Working papers series

WP ECON 22.13

Sports competitions and the Break-Even rule

Carmen Herrero
Department of Economics, Universidad de Alicante

Antonio Villar
Department of Economics, Universidad Pablo de Olavide

Keywords: sports competitions; break-even rule; teams' performance; strength; handicap; Premier League.

JEL Classification: C65, L83.

Department of Economics

Sports competitions and the Break-Even rule

Carmen Herrero

FAE, Universidad de Alicante & Ivie

Antonio Villar

Universidad Pablo de Olavide & ISEAK

ABSTRACT

Sports competitions represent an interesting family of evaluation problems involving pairwise comparisons. In this context, the alternatives are contending teams and the comparison is made in terms of outcomes. Different evaluation protocols, aimed at getting more robust estimates of the teams' worth or better predictions of their future achievements, have been proposed in the literature. We present here a protocol that makes the evaluation of a team directly proportional to the sum of the points accrued along with the competition, weighted by the worth of the competitors (its *strength*) and inversely proportional to the total points lost (its *handicap*). We call this new evaluation protocol the *break-even* rule and show that it is well-defined and easy to compute.

Keywords: sports competitions; break-even rule; teams' performance; strength; handicap; Premier League.

1. Introduction

Sports competitions constitute a special family of evaluation problems based on pairwise comparisons, whose informational inputs correspond to outcomes of matches between teams or players. Different sports have different rules to compute victories and defeats and to determine the ranking of the competing teams. Yet, there is scope for the analysis of those competitions from a conceptual viewpoint, as they represent evaluation problems with a specific structure and particular features.

The classification of the teams that enter a competition is usually determined by the sum of the points accrued during the season. Most often the team that wins a match gets credited some points whereas the loser gets nothing. Adding up those points yields the final classification, even though the way of computing those points varies from one sport to another. The literature on sports competitions suggests several richer protocols to evaluate the teams' performance. Those protocols may have multiple purposes: to help enhance competitiveness, provide estimates of the teams' commercial worth, obtain better predictions of future outcomes, define shadow prices of some assets (e.g. broadcasting rights), or gather competitors into comparable categories. Note that, in addition to the conceptual appeal of those considerations, there are many people interested in sports competitions and a great deal of money involved, and the evaluation of the teams' performance, therefore, has quite a bite.

There are two main aspects that those richer evaluation protocols introduce. On the one hand, the points won and lost by the teams in the competition are considered, rather than exclusively the points won. On the other hand, pondering the points obtained by the strength of the competing teams. Both aspects affect the incentives of the teams in the competition and help obtain a richer evaluation of the contenders. The most popular evaluation rules that consider those aspects are probably the Elo (1978) rating, the rankings proposed by Massey (1997) and Colley (2002), and Keener (1993) strength proportionality rule and its variants.

The Elo rating was initially conceived to classify chess players and was then extended to many other competitions. The basic idea is to give player i, when playing against player j, a rating r_i that reflects the probability of winning against its opponent, based on the frequency of their victories. The player's evaluation is adjusted by making the change in the rating proportional to the difference between the actual score, s_{ij}, and the expected score, $E(s_{ij})$. That is, $r_i^t = r_i^{t-1} + K\big(s_{ij} - E(s_{ij})\big)$, where $K > 0$ is a sensitivity constant, and $E(.)$ the expected score is drawn from a logistic function of the rating.

The idea behind Massey's and Colley's procedures is that the evaluations of the competing teams should be proportional to their differential outcomes. In the case of Massey, the evaluation obtains as the vector $\boldsymbol{r}$ that solves the system $\mathbf{M}\boldsymbol{r} = \boldsymbol{p}$, where $\mathbf{M} =$

$\{m_{ij}\}$ is the so-called Massey matrix, which is given by $m_{ii} = N$ (the total number of games played by team i), $m_{ij} = -n$ (where n stands for the number of games played between any two teams), and $\boldsymbol{p}$ is the vector of cumulative score differences of each team. Colley proposes to evaluate outcomes according to the solution to the system $\mathbf{C}\boldsymbol{r} = \boldsymbol{b}$, where $\mathbf{C} = \{c_{ij}\}$ is the Colley matrix, given by: $c_{ii} = 2 + N$, $c_{ij} = -n, i \neq j$. Vector $\boldsymbol{b}$ corresponds to the cumulative win-loss differentials for each team, that is, $b_i = 1 + 0.5(w_i - l_i)$, where w_i, l_i are the total games won and lost by *i*, respectively. In both cases, the evaluation derives from the solution of a linear system, applies a proportionality principle, and focuses on outcome differences.

