Journal of Industry, Competition and Trade, 3:3, 167–186, 2003

Equality of Opportunity and Equality of Outcome: Open Leagues, Closed Leagues and Competitive Balance*

LUIGI BUZZACCHI luigi.buzzacchi@polito.it
Politecnico di Torino

STEFAN SZYMANSKI szy@ic.ac.uk
Imperial College London

TOMMASO M. VALLETTI† t.valletti@ic.ac.uk
The Business School, Imperial College London, 53 Prince's Gate, Exhibition Road, London SW7 2PG, UK

Abstract. This paper compares conventional static measures of competitive balance with measures that take account of the mobility of teams into the upper ranks of professional leagues, which we call dynamic competitive balance. We use this measure to compare the open soccer leagues that permit entry by the process of promotion and relegation, to the closed leagues of North America where there is no automatic right of entry. We also identify the theoretical distribution of entrants to the top k ranks assuming that all teams have equal probabilities of winning. We find that the open leagues (OL) we study are less balanced, dynamically, than closed leagues (CL), and also that OL lie much further away from the theoretical distribution than CL.

Keywords: sports leagues, competitive balance

1. Introduction

> ''All schemes used in the United States punish excellence in one way or another. The European football approach punishes failure by promoting excellent minor league teams to the majors and demoting (relegating) poor performing major league teams back down to the minors. The revenue loss from a potential demotion to a lower class of play is severe punishment for low quality—severe enough that salary treaties, league sharing arrangements, and unified player drafts are so far thought to be unnecessary, even though star salaries are enormous. It is an interesting economic question as to which system achieves better results.''
>
> Rosen and Sanderson (2001)

Since Superbowl I in 1967 the NFL has expanded from a league of 16 teams to one of 32. During this period 17 different franchises have won the Superbowl. In Serie A, the top division of Italian soccer, 48 different teams have participated since 1967, but there have

* We thank the Editor and an anonymous referee for helpful comments.
† Corresponding author.

been only 11 different winners.[1] Serie A has had more teams not because it is a larger league—in most seasons only 18 teams compete for the championship title. However, the institution of promotion and relegation permits new teams to enter the league each year. Nonetheless, despite having more competitors, fewer teams seem to have a chance of winning—less than quarter of the teams in Serie A over the period have one, compared to half of the current NFL franchises. Moreover, a similar story emerges if any of the North American leagues are compared to the national soccer leagues of Europe.[2] In other words, soccer leagues tend to be much less balanced than the major leagues. However, this is not true when measured in the way that has been conventionally adopted in the sports literature. This paper proposes a way of measuring competitive balance that permits comparison between the North American closed leagues (CL) and the open leagues (OL) of Europe.

This enables us to address some important policy issues. It is a long-standing proposition in sports economics that ''better'' means ''more balanced'' results (*e.g.* Rottenberg, 1956). Competitive balance refers to the expectations of fans about who will be the winner. In a perfectly balanced contest fans believe all outcomes are equally possible so there is complete outcome uncertainty. In a perfectly unbalanced contest the winner is known *ex ante* with probability 1. It seems reasonable to suppose that without at least some degree of competitive balance, fans will loose interest in a competition. A stronger proposition would be that, all else equal, a more balanced contest is a more interesting one. This proposition is widely accepted (see Szymanski, 2003 for a survey of the evidence). Most importantly, it has frequently been accepted by the courts as a justification for restrictive agreements concerning the sharing of revenues or restraints in the player market (*e.g.* salary caps and roster limits).

Measuring the extent to which a competition is balanced, therefore, is a critical issue. However, this is no simple issue. The measure widely cited in the existing literature (see, *e.g.* Quirk and Fort, 1992) is the standard deviation of winning percentages in a season. According to this measure, the greater the variance in outcomes in a season, the less balanced is the contest. The principal weakness of this measure is that it takes no account of the identity of the teams across seasons. Applying this measure to North American and European leagues produces the seemingly perverse result that European leagues are if anything slightly more balanced than North American leagues. This is an artifact of the European system of promotion and relegation, by which the worst performing teams in a league in each season are demoted to the immediately junior league—so that no team can afford to give up once it is out of contention for the title.

The novel measure of competitive balance that we develop in this paper permits a comparison to be made between different sports leagues across time, in particular taking the account of the promotion and relegation system. This measure focuses on the identity of teams, and the frequency with which they approach the possibility of winning the title.

1 In fact, going back to 1945 only 12 different teams have won Serie A.

2 Soccer is by far the dominant team sport in Europe. Sports such as Rugby, cricket, basketball and ice hockey trail a long way behind in terms of popularity.

What we show is that the North American major leagues (we consider Major League Baseball, the National Football League and the National Hockey League) are much more balanced than the dominant soccer leagues of Europe (we compare the top divisions of the English, Italian and Belgian leagues) in the sense that a greater proportion of teams in the league are likely to experience any given level of success within a given period of time. However, the North American leagues are more or less closed to entry by new teams, whereas European leagues have a system of promotion and relegation which gives many more teams access to the highest level of play (the top division). In other words, we find equality of outcomes in the closed North American leagues while in Europe, we find equality of opportunity within a system that is dominated by a small number of teams.

