Overdog Team, Underdog Team and the Philosophy of Soccer Team Ranking

Sébastien Courtin*
Frédéric Chantreuil†
Olivier Dagnelie‡

The Ramanujacharyula method, applied in the field of sports, allows to determine the rank of a team accounting both for the strength of the teams it beats, as well as the strength of the teams that beat it. We analyze the changes in the final rank positions of five soccer championships (England, France, Germany, Italy and Spain), induced by the use of the Ramanujacharyula scoring method over the 1946-2016 period, and determine its quantitative impact. The Kemeny measure is then applied to establish a real-world ranking method that would best reflect the philosophy promoted by the Ramanujacharyula method. The scoring rule 2 – 1 – 0 (2 points for a win, 1 point for a draw and 0 for a loss) is found to be the best ranking method in this sense.

Ramanujacharyula method – Scoring voting methods – Soccer leagues – Tournaments

Favoris, Outsiders, et la philosophie du classement des équipes de football

La méthode de Ramanujacharyula [1964], appliquée dans le domaine sportif, permet de déterminer le rang d'une équipe en tenant compte à la fois de la force des équipes qu'elle bat, ainsi que de la force des équipes qui l'ont battue. Nous analysons les changements dans les classements finals de cinq championnats de football (Angleterre, France, Allemagne, Italie et Espagne), induits par l'utilisation de la méthode de classement de Ramanujacharyula au cours de la période 1946-2016. La mesure de Kemeny est ensuite appliquée pour établir une méthode de classement qui refléterait au mieux la philosophie promue par la méthode Ramanujacharyula. Nous montrons que la règle de classement (2 points pour une victoire, 1 point pour un match nul et 0 pour une défaite) est ce que nous considérons comme la meilleure méthode.

Méthode de Ramanujacharyula – Méthodes de votes à score – Championnats de football – Tournois

JEL Classification Numbers : Z20 ; D71

* Normandie Univ, Unicaen, IDEES, UMR CNRS 6266, France. Email: sebastien.courtin@unicaen.fr.
† University of La Reunion, CEMOI, La Reunion, France. Email: frederic.chantreuil@univ-reunion.fr.
‡ UPPA, TREE, UMR CNRS 6031, France. Email: olivier.dagnelie@univ-pau.fr.
The authors have no relevant financial or non-financial interests to disclose, whether directly or indirectly related to the work submitted. Thus, there are no competing interests to declare that are relevant to the content of this article.
The authors would like to thank an anonymous referee for their useful comments. This work also benefited from comments from participants at several conferences and seminars.

1. Introduction

"We were unlucky to lose, we should have won! Better luck next time!" What sports fan never uttered such words after the defeat of their favorite soccer team? One could argue that the beauty of soccer lies in the premise that the best team on the field doesn't always necessarily win the game. Others may defend the thesis that soccer is a beautiful game of chance where we might "see an underdog team triumph with a miraculous rebound or an undeserved penalty kick" (Tierney [2014]). On the other hand, "because of fluke goals, low scores and the many matches that end in ties, soccer is less predictable than other major sports" (Tierney [2014]). This anecdotal piece of evidence is confirmed by Scarf *et al.* [2019] who prove that in a Poisson match (a widely used model to simulate match outcomes in soccer) more scoring implies less outcome uncertainty. Despite the uncertainty of sport, in the end, there is always a winner – a Champion. Does this mean that the Champion is simply luckier than the other contestants, or is the Champion really the best?

If we look at soccer bets, Leicester (Champion of 2015/2016) beat 5000-1 odds to win the English Premier League title. A stroke of luck? During the 2015/2016 French soccer championship, Paris Saint Germain (the Champion at the time) failed twice to beat Montpellier (1 draw and 1 defeat). Could one argue that Montpellier played a better game than Paris Saint Germain that year? Or did their team get lucky twice in a season? It is, not surprisingly, possible to find similar situations all over Europe, cases in which the "worst" or the "unlucky" team in bilateral confrontations ends up winning the championship.

To account for this kind of "paradox", one must analyze a team's ability to beat underdog teams or its ability to beat overdog teams. If the Champion has demonstrated, throughout the season, a high ability to defeat underdog teams, all the while being unable to win against overdog teams, it could be concluded that the Champion is just lucky or not all that strong. Symmetrically, if a team engaged in a fierce battle against relegation has a high capacity to beat overdog teams during the season, while at the same time having the incapacity to beat underdog teams, one could conclude that this team is unlucky though not particularly weak. In the current form of soccer leagues, the strength of the opponent teams, *i.e.* their performance against other teams (strong or weak), does not matter. The only thing that matters is the result of direct confrontations. We think that the legitimacy of a champion (or of a team saved from relegation) would be reinforced if the *strength* and *weakness* of opponents were taken into account.

By modelling soccer championships, by means of a round robin tournament, a framework accounting for the previously mentioned paradoxes can be found. A tournament is a collection of games, in which one can observe the results of a team going up against another team.

Consider the following graph representing the results of a tournament comprising five soccer teams. Note that each node is a team and an arc going from node i to node j means that player i has won a game against player j.

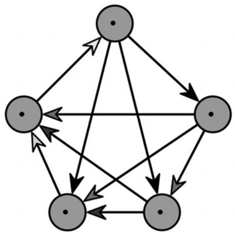
The win-loss outcomes of these matches can be conveniently displayed in a tournament matrix (adjacency matrix), by setting its i, j component to one when team i beats team j, and zero otherwise.[1]

$$A = \begin{Bmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{Bmatrix}$$

The adjacency matrix will be useful in assessing the final result of the tournament. Indeed, one of the many questions that surround the study of tournaments is how to rank teams relatively to one another, *i.e* by quantifying the strength of each team. There exists a wide variety of social choice methods used to aggregate individual rankings into a final ranking.

The most natural method would be the score method. The score method consists of comparing the number of games won by each player. In an adjacency matrix, it follows that the number of games won by player i is just the sum of the entries in the ith row of the matrix. Another well-known method is the Kendall–Wei ($K-W$) method introduced by Kendall [1955] and Wei [1952] and also referred to as the long-path method in the context of a tournament (Laslier [1997]). They extend the score method by attempting to account not only for a team's head-to-head wins but also for the strength of the defeated teams. The idea is to better reward a win against an overdog team who has earned many victories, than a victory against an underdog team defeated by everyone. To find the $K-W$ scores, we begin by giving each team a score equal to the number of games won by each player. We then give each team a second score equal to the number of wins earned by the teams it defeated. In the third iteration, each team's score equals the sum of the second round scores of the teams it defeated. The process continues and, at each stage, the team's score is equal to the sum of scores of the teams it defeated in the previous round. This results in each team having an infinite sequence of scores. As it happens, these sequences converge to what we call the $K-W$ score, from which we derive the final ranking.

Figure 1. graph example



1. Note that draws are allowed as shown in Section 2.

The previously presented method can be considered as asymmetric. Indeed, in tournaments, it might be interesting to identify the really talented team, that is to say, the one that beat the largest number of opponents but was simultaneously defeated by only a few opponents. In other words, we must account for the strength of each opponent a team defeated, but also for the strength of the opponent that defeated it (its weakness). In the previous example, team 1 and team 2 each beat three of their opponents and ranked first according to the score method. They both beat team 3 and team 4 and it is the third victory that will be decisive for the final ranking according to the $K-W$ method. Team 1 beat team 2, an overdog team (due to its three wins) while team 2 beat team 5, an underdog team with only one win. Thus, team 1 ranked ahead of team 2 on account of its third victory against a stronger opponent. However, this overlooks the fact that team 2's only defeat was against an overdog team (with three wins), while team 1 lost against a team beaten by all the other teams. Ramanujacharyula [1964] introduces a measure which indicates the level of weakness for each individual and is used with the measure of strength. According to this method (the RA method), a team i is stronger than a team j if the ratio strengthweakness is greater in team i than in team j. In the previous example, team 2 has the highest ratio and is ranked ahead of the other teams. Another method which takes strength and weakness into account is the one put forward by Hasse [1961]. Hasse and RA methods are close to each other but may produce different rankings, since the first one registers the absolute differences between the strength and weakness vectors whereas the second is based on the relative differences.

The RA method is the one considered in this paper since we think that relative differences best reflect the concept of opponents' strength and weakness (*i.e.* their performance against other teams).[2] Our aim is double, the first one being quantitative. Using data on five soccer leagues (England, France, Germany, Italy and Spain)[3] we want to quantify the major changes in these different leagues by applying the RA method. The second aim is qualitative: what are the best scoring methods for a soccer tournament? Are there any specificities related to the leagues studied? Indeed, although the RA method is very appealing, it will hardly be applicable in real life. Nevertheless, we will show that a scoring method accurately reflecting the philosophy of the RA method exists.