An alternative stream of contributions introduces the idea of pondering outcomes on the overall strength of the teams.[1] A key reference in this approach is Keener (1993), who states explicitly (p. 81), "To each participant in a contest we wish to assign a score that is based on the interactions with other participants. The assigned score should depend on both the outcome of the interaction and the strength of its opponents." To solve the evaluation problem by computing both the outcomes and the strength of the opponents derived from those outcomes, Keener recurs to a fixpoint argument, which in this context corresponds to the dominant eigenvector of a matrix $\mathbf{A} = \{a_{ij}\}$, where a_{ij} reflects what *i* obtains in its confrontation with *j*. In the simplest case, he considers $a_{ij} = 1$, if team *i* wins, equal to ½ in case of a tie, and equal to 0 if it loses. By letting r_j stand for the strength of team j, the score of team i corresponds to the sum of the outcomes weighted by the strength of the competitors, $s_i = \sum_j a_{ij} r_j$, Keener proposes that the score be proportional to the strength. That is, $\mathbf{A}\boldsymbol{r} = \lambda\boldsymbol{r}$. He also discusses different ways of defining the outcomes that can be embedded within this evaluation procedure. We here find again a proportionality principle but now consider the strength of the teams at the heart of the evaluation.

A variant of Keener's contribution is that in which the evaluation procedure consists of a Markov chain (e.g., Govan, 2008, Govan, Langville & Meyer, 2009). This procedure derives from a "voting" matrix. Each player gives a "vote" to the other in a pairwise confrontation between players *i* and *j*. Therefore, m_{ij} is the vote of *j* to *i*, and m_{ji} the other way around. The columns of the matrix are normalized so that it becomes a Markov matrix (dividing each element by the sum of the column). The evaluation is given by the stable distribution of that Markov matrix.

The reducibility of the outcome matrices, which is usually found in empirical applications, is regarded as problematic since it implies that some teams may receive the same evaluation even though they present different outcomes. Vaziri, Yih & Morin (2018)

---

[1] The Elo rule also introduces in the evaluation the relative strength of the competitors, even though it is defined pairwise and using a different strength function.

introduce the so-called $(1,\alpha)$ method to deal with that "inconvenience" (see below). In this method, the winning team gives a vote of 1 to the losing team, and the losing one gives a vote of $\alpha > 1$ to the winning team. The Markov matrix becomes, therefore, irreducible. Yet, the final evaluation depends critically on that α parameter, which is chosen arbitrarily by the evaluator. We here find similar features to Keener's evaluation protocol, but the cardinal evaluation depends on the choice of the α parameter.

Vaziri, Yih & Morin (2018) show that the $(1,\alpha)$ method satisfies three desirable properties for a suitable range of values of the parameter α: (1) To consider the strength of the opponents (*comprehensiveness*); (2) To provide incentives to always win a match (*monotonicity*); and (3) Independence of the sequence of matches (*fairness*). They also show that other Markov procedures, Elo's, or those obtained as the solutions of linear equations systems, fail to satisfy at least one of those three properties.

We propose in this paper a rule that evaluates the teams' performance by computing both wins and losses, and their strength. This evaluation rule, called the ***break-even rule***, assigns to each team a value that is directly proportional to its strength and inversely proportional to its handicap. Strength is defined as the worth of the points accrued by a team (points obtained weighted by the evaluations of the other teams). Handicap corresponds to its complement: the worth of the points conceded to the other teams along with the competition.

The paper is organized as follows. Section 2 presents the evaluation protocol and provides an application that helps visualize how it works. It refers to the 21-22 English Premier League, in which we compare the evaluation derived from the break-even rule and the official one (points accrued through the competition). Section 3 is devoted to discussing some relevant features of our evaluation rule. A few final words, in Section 4, close the work.

2. Methods and results

2.1 The evaluation of the teams' performance

A sports competition consists of a set of teams, $M = \{1,2,\dots,m\}$, that compete in pairwise confrontations a number n of rounds in a given period or *season*. We assume that competitions adopt the format of symmetric round-robin tournaments to facilitate the discussion. That is, each team plays exactly n times against each other, and all rounds are equally worthy. For each match between teams $i,j \in M$, and each round $h \in N = \{1,2,\dots,n\}$, the term $p_{ij}(h) \in \mathbb{R}_+$ describes the ***partial outcome*** obtained by i when playing against j in round h. There are several ways of defining this outcome variable (see below), but what follows does not depend on a particular choice. We shall refer to the units in which those outcomes are measured as “points”. Since no team can obtain points by playing against itself, we let $p_{ii}(h) = 0, \forall\, i, h$.