Previous studies of competitive balance have measured either ''match uncertainty'' or ''seasonal uncertainty'', and both of these measures can be compared across leagues. Match uncertainty refers simply to expectations about a particular game, and this can be measured, for example, by studying pre-match betting odds.[3] However, interest in a league competition goes beyond uncertainty about a particular match and many fans are attracted by uncertainty about the overall outcome of the championship. Seasonal uncertainty, the closeness of an overall championship race, can be measured in a number of ways, *e.g.* the number of games behind the leader that the following k teams are as the season ends, the date at which at the championship outcome becomes known with certainty, or more generally, the standard deviation of success (*e.g.* win percentage) among the teams.

These measures tell us something about competitive balance in a ''static'' sense—the balance of a particular match or season, but for many fans there is also interest in competitive balance in a ''dynamic'' sense: do particular teams dominate the championship over time? This paper proposes a natural basis for comparing dynamic competitive balance among leagues and a theoretical benchmark.[4] The natural basis is to consider the cumulative frequency of teams entering the top k ranks of a league (ranked by some measure such as win percentage). By analogy we can think of dynamic competitive balance like the spread of an epidemic—the more balanced a league the more rapidly teams enter the ranks of the top k.

This raises an important additional consideration. In a ''closed'' league of, say, 20 teams, the top k can only ever be drawn from that 20 teams.[5] The more teams, the greater the potential for entry into the top k, just as the absolute number of people succumbing to a disease must be increasing in the total population. Leagues in North America are typically closed in the sense used above (although there can be some new entry from the sale of additional franchises and league mergers). By contrast, in most other countries leagues are ''open'' in the sense that at the end of each season the worst performing

3 A review of these studies is provided by Szymanski and Kuypers (1999).

4 Humphreys (2002) and Koop (2002) have proposed alternative methods for evaluating the evolution of competitive balance over time, but in both the cases the method is not naturally adapted to inter league comparisons.

5 Abstracting, for the moment from franchise expansion.

teams are demoted to the immediately junior league and replaced by the best performing teams from that league. The European Commission (1998) has gone so far as to suggest that ''the system of promotion and relegation is one of the key features of the European model of sport''. Given this hierarchical structure it is apparent that the population of potential entrants into the major league is, over time, much larger (and possibly unlimited) compared to a closed league.

One central concern of this paper is whether this ''equality of opportunity'' translates into equality of outcome, as measured by competitive balance. Our theoretical benchmark approaches this issue by asking how many teams would be expected to enter the top k ranks in a league system where resources were so distributed that each match played were perfectly balanced. In other words, if success were purely random (because all sources of systematic variation, such as resource inequality, had been removed) how much mobility would there be in a given league structure? Calculating this benchmark for a CL with a fixed number of teams is relatively straightforward, but as we show the case of an OL is more complex. However, by deriving these values we can compare actual with theoretical mobility and derive a kind of Gini coefficient, measuring the closeness of league outcomes to a perfectly balanced ideal. It turns out that North American leagues are far closer to the theoretical ideal than their European counterparts, a phenomenon that can be accounted for by the much greater extent of resource equalization measures in North America, *e.g.* gate revenue sharing, collective merchandising, draft rules, salary caps and so on.

The paper is set out as follows. In Section 2, we survey the literature comparing North American leagues and European soccer leagues. In Section 3, we compare measures of static and dynamic competitive balance for North American major league sports and the dominant national sports leagues in Europe for soccer. In Section 4, we derive our theoretical benchmark for CL and OLs. In Section 5, we consider the difference in mobility comparing theoretical and actual measures. Section 6 concludes.

2. North American and European sports leagues

As the quote at the beginning of this paper indicates, there are substantial differences in the organizational framework of North American and European sports leagues and many of these are illustrated in Table 1.[6] First, in North America, there are four significant team sports competing for market share compared to a single dominant team sport in Europe-soccer. However, in North America, there is a single league that dominates competition in each sport, whereas in Europe the national league of each country is normally dominant in its own territory, with limited penetration elsewhere (in some ways the regional conferences of the NCAA bear some resemblance to this). Four leagues currently dominate Europe due to their large populations, wealth and traditional obsession with

6 Adapted from Hoehn and Szymanski (1999) who review the differences between North American and European soccer.

Table 1. Differences in structure of U.S. and European sports leagues.