This work is structured as follows: after a brief review of the literature on tournament ranking methods in Section 2, Section 3 introduces the general framework. Section 4 presents the five different soccer leagues. Section 5 and 6 contain the main results of our work, while Section 7 concludes the paper.

2. We acknowledge that other strength / weakness methods could be further investigated, since "there are many possible modifications of such procedures" (*i.e.* Hasse and RA) as suggested by Chebotarev and Shamis [1998].

3. Note that simulations can also be used to analyse tournament rules, see Devriesere *et al.* [2024].

2. Related literature

There is an extensive literature in social choice on the selection of the best alternative in tournaments, see Moon [1968] and Laslier [1997] for a general survey, and Devriesere *et al.* [2024] for a recent survey with many details about ranking in sports. In soccer league, the goal is not only to select the best team but also to obtain a complete ranking of the teams. Note that tie-breaking rules must sometimes be used in order to have a complete ranking in a tournament. For example, Csató [2023] shows that tie-breaking rules would have non negligible sporting effect in practice. Thus, an advantage of methods like RA, is that they rarely require further tie breaking rules, compared to scoring rules. In order to achieve a general ranking, several methods have been proposed in the literature. Herings *et al.* [2005] distinguish between local and global methods. The score of a node in global methods depends on the score of the other nodes, which is not the case in local methods. A local method uses only partial information about the structure of the graph, whereas global methods use information pertaining to the entire structure of the graph. The score method is the most famous example of a local method (see among others Behzad *et al.* [1979], Rubinstein [1980], Henriet [1985], Bouyssou [1992], Chebotarev and Shamis [1998], van den Brink and Gilles [2003] and González-Díaz *et al.* [2014]). A global method iterates the score vector, which is inspired by the $K-W$ method. The properties of these methods have been studied, among others, by Keener [1993], Eschenbach *et al.* [2000], and Kitti [2016]. The most obvious application of $K-W$ involves the ranking of sports teams, but $K-W$ has also been used in other areas. For example, Google uses a modification of the $K-W$ ratings in order to determine the order of results in response to a website search query (the PageRank method), see Brin and Page [1998], and Altman and Tennenholtz [2005]. Slutzki and Volij [2006] axiomatically characterized two methods (the fair-bets and the invariant scoring methods) derived from the $K-W$ method. These are variants that are used to rank webpages by their relevance to a query and academic journals according to their impact. Another global method is the well-known Markov solution (see Daniels [1969]). Herings *et al.* [2005] define yet another method: the positional power function in which the score of a node is determined by both the number of its successors and the score of these successors. One of the advantages of this method is that it is not restricted to a subclass of digraphs as is the case of the $K-W$ method.

It is worth noting that outside of the tournament context, an extensive literature has been devoted to the study of rules used in professional sports from a social choice perspective. For example, Truchon [1998] and Saari [2001] analyze the rule used in international figure skating competitions to aggregate the marks allocated to skaters by individual judges in order to determine a final ranking. Hammond [2007] demonstrates that the scoring method used in high school and college cross-country meets in the U.S. has some undesirable social choice properties. Moschini [2010] came up with a game-theoretic model to capture the impact of the 3-points rule for a win in soccer. The implementation of his model into datasets spanning 30 years for 35 countries shows that this system has increased the expected number of goals and decreased the fraction of drawn matches. Nevertheless, the 3-points system has led to a decrease of the number of goals in Germany, Austria and the Czech Republic, and an increase in the number of drawn matches in Italy. Moreover Guedes and Machado [2002] show that a higher reward for victory creates an incentive for the weaker team to play more defensively if the difference of strengths of the two teams is sufficiently high. In a recent paper, Kondratev *et al.* [2023] discuss the problem of choosing a scoring rule in sports from a social choice perspective.

There is also a literature that statistically compares the predictive power of rankings. A comprehensive survey of this literature is beyond the scope of this work but see, for example, Barrow *et al.* [2013], or, more recently, Szczecinski [2023], who estimates with a probabilistic model that the value of a win in soccer should be close to five points.

3. Notations and definitions

Round robin tournament

Let $N = \{1, ..., n\}$ a set of nodes (teams) and let (i, j) be an arc which points from some node $i \in N$ to some node $j \in N$. The set of directed arcs is denoted by G. A directed graph on N is defined by its set G of directed arcs, *i.e.* $G \subseteq N \times N$. If $(i, j) \in G$ we say that node i dominates node j (team i beats team j). The set of all graphs on N is denoted by $\mathcal{G}$. A graph G is said to be a round robin tournament if for any two different nodes $i, j \in N$ it holds that either $(i, j) \in G$ or $(j, i) \in G$.

A general class of round robin tournaments consists of weighted graphs. A weighted graph is given by a set of nodes N and a nonnegative $n \times n$ matrix W, the weighted adjacency matrix, where the (i, j)th element w_{ij} of W denotes the weight of the arc from node i to node j, while w_{ii} is a weight associated with an arc from a node to itself. The value w_{ij} can be seen as the number of points obtained by team i if it beats team j. In the following, for all $i, j \in N$,

$$w_{ij} = \begin{cases} 1 \text{ if } i \text{ beats } j \\ \sigma \text{ if } i \text{ ties } j \text{ with } \sigma \in \{0, 1\} \\ 0 \text{ otherwise} \end{cases}$$

Notice that $w_{ii} = 0$.

In most of the current soccer tournaments, a team earns 3 points for a win, 1 for a draw, and 0 for a loss. These tournaments can be modelled with

$$w_{ij} = \begin{cases} 1 \text{ if } i \text{ beats } j \\ \frac{1}{3} \text{ if } i \text{ ties } j \\ 0 \text{ otherwise} \end{cases}$$

In the following, $G(W)$ denotes the directed graph corresponding to W, $\mathcal{W}$ will be the set of all weight matrices and $\mathcal{G}(W)$ the set of all associated graphs.

Ranking methods

In a sports competition, a ranking method $F = (F_1, ..., F_n)$ on the collection $\mathcal{W}$ of weight matrices graphs on N assigns for any $W \in \mathcal{W}$ a real number to every team $i \in N$ which can be seen as its power. In principle, the power of a node can be any real number, positive or negative. Since the powers are invariant to shifting by a constant if they are used only for ranking, the powers can be normalised.

Score method

The simplest way to measure the power of a player i is to sum up the entries in the ith row of the matrix W. More formally, the score method is the function F^S: $\mathcal{W} \longmapsto IR^n_+$ given by $F^S_i(W) = S_i$, $i \in N$, $W \in \mathcal{W}$, with $S_i = \sum_{j=1}^{n} [W]_{ij}$. The row sum vector is then given by $S(W) = (S_1(W), S_2(W), ..., S_n(W))$, $S_i(W)$ being the number of points earned by each team.

$K - W$ method

For a given graph, the $K - W$ method considers the sequence $p^t = Wp^{t-1}$, $t = 1, 2, ...$, starting with p_0 equal to **1**, the $n -$ vector of ones. By definition of W, the vector p^1 is the row sum vector S, p^2 is the vector that assigns to any team i the sum of the scores that team i beats and so on. We can observe that if $\lim_{t \to \infty} \frac{p_t}{\sum_i p_t^i}$ is well defined, then the process converges to what we call the strength vector, $P(W) = (P_1(W), P_2(W), ..., P_n(W))$.

In order to converge to a nonzero solution, one needs to restrict the domain to irreducible matrices. Formally, matrix W is irreducible if $G(W)$ is strongly connected: for each pair of nodes i, j there exists a sequence $i_1, i_2, ..., i_k$ such that $i_1 = i$, $i_k = j$ and $(i_h, i_{h+1}) \in G$ for all $h = 1, ..., k-1$. In the context of sport, this condition holds if all the teams play against all the other ones, and each team has at least one tie (or win). Moon and Pullman [1970] show that if W is irreducible, the $K-W$ function converges to the unique, strictly positive eigenvector (up to normalization) of W. This eigenvector corresponds with the highest positive eigenvalue $\lambda(W)$ of W. Therefore, $P(W)$ is the solution to the homogeneous system of linear equations $\lambda(W)x = Wx$.

