}/{2}\right) + P_t^2 \cdot (3 \cdot N - 2)\right)$$

$$+ \left(P_w^2 \cdot \left(N^2 - N + {1}/{6}\right) + P_t^2 \cdot (-3 \cdot N^2 + 4 \cdot N - 1)\right)$$

Second step. We particularize the above expressions for the {3,1,0} scoring pattern. As $(2 \cdot P_t - P_w) = -1$, we would have:

$$\left[\sum_{i=1}^{N} p_i\right]_{q+1} - \left[\sum_{i=1}^{N} p_i\right]_q = (N - q - 1)$$

$$\left(\left[\sum_{i=1}^{N} p_i\right]_{q+1}\right)^2 - \left(\left[\sum_{i=1}^{N} p_i\right]_q\right)^2 = q^3 - 3 \cdot (N-1) \cdot q^2 - 3 \cdot (N-1) \cdot q + (N-1)^2 \cdot (2 \cdot N + 1)$$

On the other hand, as:

$$\left\{\begin{matrix} \left({P_g^2}/{3} - P_e^2\right) = \left({9}/{3} - 1\right) = 2 \\ \left(P_w^2 \cdot \left(-N + {1}/{2}\right) + P_t^2 \cdot (3 \cdot N - 2)\right) = {9}/{2} - 9 \cdot N + 3 \cdot N - 2 = -6 \cdot N + {5}/{2} \\ \left(P_w^2 \cdot \left(N^2 - N + {1}/{6}\right) + P_t^2 \cdot (-3 \cdot N^2 + 4 \cdot N - 1)\right) = 9 \cdot N^2 - 9 \cdot N + {3}/{2} - 3 \cdot N^2 + 4 \cdot N - 1 = 6 \cdot N^2 - 5 \cdot N + {1}/{2} \end{matrix}\right\}$$

$$\left[\sum_{i=1}^{N} p_i^2\right]_{q+1} - \left[\sum_{i=1}^{N} p_i^2\right]_q = 6 \cdot q^2 + (11 - 12 \cdot N) \cdot q + 6 \cdot N^2 - 11 \cdot N + 5$$

Where:

$$\left[\sum_{i=1}^{N} p_i\right]_q = \frac{-q^2 - (1 - 2 \cdot N) \cdot q + 2 \cdot N \cdot (N - 1)}{2}$$

$$\left(\left[\sum_{i=1}^{N} p_i\right]_q\right)^2 = \frac{1}{4}\left\{q^4 + 2 \cdot (1 - 2 \cdot N) \cdot q^3 + q^2 - 4 \cdot (1 - 2 \cdot N) \cdot N \cdot (N - 1) \cdot q + 4 \cdot N^2 \cdot (N - 1)^2\right\}$$

$$\left[\sum_{i=1}^{N} p_i^2\right]_q = 2 \cdot q^3 + \left(-6 \cdot N + \frac{5}{2}\right) \cdot q^2 + \left(6 \cdot N^2 - 5 \cdot N + \frac{1}{2}\right) \cdot q + N \cdot (N - 1)^2$$

Third step. From this recurrence, we can obtain the maximum by identifying the ***q**** that verifies the inequality $[\sum_{i=1}^{N} s_i^2]_{q+1} < [\sum_{i=1}^{N} s_i^2]_q$. We are seeking the value of ***q*** that complies:

$$\left[\sum_{i=1}^{N} s_i^2\right]_{q+1} < \left[\sum_{i=1}^{N} s_i^2\right]_q \overset{eq.to}{\Longleftrightarrow} \frac{\frac{[\sum_{I=1}^{N} p_i^2]_{q+1}}{[\sum_{I=1}^{N} p_i^2]_q}}{\frac{\left[\left(\sum_{i=1}^{N} p_i\right)^2\right]_{q+1}}{\left[\left(\sum_{i=1}^{N} p_i\right)^2\right]_q}} < 1$$