	U.S. Sports	Football in Europe
League system	Closed, no promotion or relegation	Open, annual promotion and relegation
	Teams compete in single league competition	Teams may compete simultaneously in many competitions
League functions	Collective sale of TV rights Centralized marketing	Collective sale of TV rights
Competition between clubs	Limited substitution by consumers	Significant potential for substitution
Competition between leagues	Numerous cases of entry by rival leagues	All leagues contained within the established hierarchy
Player market	Rookie draft Salary caps (NFL, NBA) Collective bargaining	Active transfer market
Revenue sharing	Equal division of national broadcast income	Sharing of television income
	Gate sharing (NFL 40%, Baseball average 15%, NBA 0%)	Little or no sharing of league gate revenues Some sharing of gate from cup competitions
Competition policy	Antitrust exemption for baseball	Centralized sale of TV rights under attack
	Sports Broadcasting Act exempts national TV deals from antitrust	Selected interventions (ticket allocation FIFA)

soccer: England, Germany, Italy and Spain. Another significant difference is that North American teams and players typically play in only one league or competition in any one season. In Europe however, teams can compete in two national competitions (the league and the knock-out Cup[7]), an international league or cup competition for European clubs, while in addition the players can be selected to represent the national team[8]—and national representative competition is in most cases even more popular than club competition.

Economic arrangements in North American and European team sports also differ significantly. On the revenue side there is a much lower incidence of revenue sharing in Europe. On the expenditure side there are almost no restrictions in the player market and trading for cash is the accepted norm (more than 10% of players move team each season) compared to the U.S. where rules and custom inhibit player mobility.

The principal difference that this paper focuses on is the existence of promotion and relegation. In Europe leagues are typically organized in an ascending hierarchy by a governing body vested with responsibility for the development of sport as a whole. At the end of each season the worst performing teams in each division are demoted to the

7 Playoff are not widely used. A league competition followed by playoffs can be thought of as an integrated League and Cup competition.

8 Clubs are obliged as part of their obligations to their national federation to release their players and receive no compensation. Secession from the federations is seldom considered an option. In France, for example, the national federation has a statutory monopoly over the organization of league competition.

immediately junior division and replaced by the best performing teams in the junior division. Hence, it is in principle possible for team to rise from the lowest to the highest level of competition, and vice versa. The existence of this system in Europe and not in North America is essentially an accident of history (see Ross and Szymanski, 2001 for a description of the historical underpinnings), and while the discussion in this paper focuses primarily on soccer, this system has been adopted in even in sports imported into Europe from North America, such as basketball.

A further difference between North American leagues and European leagues has been the subject of some controversy, and this is the widely held belief in Europe that North American clubs are run by profit maximizers whereas European clubs embrace purely sporting objectives, such as win maximization subject to a budget constraint (the locus classicus is Sloane (1971)), and many of the implications of these assumptions for league rules have been analyzed by Kesenne (see, *e.g.* Kesenne, 2000). However, it is also clear that until recently there was limited financial return to be gained from team sports in Europe, a situation which has been transformed dramatically in the 1990s—and some authors see a shift toward more commercially oriented policies and the increasing pursuit of profit (see, *e.g.* Andreff and Staudohar, 2000).

While it is clear that there are many differences, the comparative study of North American and European leagues is still at an early stage.[9] In two recent papers, Noll (2002) and Ross and Szymanski (2002) consider what impact the promotion and relegation system would have on North American leagues if introduced there, but further research is required to understand the consequences of institutional differences. By focusing on competitive balance, this paper aims to provide a starting point for a theoretical analysis of the properties of these institutions.

3. Static and dynamic competitive balance

Most measures of competitive balance in the literature are essentially static—they analyze the equality of winning opportunities for individual matches or for a championship season taken as a whole.[10] For example, Quirk and Fort (1995) hypothesize that if seasonal win percentages become less dispersed then a league has become more competitively balanced. They review some allegedly balance-enhancing reforms in the North American professional leagues (*e.g.* the introduction of the salary cap in the NBA in 1984, the NFL rookie draft introduced in 1936 and the beginning of free agency in baseball in 1976). In general, they find no significant change in the standard deviation after the reforms and therefore conclude that there is no evidence that competitive balance was in fact enhanced.

9 Hall et al. (2002) attempt a limited comparison of the economics of Major League Baseball and English soccer.

10 An exception is Szymanski (2001) who compares the competitive balance of two different competitions.

Table 2. Actual standard deviation of win percentages divided by idealized standard deviation.

Decade	MLB (1950–99)	NFL (1950–99)	NHL (1949–98)	England (1949–98)	Italy (1949–98)	Belgium (1953–2000)
1950s	2.23	1.48	2.04	1.15	1.33	1.26
1960s	2.05	1.63	1.93	1.33	1.50	1.45
1970s	1.88	1.60	2.61	1.44	1.47	1.54
1980s	1.66	1.46	2.08	1.48	1.34	1.67
1990s	1.68	1.51	1.83	1.40	1.61	1.67

Note: Ties (draws) are treated as half a win. European leagues refer to the top division of the national soccer league.