We denote by p_{ij} the ***outcome*** of team i with respect to team j, which corresponds to the sum of all the points obtained when playing against team *j* along with the competition, i.e., after n rounds. That is,

$$p_{ij} = \sum_{h=1}^{n} p_{ij}(h) \qquad [1]$$

All the relevant information from the competition can be summarized into a square m matrix $\mathbf{P}$ whose entries are the outcomes p_{ij} corresponding to the total points obtained by team *i* when playing against team *j*. We shall therefore identify an evaluation problem, involving m teams, with the matrix $\mathbf{P} = \left(p_{ij}\right)_{i,j=1}^{m} \in \Pi$, where $\Pi \subset \mathbb{R}_{+}^{m^2}$ denotes the set of all non-negative matrices of order m with zeroes in the main diagonal. We aim to provide an evaluation of the teams in M that participate in a competition, based on those outcomes. More precisely, we look for an ***evaluation rule*** $f: \Pi \to \mathbb{R}_{+}^{m}$, such that an m-vector $f(\mathbf{P})$ is associated with each problem $\mathbf{P}$ such that $f_i(\mathbf{P}) \geq f_j(\mathbf{P})$ implies that team i is regarded as better than or equal to team j.

Let $\mathbf{P}$ be a given an evaluation problem and $f: \Pi \to \mathbb{R}_{+}^{m}$ an evaluation rule. Then, $p_{ij} f_j(\mathbf{P})$ is the worth of the outcome *obtained* by team i when playing against team *j*, whereas $p_{ji} f_i(\mathbf{P})$ is the worth *conceded* by team i to team *j*. We now introduce the concepts of strength and handicap of a team under that evaluation rule.

The ***strength*** of team i under rule *f* is simply the sum of the worth that this team obtains along with the competition:

$$S_i(\mathbf{P}, f) = \sum_{j \neq i} p_{ij} f_j(\mathbf{P}) \qquad [6]$$

That is, the strength of a team is the weighted sum of the points accrued in the different matches, where the weights correspond to the evaluations of the other teams. A team is thus stronger when it wins more often and/or when it beats teams with higher evaluations.[2]

---

[2] The notion that the evaluation of a team’s performance must consider the strength of the competitors is a familiar principle in many environments. It expresses the principle that beating a top tier in a pairwise confrontation is more relevant than beating a low-level alternative. Think, for instance, of the evaluation of the impact of academic journals (e.g. Pinsky & Narin, 1976, Liebowitz & Palmer, 1984, Palacios-Huerta & Volij, 2004, or Albarrán et al, 2017), as now implemented by the Eigenfactor (see http://www.eigenfactor.org). A similar principle is applied in the celebrated Google search engine (Brin & Page, 1998) and appears in many evaluation protocols regarding tournaments (Daniels, 1969, Moon & Pullman, 1970, Laslier, 1997, Saaty, 2003, Slutzski & Volij 2006, Boccard, 2022).

Similarly, we define the ***handicap*** of team i under rule f, as:

$$h_i(\mathbf{P}, f) = f_i(\mathbf{P}) \sum\nolimits_{j \neq i} p_{ji} \qquad [6]$$

The handicap of a team is the worth of the points conceded to other teams along with the competition.

Definition: The ***break-even rule***, f^{BE}, is the evaluation rule that, for each problem $\mathbf{P} \in \Pi$, associates to each team the value that equals strength and handicap, that is,

$$S_i(\mathbf{P}, f^{BE}) = h_i(\mathbf{P}, f^{BE}), \forall\, i$$

The break-even rule mimics the no-arbitrage condition in the evaluation of financial assets. It says that the worth attached to a team, $f_i^{BE}(\mathbf{P})$, is directly proportional to its strength and inversely proportional to its handicap. It is therefore an intuitive evaluation rule that computes the relevance of the competing teams (strength) and their weakness (handicap). The obvious question is how to obtain those evaluations, which must be simultaneously determined for all teams. The following result gives us the answer.

Proposition: *(i) The break-even rule is a well-defined and continuous mapping such that, to each $\boldsymbol{P} \in \Pi$, associates a vector $f^{BE}(\boldsymbol{P}) \in \mathbb{R}_+^m$. (ii) The evaluation of any problem can be easily calculated by employing standard algorithms. (iii) In the subset of irreducible non-negative matrices, the break-even rule is a continuous vector-valued function.*

Proof

(i) Given a square non-negative matrix $\mathbf{P} \in \Pi$, of order m, let Q denote a scalar large enough so that $Q > max_i\{\sum_{j \neq i} p_{ji}\}$. Now define another non-negative square matrix of the same order, given by: $\mathbf{P}' = \mathbf{P} + \widehat{\mathbf{G}}(\mathbf{P})$, where $\widehat{\mathbf{G}}(\mathbf{P})$ is a diagonal matrix, whose diagonal elements are of the form: $g_{ii} = Q - \sum_{j \neq i} p_{ji}$.