By applying the expressions of recurrence we obtain:

$$1 > \frac{\frac{\left[\sum_{I=1}^{N} p_i^2\right]_{q+1}}{\left[\sum_{I=1}^{N} p_i^2\right]_q}}{\frac{\left[\left(\sum_{i=1}^{N} p_i\right)^2\right]_{q+1}}{\left[\left(\sum_{i=1}^{N} p_i\right)^2\right]_q}}$$

$$= \frac{\frac{\left[\sum_{I=1}^{N} p_i^2\right]_q + 6 \cdot q^2 + (11 - 12 \cdot N) \cdot q + 6 \cdot N^2 - 11 \cdot N + 5}{\left[\sum_{I=1}^{N} p_i^2\right]_q}}{\frac{\left[\left(\sum_{i=1}^{N} p_i\right)^2\right]_q + q^3 - 3 \cdot (N-1) \cdot q^2 - 3 \cdot (N-1) \cdot q + (N-1)^2 \cdot (2 \cdot N + 1)}{\left[\left(\sum_{i=1}^{N} p_i\right)^2\right]_q}} =$$

$$= \frac{1 + \frac{6 \cdot q^2 + (11 - 12 \cdot N) \cdot q + 6 \cdot N^2 - 11 \cdot N + 5}{\left[\sum_{I=1}^{N} p_i^2\right]_q}}{1 + \frac{q^3 - 3 \cdot (N-1) \cdot q^2 - 3 \cdot (N-1) \cdot q + (N-1)^2 \cdot (2 \cdot N + 1)}{\left[\left(\sum_{i=1}^{N} p_i\right)^2\right]_q}}$$

which are equivalent to:

$$\frac{q^3 - 3 \cdot (N-1) \cdot q^2 - 3 \cdot (N-1) \cdot q + (N-1)^2 \cdot (2 \cdot N + 1)}{\left[\left(\sum_{i=1}^{N} p_i\right)^2\right]_q} > \frac{6 \cdot q^2 + (11 - 12 \cdot N) \cdot q + 6 \cdot N^2 - 11 \cdot N + 5}{\left[\sum_{I=1}^{N} p_i^2\right]_q}$$

Fourth step. We redefine inequality, so that the critical value of $\boldsymbol{q}$ is obtained when:

$$\frac{\left(q - (N-1)\right) \cdot \left[q^2 - 2 \cdot (N-1) \cdot q - (N-1) \cdot (2 \cdot N + 1)\right]}{\frac{1}{4}\{q^4 + 2 \cdot (1 - 2 \cdot N) \cdot q^3 + q^2 - 4 \cdot (1 - 2 \cdot N) \cdot N \cdot (N-1) \cdot q + 4 \cdot N^2 \cdot (N-1)^2\}}$$

$$> \frac{\left(q - (N-1)\right) \cdot [6 \cdot q - 6 \cdot N + 5]}{2 \cdot q^3 + \left(-6 \cdot N + \frac{5}{2}\right) \cdot q^2 + \left(6 \cdot N^2 - 5 \cdot N + \frac{1}{2}\right) \cdot q + N \cdot (N^2 - 2 \cdot N + 1)}$$

Simplifying:

$$\big(q-(N-1)\big)\cdot\Big[2\cdot q^5+\Big(-10\cdot N+\frac{13}{2}\Big)q^4+\Big(14\cdot N^2-20\cdot N+\frac{15}{2}\Big)q^3$$

$$+\Big(N^3+9\cdot N^2-\frac{27}{2}\cdot N+\frac{7}{2}\Big)q^2$$

$$+\Big(-14\cdot N^4+22\cdot N^3-6\cdot N^2-\frac{5}{2}\cdot N+\frac{1}{2}\Big)\cdot q-N$$

$$\cdot(2\cdot N^4-5\cdot N^3+3\cdot N^2+N-1)\Big]$$

$$>\frac{\big(q-(N-1)\big)}{4}[6\cdot q^5+(-30\cdot N+17)q^4$$

$$+(24\cdot N^2-32\cdot N+16)q^3+(48\cdot N^3-72\cdot N^2+18\cdot N+5)q^2+4$$

$$\cdot(-6\cdot N^4+16\cdot N^3-15\cdot N^2+5\cdot N)\cdot q+N^2$$

$$\cdot(-6\cdot N^3+17\cdot N^2-16\cdot N+5)]$$

By operating inequality, the resulting polynomial would be:

$$0>\Big[-\frac{1}{2}\cdot q^5+\Big(\frac{5}{2}\cdot N-\frac{9}{4}\Big)\cdot q^4+\Big(-8\cdot N^2+12\cdot N-\frac{7}{2}\Big)\cdot q^3$$

$$+\Big(11\cdot N^3-27\cdot N^2+18\cdot N-\frac{9}{4}\Big)\cdot q^2$$

$$+\Big(8\cdot N^4-6\cdot N^3-9\cdot N^2+\frac{15}{2}\cdot N-\frac{1}{2}\Big)\cdot q-4\cdot N^5+12\cdot N^4-13\cdot N^3+6$$

$$\cdot N^2-N\Big]\ \blacksquare$$