RA method

Alongside the strength vector, Ramanujacharyula [1964] introduces a vector which indicates the level of weakness for each team. To determine this vector, one needs to apply the $K-W$ function to W^T.[4] More formally, for a given graph W^T, Ramanujacharyula first considers the sequence $q^t = W^T q^{t-1}$, $t = 1, 2, ...$, starting with q_0 equal to **1**. In this case, q^2 can be seen as the vector that assigns to any team i the sum of the scores obtained by every team j that beats it. The weakness vector is the result of $\lim_{t\to\infty} \frac{q_t}{\sum_i q_t^i}$. In other words, one needs to find the corresponding eigenvector of W^T, denoted by $Q(W) = (Q_1(W), Q_2(W), ..., Q_n(W))$.

From the strength vector and the weakness vector, one can derive the *RA* vector $R(W) = [R_1(W), ..., R_n(W)]$ with $R_i(W) = P_i(W)/Q_i(W)$ for all i. Simply put, a team i is stronger than a team j if the ratio "strength-weakness" is greater in team i than in team j.

Example

Consider the five teams tournament introduced in Figure 1 with

$$A = \begin{Bmatrix} 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \end{Bmatrix} \text{ and } A^T = \begin{Bmatrix} 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 \end{Bmatrix}$$

4. where T denotes the transposed matrix.

Score method: Note that ties are broken by lexicographical order.

$$S(A)=\begin{Bmatrix}3\\3\\2\\1\\1\end{Bmatrix}\mapsto\begin{Bmatrix}Team1\\Team2\\Team3\\Team4\\Team5\end{Bmatrix}$$

$K-W$ method:

$$p^1(A)=\begin{Bmatrix}3\\3\\2\\1\\1\end{Bmatrix},p^2(A)=\begin{Bmatrix}6\\4\\2\\1\\3\end{Bmatrix},p^3(A)=\begin{Bmatrix}7\\6\\4\\3\\6\end{Bmatrix}\dots p(A)\mapsto\begin{Bmatrix}Team1\\Team2\\Team5\\Team3\\Team4\end{Bmatrix}$$

RA method:

$$q^1(A^T)=\begin{Bmatrix}1\\1\\2\\3\\3\end{Bmatrix},q^2(A^T)=\begin{Bmatrix}3\\1\\2\\4\\6\end{Bmatrix},q^3(A^T)=\begin{Bmatrix}6\\3\\4\\6\\7\end{Bmatrix},\dots q(A^T)\mapsto\frac{p(A)}{q(A^T)}\mapsto\begin{Bmatrix}Team2\\Team1\\Team3\\Team5\\Team4\end{Bmatrix}$$

With three victory, Team 1 and Team 2 are tied (the tie is broken by lexicographical order) according to the score method. They have each beaten Team 3 and Team 4. The third victory is decisive for the final ranking according to $K-W$. Team 1 beats Team 2, while Team 2 beats Team 5. Therefore, Team 1 is ranked ahead of Teams 2 since its third victory is against a stronger opponent. Hovewer $K-W$ ignore the fact that the only defeat of Team 2 is against Team 1, while Team 1 is beaten by Team 5. This implies that Team 2 is ahead of Team 1 according to RA.

According to $K-W$, Team 5 is ranked ahead of Team 3 and Team 4 (but below according to the score method) since Team 5 beats Team 1, the strongest team. Team 3 is ahead of Team 5 according to RA, since Team 3 is beaten only by the two strongest teams, while Team 5 is beaten by the three strongest teams.

4. Official method versus *RA* method

4.1. Soccer Leagues

The UEFA country coefficients take into account the results of all clubs in each association and are used to determine the number of entries an association will be granted in the forthcoming season of the UEFA club competitions. According to these rankings, the top five soccer leagues are Spain, England, Germany, Italy and France. We focus our attention on these five soccer leagues. Table 1 summarizes, for each league, the ranking rules $(1, \frac{1}{2}, 0)$ or $\left(1, \frac{1}{3}, 0\right)$ and the number of teams is given in brackets. Most of the soccer leagues started before 1946, however, due to a one-year gap on account of"WorldWar 2", our study begins with 1946, with the exception of Germany where the national league emerged in 1963. The data used in this article was assembled by compiling information reported by "www.worldfootball.net" and is accessible online.[5]

4.2. Descriptive statistics

For each league, we first identify and report the major changes that take place when using the *RA* method instead of the scoring method. It should be emphasised that modifying the ranking method could have changed the behaviour of the teams, which is not considered in our theoretical analysis.

Table 1. *Soccer League*

League	$\left(1, \frac{1}{2}, 0\right)$	$\left(1, \frac{1}{3}, 0\right)$
England	1946 – 1980(22)	1981 – 1986(22) 1987(21) 1988 – 1990(20) 1991 – 1994(22) ≥ 1995(20)
France	1946(18) 1947(20) 1948 – 1958(18) 1959 – 1963(20) 1964 – 1965(18) 1966 – 1968(20) 1969 – 1971(18) 1972 – 1988(20) 1989 – 1994(20)	1988 – 1989(20) 1994 – 1997(20) 1998 – 2001(18) ≥ 2002(20)
Germany	1963 – 1964(16) 1965 – 1989(18)	1990(18) 1991(20) ≥ 1992(18)
Italy	1946(20) 1947(21) 1948 – 1951(20) 1952 – 1966(18) 1967 – 1987(16) 1988 – 1993(18)	1994 – 2004(18) ≥ 2004(20)
Spain	1946 – 1949(14) 1950 – 1970(16) 1971 – 1989(18) 1990 – 1994(20)	1995(22) 1996(23) ≥ 1997(20)

5. http://www.worldfootball.net/, the computational work undertaken in this research work was coded in Stata and Mata, and the corresponding programs are available from the authors upon request.

The most important changes concern the final Champion, the teams qualified for the European Leagues and the relegated teams. Let us begin with some examples of theses modifications.

4.2.1. France

During the 1997/1998 French League, RC Lens (68 points; 21,5,8) was the Champion with 21 wins, 5 draws and 8 losses. In second place came FC Metz (68 points; 20,8,6) with 20 wins, 8 draws and 6 losses. With the RA method it appears that FC Metz should have been the Champion.

In the 2003/2004 French League, the final standings were as follows: 1) Olympique Lyonnais (79; 24,7,7); 2) Paris Saint-Germain (76; 22,10,6); and 3) AS Monaco (75; 21,12,5). Using RA, the ranking would have been: 1) AS Monaco; 2) Olympique Lyonnais; and 3) Paris Saint-Germain.

One last French example concerns the Drop Zone. In 1999/2000, the battle for the last team relegated was intense since five teams were within one point of one another for the last position in the Drop Zone. Eventually, AS Nancy Lorraine (42; 11,9,14) was relegated due to scoring two goals less than Olympique de Marseille (42; 9,15,10). However, with the RA method, ESTAC Troyes (43; 13,4,17) would have been relegated.

4.2.2. Spain

Is Real Madrid not happy with the current ranking method? In 2013/2014, Atletico Madrid was the Champion (90; 28,6,4), FC Barcelona was second (90; 27,6,5) and Real Madrid came third (90; 27,6,5) because it had a lower "goal difference". Real Madrid ranked second in 2014/2015 (92; 30,3,5) close behind FC Barcelona (94; 30,4,4). The same scenario occurred in 2015/2016 with Real Madrid (90; 28,6,4) versus FC Barcelona (91; 29,4,5). Real Madrid would have been Champion three years in a row with the RA method.

Table 2 presents the final ranking of 2011/2012 season whereas Table 3 indicates the final ranking with RA.

The final Spanish example concerns the 2015/2016 season where the Rayo Vallecano (38; 9,11,18) was relegated whereas the Sporting Gigòn (39; 10,9,19) and Granada CF (39; 10,9,19) were outside the Drop Zone. Using RA, the Sporting Gigòn would have been in the Drop Zone.

Table 2. *La Liga 2011/2012*

4th	Malaga FC (58; 17,7,14)	Play-off Round Champions League
5th	Atletico Madrid (56; 15,11,12)	Group Stage C3
6th	Levante UD (56; 16,7,15)	Play-off Round C3
7th	Osasuna Pamplona (54; 13,15,10)	Nothing

Table 3. *Liga 2011/2012*

4th	Osasuna Pamplelune	Play-off round Champions League
5th	Athletico Madrid	Group Stage C3
6th	Malaga FC	Play-off round C3
7th	Levante UD	Nothing

Table 4. *Premier League 2005/2006*

1th	Chelsea FC (91; 29,4,5)	Group Stage Champions League
2th	Manchester United (83; 25,8,5)	Group Stage Champions League
3th	Liverpool FC (82; 25,7,6)	Play-off Round Champions League
4th	Arsenal FC (67; 20,7,11)	Play-off Round Champions League
5th	Tottenham Hotspur (65; 18,11,9)	First Round UEFA Cup

Table 5. *Premier League 2005/2006*

1th	Chelsea FC	Group Stage Champions League
2th	Liverpool FC	Group Stage Champions League
3th	Manchester United	Play-off Round Champions League
4th	Tottenham Hotspur	Play-off Round Champions League
5th	Arsenal FC	First Round UEFA Cup

4.2.3. England

Our first example concerns the 2005/2006 season, presenting the official ranking in Table 4 and the *RA* method in Table 5.