Horowitz (1997) uses an entropy index to measure changes in competitive balance over time in Major League Baseball (and finds that there is underlying trend toward increasing balance over the period 1903–1995). Applying the concept of entropy to a CL seems natural enough, but it is less clear how one might extend this to a league with promotion and relegation. Quirk and Fort (1992) look at balance over time by adopting another measure based on seasonal variance. If a given league were perfectly balanced the winning probability for each team in each match would be 0.5, which would also be the expected value of the seasonal win percentage. The standard deviation of this win percentage would then be $0.5/\sqrt{m}$ where m is the number of matches played. This can be used as an ''idealized'' measure of the standard deviation for a particular league. Expressing the actual standard deviation as a ratio of the idealized standard deviation thus provides a basis for comparing the degree of competitive balance of different leagues.

In this paper, we have chosen to compare three North American leagues (Major League Baseball (MLB), The National Football League (NFL) and the National Hockey League (NHL)) to three national soccer leagues in Europe (Italy, England and Belgium). In terms of revenues and broadcast audiences the first two in each region are somewhat larger than third, but within both regions the basic league structures are comparable. Above all, the North American leagues are all closed and the European soccer leagues all open. In Table 2 we compute the ratio of actual to idealized standard deviation for North American major league sports and European soccer leagues. As Quirk and Fort have noted, the data for the North American leagues indicate a trend toward competitive balance over time in baseball, but no trend in the NFL or NHL. On this measure the NFL has tended to be the most balanced of the North American leagues and the NHL the least balanced. As has been noted in the work of Kipker (2000), the European soccer leagues seem by comparison to be more balanced (although the European trend is toward less balance). Thus, in the 1950s each of the three European leagues had lower standard deviations than any of the North American leagues, and in the case of England the actual standard deviation was only 15% higher than the idealized standard deviation. However, by the 1990s the gap had narrowed considerably, and for instance, the NFL seemed more balanced than either Italian of Belgian top divisions.

The principal weakness of a static measure of competitive balance such as this is that it takes no account of the identity of the successful teams.[11] So for example, according to the data the NFL was less balanced than the top soccer division in England, but in between 1990 and 1999 6 different teams won the Superbowl whereas in England only 5 different teams won the League Championship. Moreover, in England one team won the title on 5 occasions (Manchester United), whereas the biggest winner in the U.S. won only 3 times (Dallas).[12] To consider the dynamics more fully we have looked at the number of different teams winning the league title and the number of teams entering the top ranks.[13]

We conjecture that fans care about balance in the sense that they want a reasonable prospect that the identity of the winners will change from time to time (although they may also care about the variance of success among the teams within the season). ''Turbulence'' at the top increases the interest of fans of a greater number of teams. If, say, each team experiences diminishing returns to success in terms of fan interest, then a league that is more balanced in this dynamic sense will be more successful. This point has been made elsewhere, see *e.g.* Ross and Lucke (1997) and Szymanski (2001) who finds some empirical support for the conjecture. In this paper, our aim is to compare dynamic competitive balance across open and closed league. To do this, we have looked at the number of entrants into the ranks, first over the full 50 years of data, then over 40 years, 30 years, 20 years and 10 years. The data is reported in Tables 3 and 4.

Table 3 shows the number of different teams with the highest win percentage in each season. The North American leagues have developed the post season play-off season over the period in question in order to involve more teams in the championship race for longer. The play-off system introduces more randomness in outcomes and therefore adds to uncertainty even if the teams are not well balanced competitively. Since we are interested in competitive balance rather than uncertainty itself in this paper we have restricted ourselves to considering win percentages during the regular season only (see note to Table 3).

For the North American leagues the NFL had the greatest number of teams entering the top rank for 4 out of the 5 periods considered, although in the last decade the performance

11 This fact may also produce uninformative information within a given period. For instance the standard deviation of winning percentages tends to put too much weight among weaker teams. To give an example, imagine 10 teams competing in a league in a given year. In the first scenario, there is a team much stronger than everybody else that wins every single match, while the remaining teams have identical strength and win 50% of the matches. In the second scenario, there are 5 slightly stronger teams and 5 slightly weaker teams, where a team has 50% chance of winning a match among ''equals'', while a stronger team has 80% probability of winning against a weaker team. Despite it would be natural to describe as more balanced the second scenario, the normalized standard deviation of win percentages would yield the same numerical value in both cases.

12 Rolling the data two years further forward would highlight the point even more clearly: only 4 winning teams in England, one of which won 7 of the 10 titles, while in the NFL there were 7 different winners.

13 League rank is the standard measure of performance in Europe. For North American leagues we have ranked teams according to their regular season win percentage.

Table 3. Teams that had the highest winning percentage or were winners of the league championship.