As **P'** is a non-negative square matrix whose columns add up to Q, the Perron-Frobenius Theorem (e.g. Berman & Plemmons, 1994) ensures that Q is the dominant eigenvalue of this matrix and has associated with it a non-negative eigenvector $\boldsymbol{v} \in \mathbb{R}_+^m$. That is, $\mathbf{P}'\boldsymbol{v} = Q\boldsymbol{v}$. The ith component of this eigenvector thus satisfies:

$$\left(Q - \sum\nolimits_{j \neq i} p_{ji}\right) v_i + \sum\nolimits_{j \neq i} p_{ij}\, v_j = Q v_i$$

From which it follows that:

$$v_i \sum\nolimits_{j \neq i} p_{ji} = \sum\nolimits_{j \neq i} p_{ij}\, v_j$$

That is precisely the break-even condition.

The break-even rule, f^{BE}, corresponds therefore to the dominant eigenvector of the matrix **P'** associated with matrix **P**, and is independent of the value of the constant Q. The Perron-Frobenius Theorem ensures that $f^{BE}(\mathbf{P})$ is non-empty and non-negative, for all $\mathbf{P} \in \Pi$, and varies continuously with **P**.

(ii) Computing the dominant eigenvectors of a non-negative matrix is an immediate task for which there are conventional algorithms.

(iii) When, furthermore, matrix **P** is irreducible, the dominant eigenvector is unique (except for the choice of units) and strictly positive (we apply again the Perron-Frobenius Theorem), so that f^{BE} is a continuous vector-valued function.

Q.e.d.

We can write more explicitly the evaluation of team *i* as follows:

$$f_i^{BE}(\mathbf{P}) = \frac{\sum_{j=1}^{m} p_{ij} f_j^{BE}(\mathbf{P})}{\sum_{j=1}^{m} p_{ji}}, \qquad \forall\, i \qquad [7]$$

Remark: *We assume implicitly that* $\sum_{j=1}^{m} p_{ji} > 0$ *for all i. Note that* $\sum_{j=1}^{m} p_{ji} = 0$ *implies that team* i *wins all matches, a point that will be discussed later.*

2.2 An application: The Premier League

We now apply that evaluation procedure to a real sports competition: The 2021-2022 English football league, *The Premier League*. The first division of the English football championship consists of 20 teams that compete twice against each other (i.e., two rounds) throughout the season. A team obtains 3 points when it beats another, 1 point in case of a draw, and 0 points otherwise. The competition's outcome corresponds to the sum of all those points over the season. This is, therefore, an evaluation in which the number of defeats is immaterial (except for breaking ties), and the number of points is independent of the strength of the competing teams.

We apply here the break-even rule, to the 2021-2022 English football championship, to illustrate how this evaluation operates and how it differs from the official outcomes.

Table 1 provides the evaluations according to those two criteria. We normalize the evaluations to a common maximum of 100 points for the winner. Note that the break-even rule discriminates much more than the official competition, with a coefficient of variation of 1.118, for 0.359 for the official results. There are many changes in the associated ranking of the break-even rule, and those changes include another winner of the competition: Liverpool, rather than Manchester City, would be first due to its better performance vis a vis the top tiers.

Table 1: Normalized points and rank of the Premier League 2021-22, according to the official scoring system and the break-even rule

	Official classification		Break-even rule	
Team	Rank	Points	Rank	Points
Arsenal	5	74	5	22
Aston Villa	14	48	16	9
Brentford	13	49	13	12
Brighton	9	55	9	18
Burnley	18	38	17	8
Chelsea	3	80	3	41
Crystal Pal	12	52	8	19
Everton	16	42	15	9
Leeds Un	17	41	18	6
Leicester City	8	56	10	16
Liverpool	2	99	1	100
Man City	1	100	2	98
Man Utd	6	62	7	20
Newcastle	11	53	14	11
Norwich	20	24	20	3
Southampton	15	43	11	13
Tottenham	4	76	4	39
Watford	19	25	19	3
West Ham	7	60	6	21
Wolverhampton	10	55	12	13

Regarding the implications of using the break-even rule, we can think of the effect on incentives and the market value of some of the teams' assets. In terms of incentives, one would expect applying this rule will induce tougher competition. This is so because losing not only implies that the team does not add points but also that it is penalized. Moreover, weak teams will have a premium when beating top tiers and so will be expected to fight harder (and strong teams would know that which also induces greater effort as the distance between teams will shorten in a convex way). Concerning market values, those evaluation protocols may provide a basis to define the range of certain financial variables, from the worth of a team to the allocation of broadcasting rights, for example.