One can observe that, with the *RA* method, Manchester United swapped places with its historical opponent – Liverpool FC. Concurrently, Arsenal FC would have ranked behind Tottenham Hotspur, its “best opponent”.

Table 6. *Bundesliga 2000/2001*

1th	Bayern München (63; 19,6,5)	Group Stage Champions League
2th	FC Schalke 04 (62; 18,8,8)	Group Stage Champions League
3th	Borussia Dortmund (58; 16,10,8)	Play-off Round Champions League
4th	Bayer Leverkusen (57; 17,6,11)	Play-off Round Champions League
5th	Hertha BSC (56; 18,2,14)	First Round UEFA Cup
6th	SC Freiburg (55; 15,10,9)	First Round UEFA Cup

Table 7. *Bundesliga 2000/2001*

1th	FC Schalke 04	Group Stage Champions League
2th	Bayern München	Group Stage Champions League
3th	Borussia Dortmund	Play-off Round Champions League
4th	SC Freiburg	Play-off Round Champions League
5th	Bayer Leverkusen	First Round UEFA Cup
6th	Hertha BSC	First Round UEFA Cup

In 2007/2008, Manchester United (87; 27,6,5) was the Champion ahead of Chelsea (85; 25,10,3) and Arsenal FC (83; 24,11,3). With the RA method, Manchester United would have only ranked second, while Arsenal FC would have been the Champion. In 2010/2011, Wolverhampton Wanderers (40; 11,7,20) saved its position in the Premier League against Birmingham City (39; 8,15,15) and Blackpool FC (39; 10,9,19). This is due to the ranking method since applying RA would have saved Birmingham City's position.

4.2.4. Germany

The 2000/2001 Bundesliga is represented by the official ranking displayed in Table 6, while Table 7 provides the ranking with the RA method.

With the RA method, two teams improve their rank order: Schalke 04 would have been the Champion and SC Freiburg would have qualified for the Play-off Round of the Champions League.

In 2014/2015, Hamburger SV (35; 9,8,17) went into a play-off match and saved its position in the First League. However, with the RA method, Hamburger SV would have ranked 17th and been directly relegated. One can also note that Freiburg (34; 7,13,14), in 19th place in the official ranking, would not have been relegated when using the RA method.

Table 8. *Serie A 1995/1996*

3th	Lazio Roma (59; 17, 8, 9,+28)
4th	ACF Fiorentina (59; 17, 8, 9, +12)
5th	AS Roma (58; 16, 10, 8 51, +17)
6th	Parma AC (58; 16, 10, 8, +13)

Table 9. *Serie A 1995/1996*

3th	ACF Fiorentina
4th	Parma AC
5th	Lazio Roma
6th	AS Roma

4.2.5. Italy

The 2000/2001 Serie A saw many teams fight their way out of the Drop Zone: 6 clubs (Udinese Calcio, US Lecce, Hellas Verona, SSC Napoli, Reggina Calcio and Vicenza Calcio) found themselves in a battle to survive. In the end, Udinese Calcio was preserved, among 18 teams, while the *RA* method would have directly relegated Udinese Calcio.

Tables 8 and 9 summarize the 1995/1996 Serie A situation under the official ranking and the *RA* method.

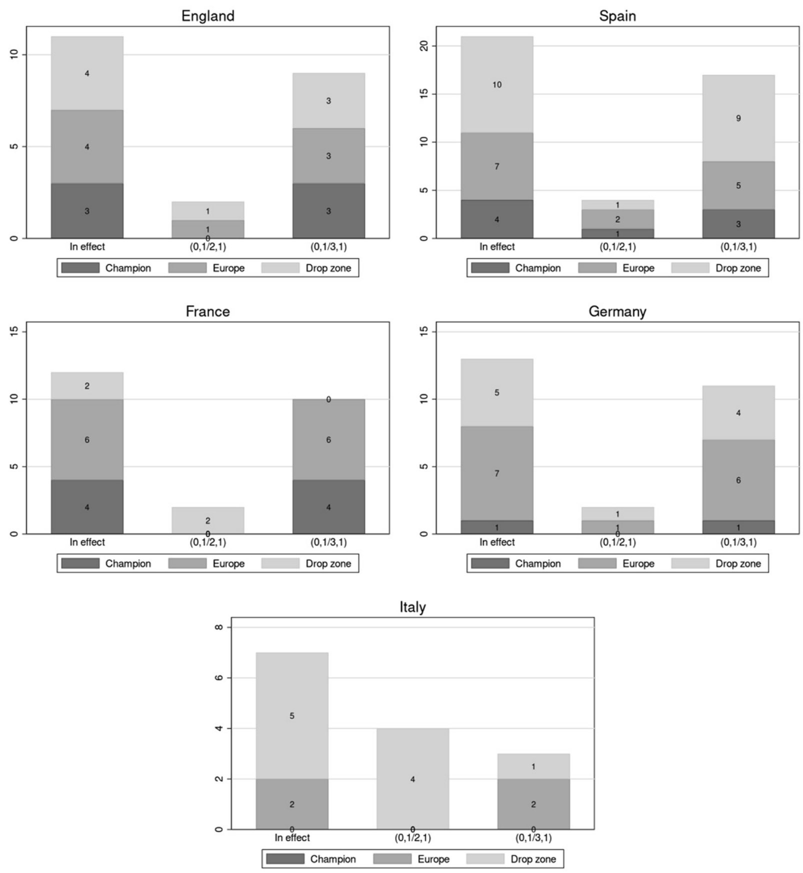
These tables show that the teams that qualified for the European Leagues, during the 1995/1996 Serie A, would have been rather different under the *RA* method.

In 1963/1964, a play-off match took place between Bologna FC and Internazionale to determine the Champion for the first and only time in Serie A. Bologna FC won the game but Internazionale would have been the Champion with *RA* without a play-off.

To conclude this subsection, let us present Figure 2 which indicates, for each league, the total number of changes[6] in the most important ranking zones (Champion, European Zone, Drop Zone).

6. We count 1 when a team is not in the zone with the official method but is in the zone under the *RA* method. Note that cases where *RA* only breaks a tie in this zone are not taken into account.

Figure 2. Changes in ranking



The first column is the total number of changes whatever the ranking rule, the second and the third columns concern the change for the $\left(1, \frac{1}{2}, 0\right)$ rule and the $\left(1, \frac{1}{3}, 0\right)$ rule. There are 69 seasons for all leagues except for Germany with 52 seasons.

It appears that the changes that occurred in the important zones are more significant during the $\left(1, \frac{1}{3}, 0\right)$ period than the $\left(1, \frac{1}{2}, 0\right)$ period, suggesting that RA seems closest to $\left(1, \frac{1}{2}, 0\right)$. One explanation could be that RA considers a draw as a half win. Therefore two draws for a team are equivalent to a win and a loss. For example, a team with six draws over six games in a given tournament will be seen by RA like 3 wins and 3 losses, *i.e.* neither as an overdog team, nor as an underdog team. Therefore, its ranking (relative to the other teams) will be a middle one, which will also be the case with the $\left(1, \frac{1}{2}, 0\right)$ scoring method.