Period	MLB	NFL	NHL	England	Italy	Belgium
50–99	16	20	13	16	12	10
60–99	16	18	13	13	12	8
70–99	14	14	12	9	10	8
80–99	12	9	10	7	7	7
90–99	6	7	7	5	4	4

Note: North American teams selected on the basis of regular season win percentage. European teams selected on the basis on actual championship wins. Traditionally, 2 points were awarded for win, 1 for a draw (tie). Tied winning percentages were then decided on goal difference. However, from the 1980s onward leagues introduced the award of 3 point for a win and 1 for a draw. In the data, the champions always had the highest win percentage, but in 9 out of the 150 championships considered the champions were tied in win percentage with the team ranked second (on goal difference). In 1995, Blackburn Rovers won the English championship on the basis of the new points system but would have tied on the old points system (which is the same as our measure of win percentage) and had an inferior goal difference to the team ranked second.

of all 3 leagues looks remarkably similar on this measure (notwithstanding the recent dominance of the New York Yankees in baseball).

For any of the 5 ranges considered, both the Italian and Belgian leagues had less variation in the number of teams appearing in this rank than any of the North American leagues. Only when the last 40 or 50 years are considered did the English league have as many teams entering the top rank, and for both of these ranges the NFL had more teams achieving the highest win percentage. Thus, despite the greater opportunity through promotion and relegation for teams to reach the highest rank, there seems to be relatively less turnover at the very top in open European leagues than in the closed North American leagues. On average over the last 30 years there have been 50% more teams achieving the highest rank in North America compared Europe.[14]

The story told by Table 4 is slightly different. Looking at the last 10 years, a very large fraction of all teams in the each of the North American leagues managed a top 5 finish in term of win percentage. Forty-six per cent more teams in the CLs achieved this feat on average than in the open European leagues. However, as we go back further the number of teams entering the top ranks does not increase significantly for the North American case, but does in the European case. Clearly, once all the members of a closed league have entered the top 5, the population of entrants can only increase through franchise expansion. In an OL, however, there need be no limit to the increase in the population. Thus, in the case of England there were only 16 teams entering the top 5 ranks over the last decade, but 34 teams entered over the entire 50 year period, more than for any league in the sample (during this period the top division was restricted to 22 or fewer members in each season). Over this lengthy period the number of teams entering the top 5 in North America and Europe is almost identical. Thus, openness in Europe seems to give roughly

14 In writing this paper we have looked at some of the descriptive statistics for other European leagues such as Germany, Portugal, Scotland and Spain. In all cases, a similar pattern emerges to that described here.

Table 4. Teams that entered the top 5 ranks.

Period	MLB	NFL	NHL	England	Italy	Belgium*
51–00	28	31	21	34	19	32
61–00	28	30	21	27	18	28
71–00	28	29	21	24	18	24
81–00	28	28	21	22	14	21
91–00	23	26	18	16	12	18

*1953–2000 only.
Note: all teams selected on the basis of regular season win percentage.

similar opportunities over a very long period time, even if there are fewer opportunities over relatively short timespans. Whether this is enough to make an open European league as competitively balanced as a closed North American league in the eyes of the fans must be doubtful at least.

In order to set this picture in a proper perspective, we now consider the theoretical probability of teams entering the top k of ranks for open and closed leagues, under the hypothesis that in each season each contestant in a division has an equal probability of winning each match.

4. Entry in the top k ranks

4.1. Closed leagues

We consider first the case of n teams that are grouped together and compete in a CL with no system of promotion and relegation. Under the hypothesis that the outcome of the championship in a given year is purely random, each team has the same probability $1/n$ of being ranked 1st, 2nd, ..., n-th. The probability that a team is ranked in the top k places in a generic year is then $w(k) = k/n$. After T years, the probability that a given team has been placed at least once in the top k places is $w(k,T) = 1 - (1 - w(k))^T = 1 - [(n-k)/n]^T$. Finally, the expected number of teams that has won one of the top k positions in the first T years is:

$$y^{\mathrm{CL}}(k,T) = nw(k,T) = n - \frac{(n-k)^T}{n^{T-1}} \tag{1}$$

Equation (1) represents our benchmark for a CL. In particular, the expected number of teams that has won the title at least once after T years is simply $y^{\mathrm{CL}}(1,T) = [n^T - (n-1)^T]/n^{T-1}$, increasing at a decreasing rate over time, from 1 when $T = 1$, to n when T tends to infinity.

4.2. Open leagues

We now consider the typical European way of organizing a team contest. There are L leagues ordered from the top division to the lowest one: league 1 is the ''premier'' league

that awards the championships while league L is the lowest league. League l consists of n_l teams, $l = 1, \ldots, L$. In a generic period, a team in league l can either remain in the same league, or go to an ''adjacent'' league. We denote respectively by $p(l)$ and by $r(l)$ the total number of promotions to the league above and the total number of teams relegated to the league below league l.[15] If the outcome of each league is random, the probability that a team is in division l at time t is:

$$d(l,t) = d(l,t-1)\frac{n_l - r(l) - p(l)}{n_l} + d(l-1,t-1)\frac{r(l-1)}{n_{l-1}} + d(l+1,t-1)\frac{p(l+1)}{n_{l+1}} \tag{2}$$

where $l = 1, \ldots, L$ and $r(L) = p(1) = 0$, $d(0, t) = d(L+1, t) = 0$.[16] It can be verified that

$$\sum_{l=1}^{L} d(l,t) = \sum_{l=1}^{L} d(l,t-1) = 1$$

since a team starts at $t = 0$ in some league with probability 1, *i.e.* $d(l, 0)$ is 1 for only one value of l and 0 otherwise. In order to take into account the initial distribution of teams, we denote with a subscript l the league where a team starts at the beginning, $d_l(j, 0) = 1$ if $j = l$ and 0 if $j \neq l$.