3. Discussion

We have presented an evaluation rule for sports competitions that makes each team's rating a function of its handicap and its strength. Both elements affect the incentives and tend to enhance the differences between the teams' ratings, as they make each victory worthier and each defeat more harmful. Moreover, they introduce a link between each particular outcome and all the rest. We now discuss several aspects of this rule, which help us understand its meaning and scope.

3.1 Sequential format

One may argue that the break-even rule cannot be used as a standard evaluation procedure in competitions, as the season proceeds, because the necessary information is only available once the competition ends. This is only partly true, as one can give this rule a dynamic format that gets updated as the competition unfolds, as with all standard classification procedures. Let us briefly address this point here.

Suppose that the competition consists of a single round, to simplify the discussion. Each team must play $(m-1)$ matches in a certain order, which implies that there must be $(m-1)$ "days" of competition and each of those "days" will involve $(m-1)$ pairwise encounters.

Before the competition begins, we take a reference matrix $\mathbf{P}^0$ given by:

$$p_{ij}^0 = (m-1), \forall\, i \neq j, \;\; p_{ii}^0 = 0, \forall\, i$$

Those initial values p_{ij}^0 correspond to what the teams would get if they had ended the season with a draw in each match, under the rules of European football (one point per match). Needless to say, all teams obtain the same evaluation at this stage.

After the first day of the competition, we build a new matrix $\mathbf{P}^1$ as follows:

$$p_{ij}^1 = \begin{cases} p_{ij}, if\ i\ played\ against\ j \\ (m-2)\ , otherwise \end{cases} \qquad , \forall\, i \neq j$$

With $p_{ii}^1 = 0, \forall\, i$. Applying the break-even rule to this matrix provides an evaluation of the teams after the first day.

We proceed along this path as the season goes on. After t days of competition, we have a new matrix, $\mathbf{P}^t$, given by:

$$p_{ij}^t = \begin{cases} p_{ij}, if\ i\ played\ against\ j\ on\ day\ t \\ p_{ij}^{t-1}, if\ i\ had\ played\ against\ j\ in\ a\ former\ day \\ (m-t-1)\ , otherwise \end{cases} \qquad , \forall\, i \neq j$$

With $p_{ii}^t = 0, \forall\, i$.

After completing the season, where each team had played $(m-1)$ matches, we would have: $\mathbf{P}^{m-1} = \mathbf{P}$.

This construction allows an evaluation of the teams to be obtained and which is updated as the competition develops. Note how the break-even rule produces an evaluation in which the outcome of each new pairwise encounter affects the evaluation of *all* teams, as each new match modifies the outcome matrix $\mathbf{P}^t$ and hence the associated solution.

3.2 Outcome variables

Most sports apply a lexicographic criterion to determine the competition results: winning or losing the matches is the key criterion, with the specific results of the matches used only as a secondary principle, if at all (e.g., football, basketball). So, we typically find that winning a match gives the winner some points, a draw gives a smaller number of points (when draws are allowed), and losing gives the loser no points. All those points are added along with the competition and the final classification is determined by those sums, while the specific results achieved in the matches might be used to undo the ties.

Following this criterion, let us assume that each $p_{ij}, i \neq j$, corresponds to the result of applying a scheme by which the team that wins the match gets $\alpha > 0$ points, the team that loses the match gets 0 points, and both teams get β points, with $0 < \beta < \alpha$, when there is a draw. So here we have $p_{ij}(h) = \alpha$ with $p_{ji}(h) = 0$ if i beats j in round h, $p_{ij}(h) = 0$ with $p_{ji}(h) = \alpha$ when i is beaten by j, and $p_{ij}(h) = p_{ji}(h) = \beta$ in case of a draw. Common schemes of this type are: (1, ½, 0), as in Keener's baseline model, (3, 1, 0) as in The Premier League and most European football competitions, or (4, 2, 0) as in rugby contests.

Let w_{ij} denote the number of victories when i plays against j, in all the rounds, and d_{ij} the corresponding number of draws. Then, we can write $p_{ij} = \left(w_{ij}\alpha + d_{ij}\beta\right)$ and the break-even rule adopts the form:

$$f_i^{BE}(\mathbf{P}) = \frac{\sum_{j\neq i}\left(w_{ij}\alpha + d_{ij}\beta\right) f_j^{BE}(\mathbf{P})}{\sum_{j\neq i}\left(w_{ji}\alpha + d_{ij}\beta\right)}\ , \qquad i, j = 1, 2, \dots, m \qquad [2']$$

So that when $d_{ij} = 0$ for all $j \neq i$ (no draws), we get:

$$f_i^{BE}(\mathbf{P}) \ = \frac{\sum_{j\neq i} w_{ij} f_j^{BE}(\mathbf{P})}{\sum_{j\neq i} w_{ji}}\ , \qquad i, j = 1, 2, \dots, m$$

That is the number of victories weighted by the strength of the corresponding teams, over the number of defeats.