4.3. Differences between rankings

In this subsection, the discrepancies between the official ranking and the one issued from the RA method are analyzed by implementing a measure used by Truchon [1998], which is based on the Kemeny measure (Kemeny [1959]).[7]

More specifically, given the two rankings, r and $r^* \in IR$ and two teams i and $j \in N$ let us define:

$$\Phi_{ij}(r, r^*) = \begin{cases} 1 \text{ if } r_i < r_j \text{ and } r_j^* < r_i^* \\ \frac{1}{2} \text{ if } r_i = r_j \text{ and } r_j^* < r_i^* \\ \frac{1}{2} \text{ if } r_i < r_j \text{ and } r_j^* = r_i^* \\ 0 \text{ otherwise} \end{cases}$$

The measure of the difference between r and r^* is given by:

$$G(r, r^*) = \sum_{i \in N} \sum_{j \in N} \Phi_{ij}(r, r^*)$$

Consider two rankings made of four teams: $r = \begin{Bmatrix} a \\ b \\ c \\ d \end{Bmatrix}$ and $r^* = \begin{Bmatrix} c \\ b \\ a \\ d \end{Bmatrix}$. The paired differences will be the following: $\Phi_{ab}(r, r^*) = 1$; $\Phi_{ac}(r, r^*) = 1$; $\Phi_{ad}(r, r^*) = 0$; $\Phi_{bc}(r, r^*) = 1$; $\Phi_{bd}(r, r^*) = 0$; and $\Phi_{cd}(r, r^*) = 0$. Therefore, the difference between r and r^* is given by $G(r, r^*) = 3$. The higher $G(r, r^*)$, the greater the difference between the two rankings. For each league, we present now the discrepancies between the official rules ($w_{ij} = \left(1, \frac{1}{2}, 0\right)$ or $\left(w_{ij} = 1, \frac{1}{3}, 0\right)$) and the RA rules according to the Kemeny measure. Our aim is to compare the official rules to the RA rules using the same vector of points. Since the number of teams in each league can be different, we do not compare Kemeny's scores between the leagues.

7. There exist other measures that can be used to evaluate the difference between two rankings. One example is the weighted measure introduced by Can [2014], in which a difference at the top of the ranking has a higher weight than a difference at the bottom of the ranking.

Figure 3. Kemeny comparison

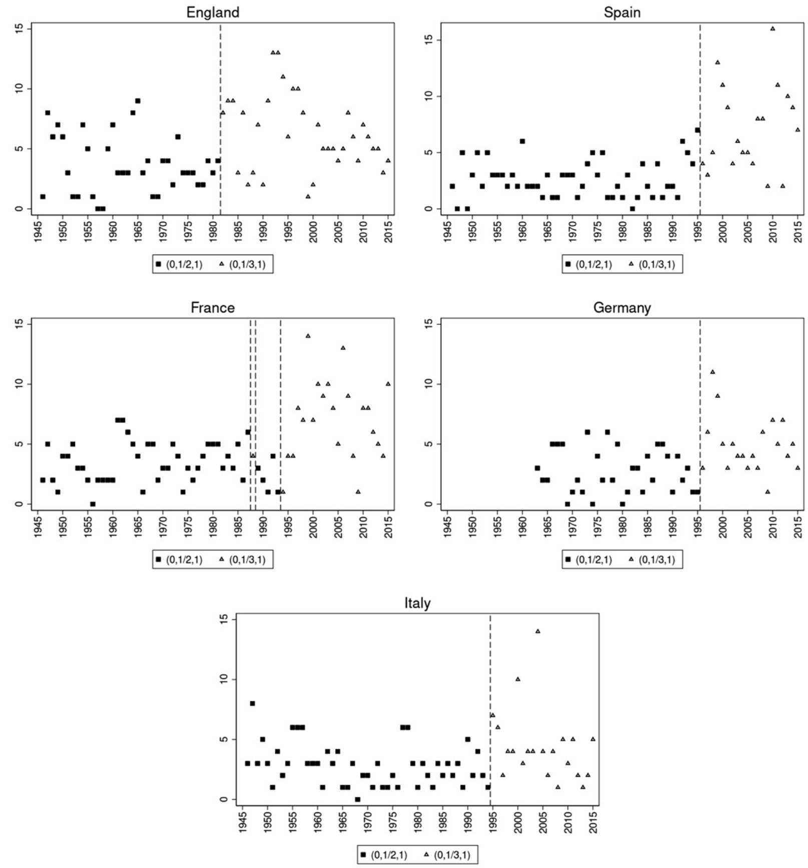


Figure 3 shows that, for a number of years, the RA ranking was exactly the same as the one obtained with the real-world ranking method, whereas for other years the difference between the two ranking methods was notable. For example, in Spain, the Kemeny distance is equal to 16 in 2010 whereas it is equal to 0 in 1982.

Nevertheless, the ranges of the Kemeny distance are smaller when the ranking method is $w_{ij} = \left(1, \frac{1}{2}, 0\right)$. The corresponding ranges for this ranking method lie between 0 and 9, while the ranges for the ranking method $w_{ij} = \left(1, \frac{1}{3}, 0\right)$ go from 1 to 16. This is confirmed by tests of difference of means for each country.[8] These tests are significant at a 1% level for all countries except Italy, for which the test is significant at a 5% level.

5. The best scoring method according to *RA*

The closeness observed between RA and scoring methods $w_{ij} = \left(1, \frac{1}{2}, 0\right)$ is studied in detail in this section. We look for the ranking method that best fits RA, proceeding in three steps:

1. We fix the scoring rule $w_{ij} = \left(1, \frac{1}{2}, 0\right)$ (or $w_{ij} = \left(1, \frac{1}{3}, 0\right)$) and determine the RA rule $w_{ij} = (1, \alpha, 0)$ that minimize our Kemeny measure.
2. We fix the RA rule $w_{ij} = \left(1, \frac{1}{2}, 0\right)$ (or $w_{ij} = \left(1, \frac{1}{3}, 0\right)$) and determine the scoring rule $w_{ij} = (1, \beta, 0)$ that minimize our Kemeny measure.
3. We consider the RA rule $(1, \gamma, 0)$ and the scoring rule $(1, \gamma, 0)$ and determine the γ that minimizes our Kemeny measure.

5.1. First step: α minimizing the Kemeny measure

Figure 4 shows, for all leagues and for both $w_{ij} = \left(1, \frac{1}{2}, 0\right)$ and $w_{ij} = \left(1, \frac{1}{3}, 0\right)$, the distributions of α that minimize our Kemeny measure. For each year, the α min is not always unique and the distributions of α min are not always continuous. The comparison of the distributions of α min when $w_{ij} = \left(1, \frac{1}{2}, 0\right)$ and for $w_{ij} = \left(1, \frac{1}{3}, 0\right)$, summarized in Table 10, shows that the intersection of both distributions is often empty (*i.e* there is not an α min that minimizes Kemeny for the two distributions). Figure 4 also shows that there are a few years for which the distributions of α min are identical both for $w_{ij} = \left(1, \frac{1}{2}, 0\right)$ and $w_{ij} = \left(1, \frac{1}{3}, 0\right)$ (*i.e* the α min that minimizes Kemeny for the two distributions are the same). Similarly, a same and unique α min for $w_{ij} = \left(1, \frac{1}{2}, 0\right)$ and $w_{ij} = \left(1, \frac{1}{3}, 0\right)$ is rather unusual.