The probability that a team is ranked in one of the top k places of the premier division in a generic year t is given by the joint probability $d_l(1,t)k/n_1$. We are now in a position to calculate the probability that, after T years, a team that started in league l in the initial period $t = 0$ has been placed at least once in the top k places of the top division. This probability depends on initial conditions and corresponds to the complement to 1 of the probability that such team has never been placed in the top k positions, *i.e.* the team was either in a lower division or in the top one but never ''picked'' one of the top placements:

$$w_l(k,T) = 1 - \prod_{t=0}^{T}\left[\sum\nolimits_{l=2}^{L} d_l(l,t) + \frac{n_1 - k}{n_1} d_l(1,t)\right] = 1 - \prod_{t=0}^{T}\left[1 - \frac{d_l(1,t)k}{n_1}\right] \tag{3}$$

The expected number of teams that has been placed in the top k positions after T years is:

$$y^{\mathrm{OL}}(k,T) = \sum_{l=1}^{L} n_l w_l(k,T) \tag{4}$$

15 In principle, both $p(\,)$ and $r(\,)$ should depend on t; however, we can drop the dependency from time under the hypothesis of random ranking as long as the number of teams in a given league is constant over time. In practice, the number of promotions and relegations can change between periods and this feature can be easily accommodated in our framework.

16 To ensure that the number of teams in a given league is constant overtime, we assume $p(l) = r(l-1)$.

Equations (2), (3) and (4) represent the benchmark for an OL and it is the counterpart to Equation (1). Once it is known the number of teams in each league, as well as the number of teams promoted and relegated to adjacent leagues and the initial conditions, it is immediate to obtain the value of the expected number of teams observed in the top positions after T years. For instance, if a total number n of teams is split equally among L leagues, the teams are ordered in a way such that $d_1(1,0) = 1$ for the first group of n/L teams, $d_2(2,0) = 1$ for the second bunch of n/L teams and so on, and if 1 team is promoted and 1 team is relegated in any period, the expected number of teams that has won the premier league at least once after T years is given by[17]

$$y^{\mathrm{OL}}(1,T) = \left(\frac{n}{L}\right)\sum_{l=1}^{L}\left(1 - \prod_{t=0}^{T}[1 - d_l(1,t)L/n]\right)$$

$$d_l(1,t) = [d_l(1,t-1)(n/L-1) + d_l(2,t-1)]L/n$$

$$d_l(l,t) = [d_l(l,t-1)(n/L-2) + d_l(l-1,t-1) + d_l(l+1,t-1)]L/n$$

$$l = 2,3,\ldots,L-1$$

$$d_l(L,t) = [d_l(L,t-1)(n/L-1) + d_l(L-1,t-1)]L/n$$

5. Mobility in theory and in practice

In the previous sections, we have looked at the actual number of entrants into top ranks and derived the theoretical distributions of teams appearing in top positions in closed and open leagues under the assumption of equal winning probabilities. In this section, we compare the difference between the actual and theoretical distributions.

We apply the theory of the previous section to the precise structure of each league. Table 5 shows the number of teams that would have been expected to achieve the highest seasonal win percentage (if all teams had equal win probabilities) over the same periods considered in Table 3. Table 6 shows the theoretical prediction of entrants into the top 5 of win percentages (if all teams had equal win probabilities), analogous to the actual data of Table 4.[18]

In general, there are two conflicting effects that produce differences in the theoretical predictions for the closed North American and open European leagues. First, in recent decades the expansion of North American leagues to around 30 teams has increased the

17 With a simple spreadsheet it is immediate to confirm that with the same total number of teams $y^{\mathrm{OL}}(1,T) < y^{\mathrm{CL}}(1,T)$ for any $T > 1$. The difference between the two expected numbers of winning teams becomes smaller as T grows, or if ''turbulence'' is increased by increasing the number of teams promoted/relegated in any period.

18 Numbers in Tables 4 and 5 are rounded. They were obtained taking into account entry and exit of teams, as well as variations in the number of promotions and relegations over time. For instance, in England the third division was split in 1959 into a 3rd and a 4th division. Our calculations do take into account all such institutional features and are available on request.

Table 5. Theoretical number of teams with highest seasonal winning percentage under equal playing strength.

