Physica A 392 (2013) 5376–5386

Contents lists available at ScienceDirect

Physica A

journal homepage: www.elsevier.com/locate/physa

Hidden power law patterns in the top European football leagues

Sergio Da Silva [a,*], Raul Matsushita [b], Eliza Silveira [a]

[a] Department of Economics, Federal University of Santa Catarina, Florianopolis SC, 88049-970, Brazil
[b] Department of Statistics, University of Brasilia, Brasilia DF, 70910-900, Brazil

HIGHLIGHTS

- We suggest football (soccer) follows power law progress curves.
- This is shown for data collected from the top fifteen European football leagues for the clubs participating in the latest seasons.
- We adjust power laws showing a close relationship between rank and points won by the clubs.
- We also propose the use of Shannon entropy for ranking the leagues according to their competitiveness.

ARTICLE INFO

Article history:
Received 28 December 2012
Received in revised form 23 May 2013
Available online 18 July 2013

Keywords:
Power laws
Wars
Sports
Terrorism
Football
Soccer

ABSTRACT

Because sports are stylized combat, sports may follow power laws similar to those found for wars, individual clashes, and acts of terrorism. We show this fact for football (soccer) by adjusting power laws that show a close relationship between rank and points won by the clubs participating in the latest seasons of the top fifteen European football leagues. In addition, we use Shannon entropy for gauging league competitive balance. As a result, we are able to rank the leagues according to competitiveness.

1. Introduction

Sports are, arguably, stylized combat because they are contests of aiming, chasing, or fighting, complete with victors and the vanquished [1]. But they transcend that. Because sports are competition without killing, they fulfill a positive role of assuaging bitterness, seeking reconciliation, attempting conciliation, and pursuing courtesy. In this sense, sports are a useful substitute for war. They can divert nations from military aggression because of the heightened national self-esteem due to sporting success [2]. Whatever the deepest meaning of sports, however, their mechanics undeniably mimics that of war. What is more, while wars do not conserve the number of participants, sports do. This ultimately makes the analogy between wars and sports non trivial in that sports provide an ideal laboratory for studying competitions [3]. Unlike wars, the records of sports events are accurate, complete, and widely available [4].

Here, we are particularly interested in the mechanics of football (soccer), the number one sport. Apart from the United States and a handful of countries, soccer is the most popular game on Earth. We take data from the latest seasons of the top fifteen European football leagues, show similarities with the formal results already found for war in the literature, and present novel results.

* Corresponding author. Tel.: +55 48 8804 7489.
E-mail addresses: drsergiodasilva@hotmail.com, professorsergiodasilva@gmail.com (S. Da Silva).

0378-4371/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physa.2013.07.008

Earlier studies on war [5] found that wars with low death tolls far outnumber high fatality conflicts. Although this seems to be highly expected, what is not obvious is the finding that the number of wars with a given number of fatalities follows an approximate power law statistical distribution as a function of the number of fatalities. The link between the severity and frequency of conflicts follows a power law. This implies that extreme events such as the world wars cannot be considered as anomalies; they are expected to occur occasionally, given the frequency with which conflicts take place. Such power law also applies to individual clashes and acts of terrorism [6–8], and even to scientists' career progress [9]. The possible explanation for the phenomenon is the fact that, for a wide range of human activities, the time taken to complete a given challenging task decreases with successive repetitions, following an approximate power law progress curve [8]. We start our analysis by suggesting that the same fact extends to football, given that it is stylized combat.

The rest of this article is organized as follows. Section 2 describes the data used, Section 3 carries out the analysis, Section 4 discusses the results, and Section 4 concludes the study.

2. Description of data

We collect data for the top fifteen European football leagues for the clubs participating in the latest seasons. The leagues we consider are the English, Italian, Spanish, German, Dutch, Portuguese, French, Scottish, Greek, Turkish, Belgian, Austrian, Danish, Polish, and Irish. The data source was. Table 1 shows every league considered, the season for which data are available, total number of clubs participating in one league, one league's rounds per season, the number of datapoints per season, and total number of datapoints.

At the end of each season j of league k, we recover from the data the total points won by each team i and denote these by p_{ikj}. Then, we rank them in decreasing order, $p_{ik(1)} \geq p_{ik(2)} \geq \cdots \geq p_{ik(n)}$, where n is the number of clubs. For further convenience we denote $p_r \equiv p_{ik(r)}$, where r is the rank. Actually, we first take the fraction of points won. Then, using the percentages of wins, draws and upsets we get the total points won through linear transformation. The win percentage is arguably a better measure of team strength than the points won are, because there are small variations between the numbers of matches played by various clubs across the leagues [4].

Fig. 1 shows the outcomes of the top fifteen European football leagues for the 2011/12 season. There is a clear pattern of green squares (meaning matches won) for the clubs that ended up on top. And there is still a hidden power law pattern, as we will show next.

3. Analysis

In a league format of competition, every club plays every other club. In particular, each club hosts every other club exactly once during a season. Of course, the outcome of a match is subject to location, weather, injuries, red cards in the previous match, and a multitude of other factors. These make the outcome of a single game unpredictable to some degree, and, thus, this is so for the championship outcome. Thus, randomness is inherent to such a form of competition [4].

For this reason, we first take the normalized distributions of the points won p_r^* by the clubs participating in the leagues:

$$p_r^* = \frac{p_r}{\sum\limits_{r=1}^{N} p_r}, \tag{1}$$

where N stands for the total number of clubs in one league, and $\sum_{r=1}^{N} p_r$ is the actual points won. Thus, we have

$$\sum_{r=1}^{N} p_r^* = 1, \tag{2}$$

so that Eq. (1) represents a probability distribution.

Second, we conjecture that there is a law describing a close relationship between rank r and the normalized points won p_r^* as a power law of the form

$$p_r^* = a \cdot r^b, \tag{3}$$

where p_r^* changes as if it were a power of r. The problem is then to verify the conjecture and determine a and b. The coefficient a is the normalizing constant, and considering (3) and (2), we have

$$a^{-1} = \sum_{r=1}^{N} r^b. \tag{4}$$

Rather than taking logs on both sides of (3), in order to satisfy (2) and (4) we directly carry out a fit of (3) using nonlinear regression.

Fig. 2 displays the empirical values of p_r^* along with the fitted power laws for the leagues. The coefficients a and b were found by running nonlinear regressions on (3) for each league (Table 2). As p_r^* represents a probability distribution we can