8. We used a "Two-sample t test with unequal variances" with Stata.

Figure 4. α **min**

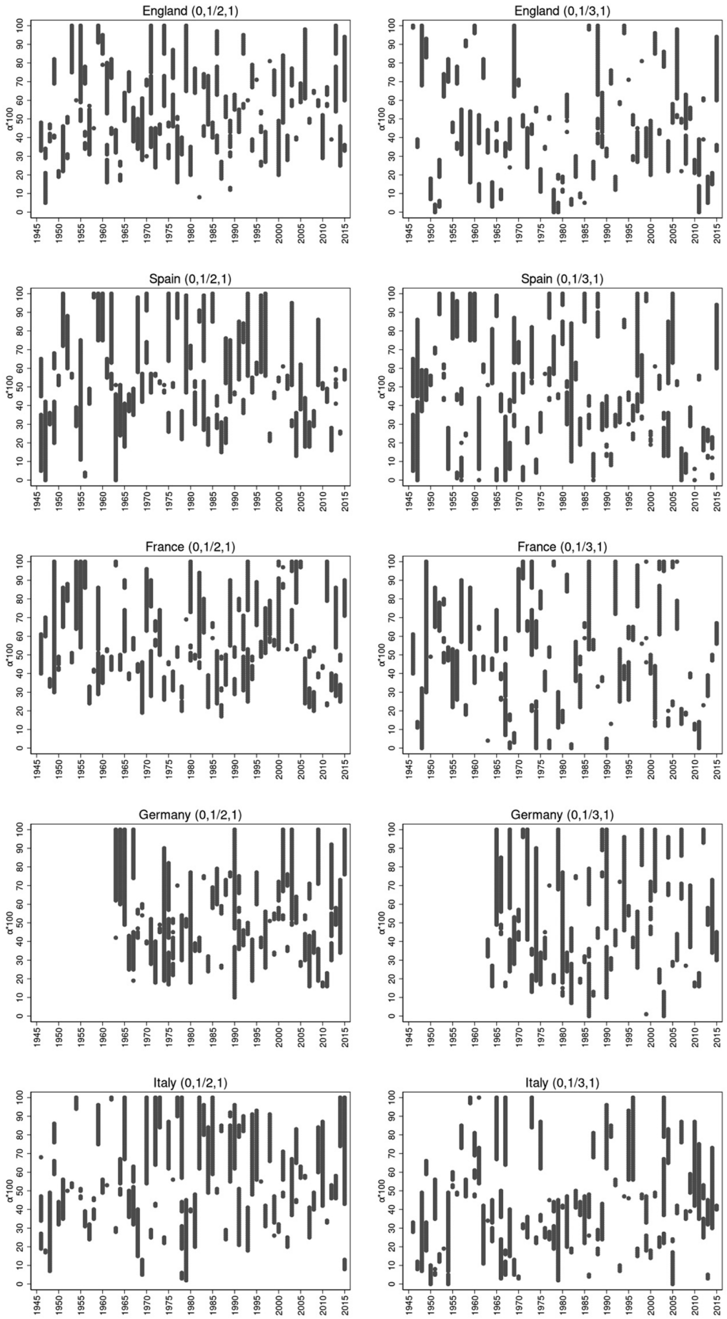


Table 10. *α comparison*

	Empty intersection	Identical Distribution	Unique α min
France	$\frac{46}{69} = 0.67$	9	
England	$\frac{51}{69} = 0.74$	8	1995(0.71), 1998(0.81)
Germany	$\frac{30}{52} = 0.58$	13	1977(0.70)
Italy	$\frac{48}{69} = 0.70$	10	
Spain	$\frac{45}{69} = 0.65$	10	2001 (0.61)

Table 11. **β comparison**

	Empty Inter	Iden Distri	Unique β min
France	$\frac{34}{69} = 0.49$	25	05(0.51), 04(0.46), 98(0.43), 63(0.45)
England	$\frac{41}{69} = 0.59$	14	03(0.56), 59(0.49), 54(0.56), 50(0.47)
Germany	$\frac{20}{52} = 0.38$	20	01(0.65), 71(0.50)
Italy	$\frac{28}{69} = 0.41$	28	07(0.50), 00(0.50), 70(0.55), 69(0.52), 51(0.50)
Spain	$\frac{34}{69} = 0.49$	21	15(0.53), 64(0.50), 54(0.50), 51(0.55), 48(0.50)

5.2. Second step: β minimizing the Kemeny measure

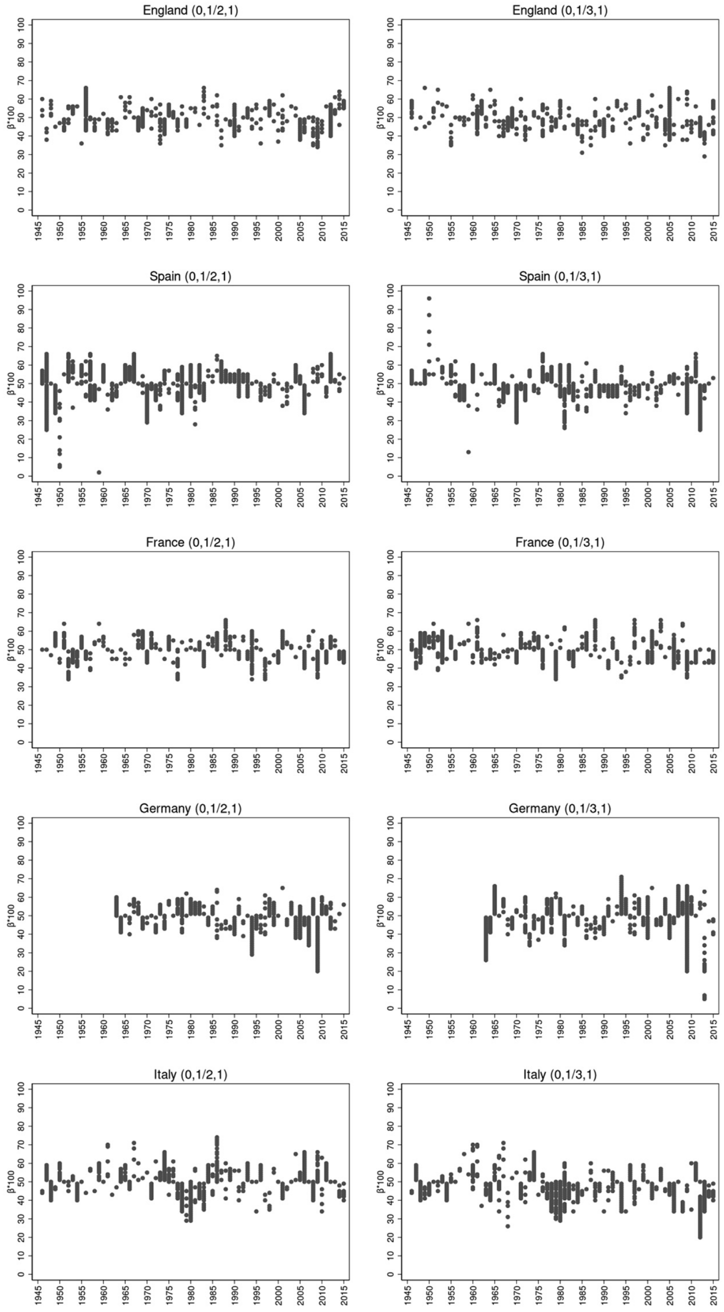
We now investigate the distribution of β that minimizes our Kemeny measure for each league. The β min are given in Figure 5.

It is very interesting to observe that the distributions of β are much more concentrated than those of α. As an example, the β min for France varies from 0.30 to 0.70. Once again, the β min is not always unique and the distributions are not always continuous.

Table 11 summarizes the main information between a distribution $w_{ij} = \left(1, \frac{1}{2}, 0\right)$ and a distribution $w_{ij} = \left(1, \frac{1}{3}, 0\right)$.

Note that the proportion of identical distributions is higher for β than for α. A unique β min is also more frequent. In other words, it is more common to find identical (possibly unique) distributions of β min in cases $w_{ij} = \left(1, \frac{1}{2}, 0\right)$ and $w_{ij} = \left(1, \frac{1}{3}, 0\right)$ than for α. We note that if there is a unique β min, its value is closest to $\frac{1}{2}$.

Figure 5. ***β* min**



One can also observe the Kernel densities of β min in Figure 6. For all leagues, the most likely β min is around $\frac{1}{2}$. This is reflected by inverted V-shaped curves for each league taken into account. This suggests that a scoring rule that gives $\frac{1}{2}$ points to a draw is the best one to represent the strength and the weakness of the teams.