One can think of different ways of introducing the results of the matches in the evaluation (see Keener, 1993, Redmond, 2003, Dabadghao & Vaziri, 2021). A simple way of doing that, while preserving the lexicographic nature of victories, is the following.[3] Let c_i, c_j denote the "goals" obtained by teams *i* and *j* in a pairwise encounter. Then define:

$$p_{ij} = w_{ij}\left(\alpha + h\left(c_i, c_j\right)\right) + d_{ij}\beta$$

Where $h\left(c_i, c_j\right)$ is a function increasing in c_i and decreasing in c_j. An obvious example of this function is $h\left(c_i, c_j\right) = max\left\{0,\ \tau \times \left(c_i - c_j - k\right)\right\}$ for some scalars $\tau, k > 0$, that regulate the number of points that will be added, depending on the score difference. In the case of football competitions, for instance, we can think of $k = 1, \tau = 1/2$ so that the winning team gets an extra point when there is a difference of three goals, and proportionally more or less depending on that difference (with no extra point when there is a single goal of difference).

3.3 The reducibility problem

Note that $p_{ij} = 0,\ \forall j$ implies $f_i^{BE}(\mathbf{P}) = 0$. Yet, the converse is not true, that is, some teams can be given a zero rating and still get $p_{ij} > 0$ for some *j*. This will happen when the matrix **P** is reducible, which implies the existence of different subsystems that are globally ranked (i.e., all teams in the top subsystem are better than those in a different subsystem because no team of the lower subsystem has been able to beat or draw with a team from the top one). The dominant eigenvector of a decomposable matrix of this type gives value zero to all teams in the dominated subsystem. Therefore, all teams in the inferior categories are equally valued and rated zero in the overall comparison.

The reducibility of the outcome matrix has been regarded as problematic by many authors, as mentioned above because it implies making all those teams indistinguishable in the dominated subsystem and induces that winning or not be irrelevant for those teams regarding the final evaluation. Moreover, those procedures based on solving linear equation systems must ensure the existence of the inverse matrix to get unique solutions. Yet, in our view, reducibility is a structural property of the system that provides relevant information on the competition, which can be easily handled. Indeed, when we have a reducible matrix, we can apply the same evaluation procedure to a dominated subsystem, considered in isolation, and obtain the evaluation of those teams *within* that group. The teams of an inferior league will typically get non-zero ratings, as we may find values $p_{ij} > 0$ for some *j* within its category.

---

[3] This type of scheme is applied in thee Six Nations rugby tournament, a yearly international competition involving the best European rugby teams: England, France, Ireland, Italy, Scotland, and Wales. It adopts the form of a league, that is, each team competes against all others in pairwise encounters. Since 2017 the scoring system for the matches is the following: a victory yields 4 points, a draw 2 points, and a defeat 0 points. There are also some bonus points, to make the competition livelier. There is an extra (offensive) point awarded when more than 4 essays are realized in the match, and another extra (defensive) point if the defeat is by 7 or less points.

When there is a subset of dominated teams, therefore, we can proceed to rate them in two different groups. First, those teams that in the joint evaluation appear with positive ratings, which would define the top class. Then, the teams that appear initially with zeroes in the dominant eigenvector of **P'** would define the bottom class, in the understanding that all teams in the top class are regarded as better than anyone in the bottom class.[4]

An extreme case of reducibility appears when $\sum_{j=1}^{m} p_{ji} = 0$, which will be observed when team *i* wins all matches. In this case, the break-even rule is not defined for *i* and this indicates that there is a single team that is above all others and stands on its own in a different category. So, it is not meaningful to attach any relative value to that team, or one can say that its relative value is +∞, as a way of expressing the notion of *hors catégorie*. The evaluation can be applied to the remaining teams, in the understanding that they all are below *i*.

Note that many sports competitions define different divisions so that only the teams that belong to the same division play against each other throughout the season. A specific protocol determines how to move up and down between those divisions. This setting can be regarded as an application of the notions of computing the strength of the competitors and using a form of reducibility. In this case, the notion of strength is introduced implicitly in a very elementary way, which determines that only teams with similar strengths (those in the same division) play against each other. The configuration of the different divisions can also be regarded as the application of a form of reducibility since all teams playing in a higher division are regarded as better than those playing in a lower one, no matter their relative outcomes.

3.4 Comparison with other rules

We now compare the break-even rule with two rules with similar features, the Elo rule, and Keener's evaluation protocol.