Period	MLB	NFL	NHL	England	Italy	Belgium
51–00	22	23	18	37	40	39
61–00	21	21	18	32	33	31
71–00	18	19	17	26	26	24
81–00	14	15	14	18	18	17
91–00	9	9	9	9	9	9

number of potential winners relative to the European leagues where the size of the top division varies in size between 16 and 22 teams. The second effect is that promotion and relegation gives more teams an opportunity to enter the major league. With equal winning probabilities it can be seen that these two effects would have canceled each other out over the last decade and the CL would have produced as many winners as the open leagues. Over time however, the promotion and relegation effects increasingly dominates the expansion effect and over a 50 year period the open leagues should have produced around twice as many winners as the closed leagues.

In all cases, the actual number of winners in each cell of Table 3 is smaller than the theoretical prediction, but the shortfall is much more pronounced for the open leagues.

Table 6 illustrates a sharper contrast between the OL and CLs. Even in a relatively short period of time the promotion effect dominates the expansion effect so that under equal win probabilities the OLs could have been expected to see more entrants into the top 5 ranks. This contrasts with Table 4 where it was shown that in reality the situation was the reverse—more teams entered the top 5 ranks in the CLs compared to the OLs. Even within a 10 year period most teams in a CL should enter the top 5 ranks—so over a longer period of time the theoretical number of entrants does not increase by much. However, for open leagues the theoretically possible number of entrants increases rapidly, so that over 50 years the number of entrants under equal win probabilities is around one hundred. Once again, the gap between theory and reality is much greater for the open leagues.

To illustrate the size of the gap Figures 1 and 2 show the relationship between theoretical and actual entry to the top for NFL while Figures 3 and 4 offer the same comparison for Serie A, the top soccer division of Italy. In each figure, the broken lines represent the theoretical number of entrants for each of the 5 time ranges, while the solid lines illustrate the actual rate of entry. The figures also provide some perspective on the

Table 6. Theoretical number of teams with top 5 seasonal winning percentage under equal playing strength.

Period	MLB	NFL	NHL	England	Italy	Belgium
51–00	28	29	25	82	104	100
61–00	28	29	25	78	94	84
71–00	28	29	25	70	78	67
81–00	28	29	25	57	58	48
91–00	25	25	24	32	33	28

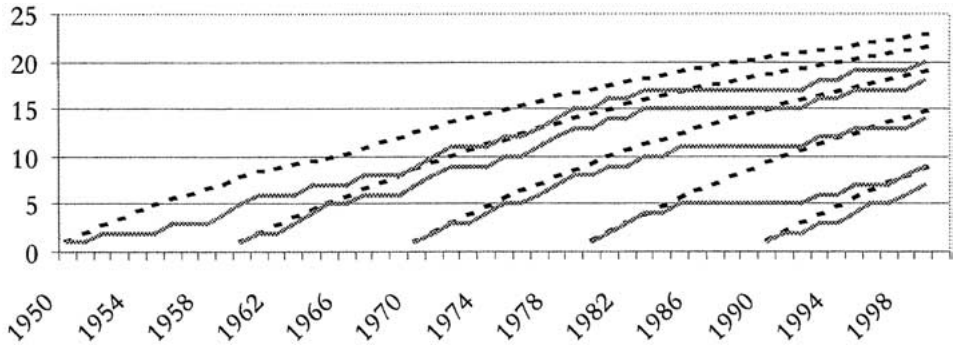

Figure 1. Entry to the highest rank in the NFL.

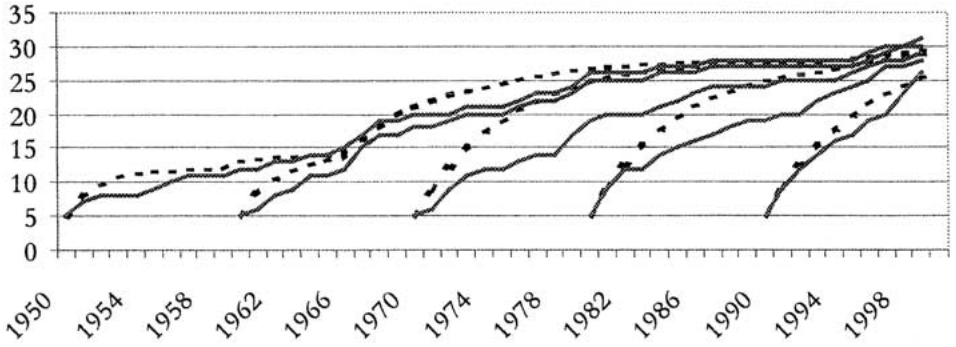

Figure 2. Entry to the top 5 ranks in the NFL.