Figure 6. β Kernel densities

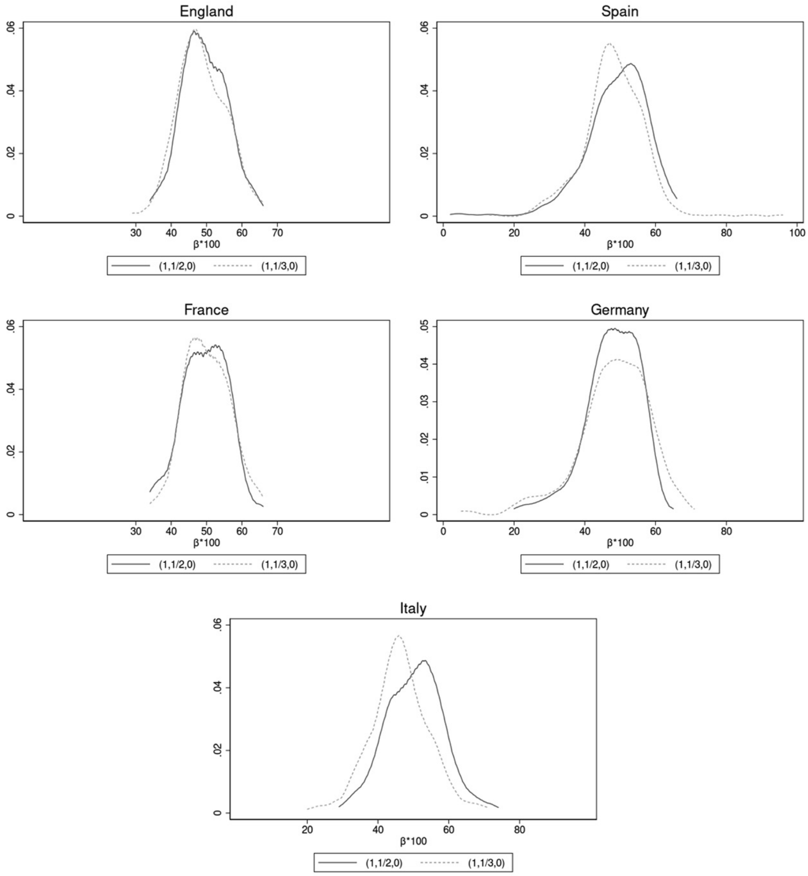
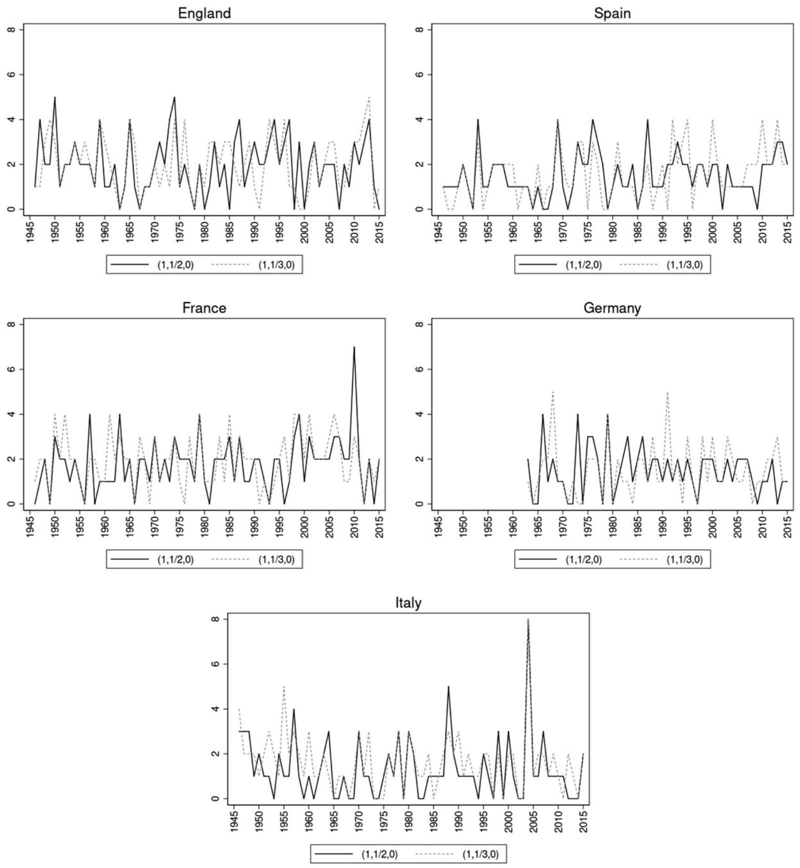


Figure 7 presents the Kemeny measures associated with each β min and shows that for many win-loss matrix distributions a β min implies a null Kemeny score. On the other hand, for some years, the Kemeny measure was very high with $w_{ij} = \left(1, \frac{1}{3}, 0\right)$. Figure 7 also shows that the minimal Kemeny score corresponding to the β min is often lower with $w_{ij} = \left(1, \frac{1}{2}, 0\right)$ (the blue curve is often below the red curve). This again suggests that $w_{ij} = \left(1, \frac{1}{3}, 0\right)$ does not reflect well the RA method.

Figure 7. Kemeny min for β



The second step provides a tentative answer: the scoring rule $w_{ij} = \left(1\,; \frac{1}{2}, 0\right)$ would best reflect the RA method.

5.3. Third step: γ minimizing the Kemeny measure

The previous two steps seem to indicate that choosing an α and a β close to $\frac{1}{2}$ reduces the distance between the RA method and the score methods. This third and final step confirms our previous observations.

It is clear that the distributions of γ (see Figure 8) for all leagues converge to $\frac{1}{2}$. This result is corroborated by the Kernel densities of γ min (see Figure 9) where we have inverted V-shaped curves with an optimum at $\frac{1}{2}$. Consequently, this third step leaves no doubt that a scoring rule with $\frac{1}{2}$ to a draw minimizes the difference between the scoring rule and the RA method, indicating that it best represents the strength and the weakness of the teams.

6. Conclusion

This paper appraises the ability of ranking methods commonly used in soccer championships to account for both the strength of the teams defeated by a given team j and the strength of the teams that beat team j.

Taking into account the five major UEFA championships over the period of 1945-2016, we first investigate the likelihood of at least one change occurring in the final ranking. We show that the number of differences between the Ramanujacharyula method and the scoring rule $\left(1\,;\frac{1}{2},0\right)$ is smaller than that between the Ramanujacharyula method and the scoring rule $\left(1\,;\frac{1}{3},0\right)$.

The first pattern is reinforced by the optimal value of γ, shown as equal to 1/2. In other words, a scoring rule with $\gamma=\frac{1}{2}$ for a draw minimizes the difference between the scoring rule $(1,\gamma,0)$ and the Ramanujacharyula method and, therefore, best represents the strength and the weakness of the teams.

Nevertheless, our research also sets forth that the scoring rule $\left(1\,;\frac{1}{2},0\right)$ does not perfectly fit the Ramanujacharyula method (as illustrated by Figure 9). This raises the question of the existence of alternative scoring methods that would exhibit a better fit than the one obtained with the scoring rule $\left(1\,;\frac{1}{2},0\right)$. Assessing this possibility would be feasible by implementing an additional rule to the scoring method $\left(1\,;\frac{1}{2},0\right)$, such as the bonus and/or malus system currently used in rugby. Another solution would be to jointly use a bonus and/or malus points rule with the UEFA club coefficient taking part in the championship in order to account for their strength. Note that

Csató [2024] showed recently that a variant of the Elo rating is a more accurate predictor of performance in the UEFA Champions League than the UEFA club coefficient.

Figure 8. γ **min**

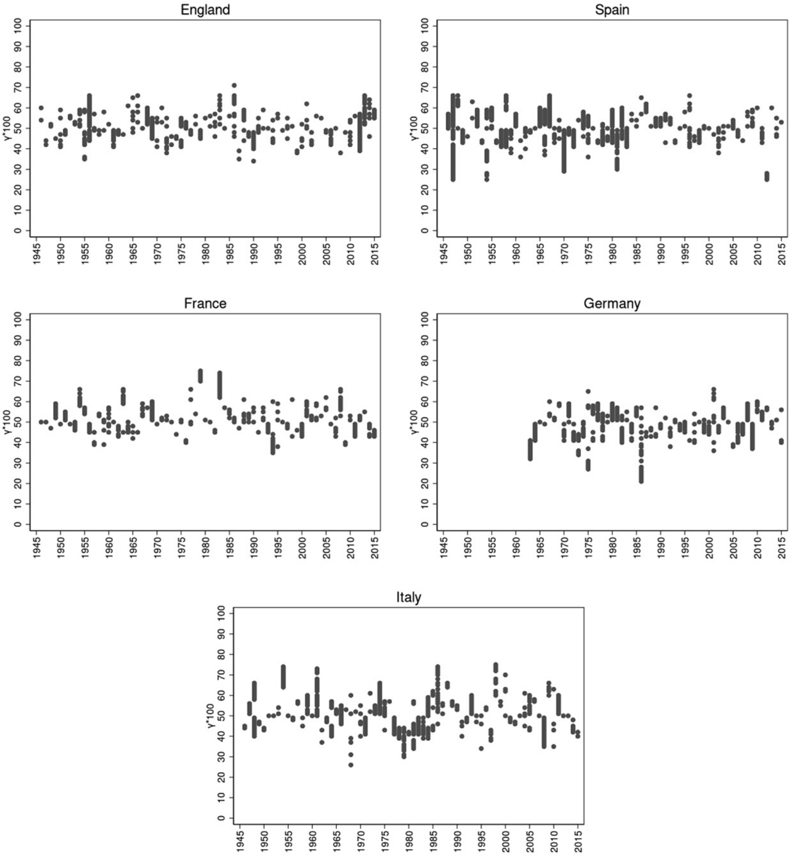
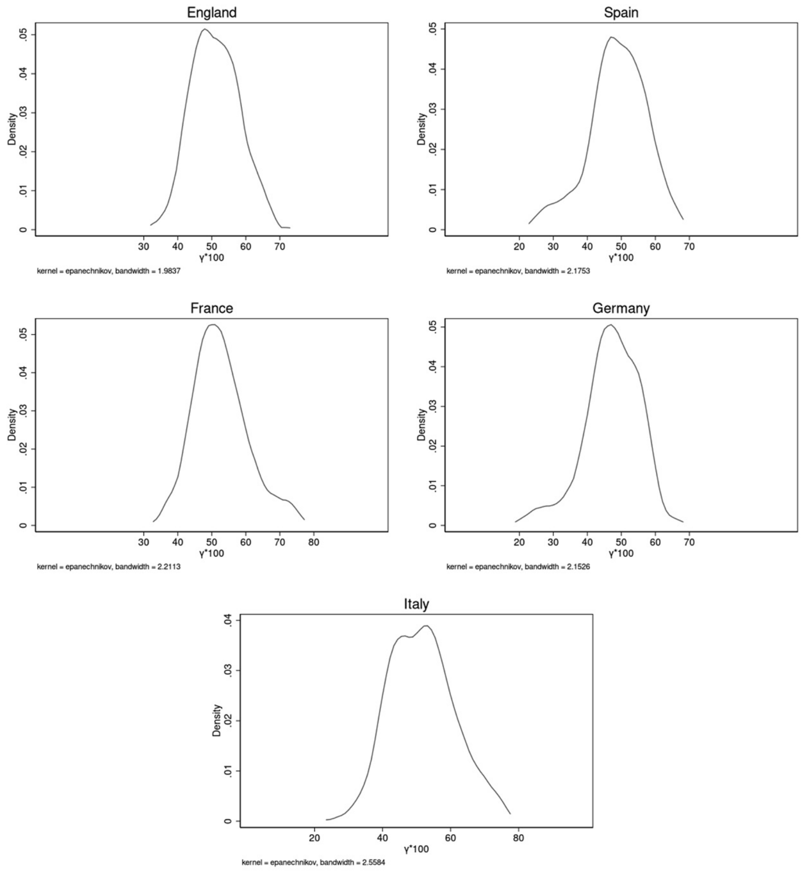