There are some common features worth mentioning between the break-even rule and the Elo rating system. Elo's rating is a dynamic process that keeps changing the teams' evaluations as they compete against each other. The key idea is that if team *i* performs as expected against team *j*, it gains nothing, whereas if it performs better (resp. worse) than expected, it is rewarded (resp. penalized). This principle induces an adjustment process in which ratings are modified after each match, proportionally to the difference between the actual and the expected score. That is, when team *i* confronts team *j* at time *t*, the outcome will affect *i*'s rating as follows:

$$r_i^t - r_i^{t-1} = K\big(s_{ij} - E(s_{ij})\big)$$

---

[4] It may be that the subset of alternatives that are not in the top class define in turn a decomposable submatrix, which indicates that there are more than two "categories".

where s_{ij} is the outcome of the match (typically 1 for a victory, 0 for a defeat, and ½ for a draw), and $E(s_{ij})$ the expected outcome, which is, in turn, a function of the teams' rankings.[5] Note that the adjustment process depends on two constants that are to be determined from the outside. Also observe that none of the evaluations of the rest of the teams will be affected by the outcome of the match between *i* and *j* (hence, avoiding the competition may be a dominant strategy for both teams in some cases).

We have already shown that the break-even rule can be given a sequential format, which facilitates the comparison with Elo's rule. From this perspective, the break-even rule exhibits an updating procedure such that the outcome of each new round of matches affects the evaluation of *all* teams, as each new match modifies the outcome matrix $\mathbf{P}^t$ and hence the associated solution. Yet, we can also think of the break-even rule as applying a variant of Elo's principle, in the following sense. The default value attached to each team when meeting another is a draw (1 point, say). Then, if all teams draw in their corresponding matches on day t, we would have $f^{BE}(\mathbf{P}^t) = f^{BE}(\mathbf{P}^{t-1})$, whereas those teams that win their matches will improve their evaluations.

The break-even rule can also be regarded as the application of Keener's proportionality principle to a modified outcome matrix, whose elements are the relative outcomes, defined as follows:

$$r_{ij} = \frac{p_{ij}}{\sum_{j=1}^{m} p_{ji}}$$

Let us call $\mathbf{R} = (r_{ij})_{i,j=1}^{m}$ to the matrix of relative outcomes. It is easy to check that the dominant eigenvector of matrix **R** yields an evaluation that satisfies a slight variant of the break-even condition, given by: $f_i(\mathbf{P}) \sum_{j\neq i} p_{ji} = \lambda \sum_{j\neq i} p_{ij} f_j(\mathbf{P})$, where λ is the dominant eigenvalue of matrix **R**. Therefore, we can interpret that what the break-even rule does is incorporate the role of the losses to Keener's strength proportionality principle.

When we compare teams or athletes in terms of the ordinal evaluations provided by a collection of judges, the terms p_{ij} would correspond to the number of judges who prefer *i* to *j* (plus ½ of those who are indifferent, if we admit weak orderings). From this viewpoint the break-even rule can be regarded as an extension of the Borda-Condorcet rule (Herrero & Villar, 2021), dispensing with the transitivity of the judges' evaluations.

Finally, one may also wonder whether the break-even rule satisfies the three criteria proposed by Vaziri *et al* (2018), referred to in the Introduction. In the context of symmetric round-robin tournaments, this rule satisfies fairness (i.e., it is independent of

---

[5] The expected score is calculated by (where ζ is a constant):

$$E(s_{ij}) = \frac{10^{r_i/\zeta}}{10^{r_i/\zeta} + 10^{r_j/\zeta}}$$

Note that this expected score can be interpreted as a measure of the strength of team *i* relative to team *j*.

the sequence of the matches). It also satisfies comprehensiveness (i.e., it is sensitive to the strength of the competitors) and weak monotonicity (no team will find it beneficial to lose a match or to score fewer points).

3.5 Wrapping up

The break-even rule is a meaningful evaluation criterion that can be computed using standard algorithms and can also be applied sequentially. Moreover, this is a rule that enhances competitiveness in sports and discriminates more between the alternatives than the standard methods. This may be relevant for several reasons: To have a better picture of alternatives that may appear close together; to provide different shadow prices for the alternatives, depending on the features regarded as relevant; to define a range of values for those alternatives, that provide a robustness check of their evaluation; or to provide incentives to foster competition between contenders.