Figure 9. γ **Kernel densities**



It seems clear that the Ramanujacharyula method is not usable in its current form in the real world. However, the prevailing rules for breaking a tie score not being flawless, the Ramanujacharyula method, although complicated, could be used to break a tie when considering the Champion, the European Positions or the Drop Zone.[9] Other global methods could also be further investigated, since "there are many possible modifications of such procedures" (Hasse and RA) as suggested by Chebotarev and Shamis [1998].

9. Some tournaments use rather complicated tie-breaking rules, see Freixas [2022] for an example.

References

ALTMAN A., TENNENHOLTZ M. [2005], Ranking systems: the PageRank axioms, in: *Proceedings of the 6th ACM conference on electronic commerce*, New York, ACM Press, 1-8.

BARROW D., DRAYER I., ELLIOTT P., GAUT G., OSTING B. [2013], Ranking rankings: An empirical comparison of the predictive power of sports ranking methods. *Journal of Quantitative Analysis in Sports*, 9(2), 187-202.

BEHZAD M., CHARTRAND G., LESNIAK-FOSTER L. [1979], *Graphs and digraphs*., Boston, Prindle Weber and Schmidt.

BOUYSSOU D. [1992], Ranking methods based on valued preference relations: A characterization of the net flow method, *European Journal of Operation Research*, 60(1), 61-67.

BRIN S., PAGE L. [1998], The anatomy of a large-scale hypertextual web search engine, *Computer Network and ISDN Systems*, 30, 107-117.

CAN B. [2014], Weighted distances between preferences, *Journal of Mathematical Economics*, 24, 109-115.

CHEBOTAREV P. Y., SHAMIS E. [1998], Characterizations of scoring methods for preference aggregation, *Annals of Operations Research*, 80, 299-332.

CSATÓ L. [2023], How to avoid uncompetitive games? The importance of tie-breaking rules, *European Journal of Operational Research*, 307(3), 1260-1269.

CSATÓ L. [2024], Club coefficients in the UEFA champions league: Time for shift to an elobased formula, *International Journal of Performance Analysis in Sport*, 24(2), 119-134.

DANIELS H. E. [1969], Round robin tournament scores, *Biometrika*, 56(2), 295-299.

DEVRIESERE K., CSATÓ L., GOOSSENS D. [2024], Tournament design: A review from an operational research perspective, *DOI: 10.48550/arXiv.2404.05034.*

ESCHENBACH C., HALL F., HEMASINHA R., KIRKLAND S., LI Z., SHADER B., STUART J., WEAVER J. [1982], Properties of tournaments among well-matched players , *The American Mathematical Monthly*, 107, 881-892.

FREIXAS J. [2023], The decline of the Buchholz tiebreaker system: A preferable alternative, in: Nguyen N. T., Kowalczyk R., Mercik J., Motylska-Kuźma A. (eds.), *Transactions on Computational Collective Intelligence XXXVII*, Heildeberg, Springer, 1-20.

GONZÁLEZ-DÍAZ J., HENDRICKX R., LOHMANN E. [2014], Paired comparisons analysis: An axiomatic approach to ranking methods, *Social Choice and Welfare*, 42(1), 139-169.

GUEDES J. C., MACHADO F. S. [2002], Changing rewards in contests: Has the three-point rule brought more offense to soccer? *Empirical Economics*, 27(4), 607-630.

HAMMOND T. H. [2007], Rank injustice: How the scoring method for cross-country running competitions violates major social choice principles, *Public Choice*, 133(3), 359-375.

HASSE M. [1961], Uber die behandlung graphentheoretischer probleme under verwendung der matrizenrechnung, *Wissenschaftliche Zeitschrift Technischen Universitat Dresden*, 10, 1313-1316.

HENRIET D. [1985], The Copeland choice function: An axiomatic characterization, *Social Choice and Welfare*, 2(1), 49-63.

HERINGS J. J., VAN DER LAAN G., TALMAN D. [2005], The positional power of nodes in digraphs, *Social Choice and Welfare*, 24(3), 439-454.

KEENER J. P. [1993], The Perron–Frobenius theorem and the ranking of football teams, *SIAM Review*, 35(1), 80-93.

KEMENY J. [1959], Mathematics without numbers, *Daedalus*, 88(4), 571-591.

KENDALL M. G. [1955], Further contributions to the theory of paired comparisons, *Biometrics*, 11(1), 43-62.

KITTI M. [2016], Axioms for centrality scoring with principal eigenvectors, *Social Choice and Welfare*, 46(3), 639-653.

KONDRATEV A. Y., IANOVSKI E., NESTEROV A. S. [2023], How should we score athletes and candidates: Geometric scoring rules, *Operations Research, in press DOI:10.1287/opre.2023.2473.*

LASLIER J.-F. [1997], *Tournament solutions and majority voting*, Berlin, Springer.

MOON J. W., PULLMAN N. J. [1970], On generalized tournament matrices, *SIAM Review*, 12(3), 384-399.

MOON J. W. [1968], *Topics on toumament*, New York, Holt, Rinehart and Winston.

MOSCHINI G. [2010], Incentives and outcomes in a strategic setting: The 3 points for a win system in soccer, *Economic Inquiry*, 48(1), 65-79.

RAMANUJACHARYULA C. [1964], Analysis of preferential experiments, *Psychometrica*, 29(3), 257-261.

RUBINSTEIN A. [1980], Ranking the participants in a tournament, *SIAM Journal of Applied Mathematics*, 38(1), 108-111.

SAARI D. [2001], *Chaotic elections! A mathematician looks at voting*. Providence, American Mathematical Society.

SCARF P., PARMA R., MCHALE I. [2019], On outcome uncertainty and scoring rates in sport: The case of international rugby union, *European Journal of Operational Research*, 273(2), 721-730.

SLUTZKI G., VOLIJ O. [2006], Scoring of webpages and tournaments axiomatizations, *Social Choice and Welfare*, 26(1), 75-92.

SZCZECINSKI L. [2023], Why winning a soccer game is worth 5 points, *DOI: 10.48550/arXiv.2303.15305*.

TIERNEY J. [2014], Soccer, a beautiful game of chance, *The New York Times*, July 7 2014.

TRUCHON M. [1998], Figure skating and the theory of social choice, *Cahier de recherche 9814,* Université Laval.

VAN DEN BRINK R., GILLES R. P. [2003], Ranking by outdegree for directed graphs, *Discrete Mathematics*, 271(1-3), 261-270.

WEI T.-H. [1952], *The algebraic foundation of ranking theory*. Doctoral Thesis, Cambridge, Cambridge University.