Let us conclude by noting that there are always concerns about the possibility of manipulating the results of sports competitions. Recently, the United Nations Office of Drugs and Crime, together with the International Olympic Committee, published a guide on tackling such manipulation (UNODC, 2021). Different rules exhibit different degrees of manipulability, in the sense that simpler rules facilitate computing what to do to get a given result. For instance, it has been proved that using the conventional evaluation, $\sum_{j=1}^{m} p_{ij}$, manipulation can be computed in polynomial time (Russell & Walsh, 2009). Increasing the complexity of the rules may help reduce manipulation, as it is more difficult to anticipate the precise implications of individual actions on the evaluation (Russell & van Beek, 2012). In this respect, therefore, introducing losses and strength in the evaluation will render manipulation more difficult.

REFERENCES

Albarrán. P., Herrero, C., Ruiz-Castillo, J, Villar, A. (2017). The Herrero-Villar approach to citation impact. **Journal of Informetrics**, 11: 625-640.

Berman, A. & Plemmons, R. J. (1994), **Nonnegative matrices in the mathematical sciences**. Society for Industrial and Applied Mathematics.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. **Computer Networks and ISND systems,** 30: 107-117

Boccard, N. (2022), Ranking athletes and teams across epochs (looking for the G.O.A.T.) SSRN: https://ssrn.com/abstract=2945389 or http://dx.doi.org/10.2139/ssrn.2945389

Colley, W.N. (2002), *Colley's Bias Free College Football Ranking Method: The Colley Matrix Explained* [online] http://www.colleyrankings.com/matrate.pdf .

Daniels, H. E. (1969), Round-robin Tournament Scores, **Biometrika**, 56: 295-299.

Elo, A. E. (1978), The rating of chessplayers, past and present. Arco Pub.

Govan AY (2008) Ranking theory with application to popular sports. **ProQuest**

Govan AY, Langville AN, Meyer CD (2009) Offense-defense approach to ranking team sports, **Journal of Quantitative Analysis in Sports**, 5(1). https://doi.org/10.2202/1559-0410.1151

Herrero, C. & Villar, A. (2021), Group decisions from individual rankings: The Borda-Condorcet rule, **European Journal of Operational Research**, 291: 757-765.

Keener, JP. (1993), The Perron-Frobenius Theorem and the ranking of football teams, **SIAM Review**, 35 (1): 80-93 .https://doi.org/10.1137/1035004

Laslier, J. (1997), **Tournament solutions and majority voting**. Springer, Berlin, Heidelberg, New York.

Liebowitz, S. J., Palmer, J. P. (1984), Assessing the relative impacts of economics journals, **Journal of Economic Literature**, 22: 77- 88.

Massey, K. (1997), **Statistical Models Applied to the Rating of Sports Teams**, Bachelor's thesis, Bluefield College, Bluefield, VA.

Moon, J.W., N.J. Pullman, (1970), On Generalized Tournament Matrices, **SIAM Review**, 12(3): 384-399.

Palacios-Huerta, I., O. Volij (2004), The Measurement of Intellectual Influence, **Econometrica**, 72(3): 963-977.

Pinsky, G., F. Narin (1976), Citation Influence for Journal Aggregates of Scientific Publications: Theory, with applications to Physics, **Information Processing, and Management**, 12: 297-312.

Russell, T., & van Beek, P. (2012). Detecting Manipulation in Cup and Round Robin Sports Competitions. *2012 IEEEE 24th International Conference on Tools with Artificial Intelligence*, (pp. 285-290).

Russell, T., & Walsh, T. (2009). Manipulating tournaments in cup and round-robin competitions. **First International Conference on Algorithmic Decision Theory**, (pp. 26-37).

Saaty, T.H. (2003), Decision-making with the AHP: Why is the principal eigenvector necessary, **European Journal of Operational Research**, 145: 85-91.

Slutzski, G, O. Volij, (2006), Scoring of web pages and tournaments: axiomatizations, **Social Choice and Welfare**, 26 (1): 75-92.

UNODC. (2021). **Legal approaches to tackling the manipulation of sports competition: a resource guide**. Viena: United Nations.

Vaziri, B., Dabadghao, S., Yih, Y., & Morin, T. L. (2018), Properties of sports ranking methods, **Journal of the Operational Research Society**, 69(5), 776-787. https://doi.org/10.1057/s41274-017-0266-8

Vaziri B, Yih Y, Morin T (2018) A proposed voting scheme to reduce the sensitivity of the Markov method, **International Journal of Operations Research,** 32(1):24–40

Acknowledgments

The first author acknowledges financial support from the Spanish Ministerio de Ciencia e Innovación, (project PID2019-107081GBI00) and the Generalitat Valenciana (Prometeo 2021/073).

The second author acknowledges financial support from the Spanish Ministerio de Ciencia e Innovación (project PID2019-104452RB-I00) and the Junta de Andalucía (projects FEDER UPO-1263503 and P18-RT-2135).