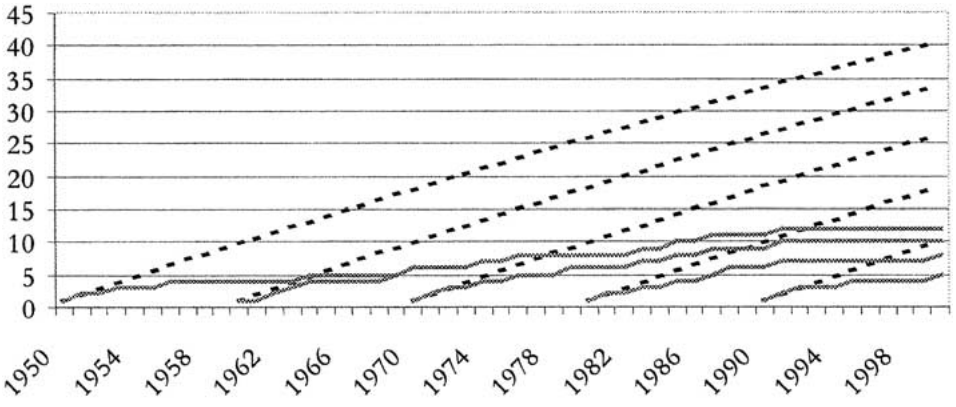

Figure 3. Entry to the highest rank in Italy.

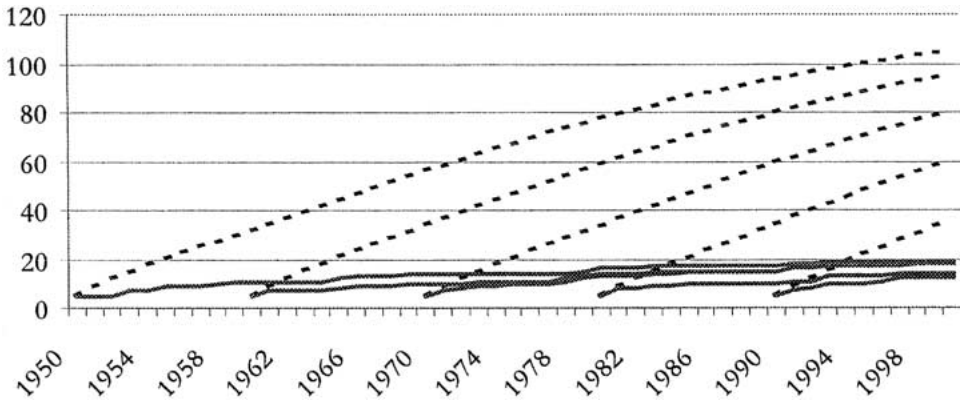

Figure 4. Entry to the top 5 ranks in Italy.

expected and actual entry on a year by year basis. Figures 1 and 2 show that actual entry is quite close to theoretical entry assuming equal playing strengths, suggesting that in dynamic terms the NFL is a fairly balanced competition (particularly looking at entry into the top 5 win percentages). On the other hand, Figures 3 and 4 demonstrate a large gap between theoretical and actual entry in Italy. Actual entry increases only very slowly, both for into the group of champions and into the top 5, and there is no evidence of convergence towards the theoretical limit. This suggests that equality of opportunity in open leagues in Europe has not translated into any equality of outcomes.

While these charts paint a very clear picture, it is desirable to quantify the differences between the OL and CLs in some way. We propose a Gini-type index that relates theoretical to actual entry. Thus, we calculate an index G where

$$G(T^*) = \frac{\sum_{T=1}^{T^*} y^L(k,T) - \sum_{T=1}^{T^*} y_a^L(k,T)}{\sum_{T=1}^{T^*} y^L(k,T)} \tag{5}$$

where T^* is the range of years considered and $y^L(k,T)$ and $y_a^L(k,T)$ are respectively the theoretical and the actual number of teams appearing in rank k or higher in a given league $L=\{\text{CL},\text{OL}\}$ over a period of T years. Thus a value of G close to zero indicates a perfectly balanced league while a value of G close to unity indicates a perfectly unbalanced league.[19] Table 7 reports the G-index for the highest seasonal win percentage across the leagues.

For every period considered the G-index for the closed leagues is lower than the G-index for the OLs, suggesting that the closed leagues were closer to the theoretical distribution under equal winning probabilities. In fact, the G-index for the CLs never rose above 0.5, while for the OLs value is either above or close to 0.5 for almost all periods.

19 This indicator depends on the starting year; moreover the longer the time series the less informative is the more recent data.

Table 7. *G*-index for teams with highest seasonal winning percentage.

Period	MLB	NFL	NHL	England	Italy	Belgium
51–00	0.34	0.19	0.42	0.44	0.66	0.70
61–00	0.30	0.15	0.36	0.44	0.55	0.70
71–00	0.30	0.21	0.40	0.54	0.48	0.62
81–00	0.05	0.39	0.39	0.57	0.44	0.51
91–00	0.19	0.24	0.12	0.24	0.38	0.54

Table 8. *G*-index for teams with top 5 seasonal winning percentage.

Period	MLB	NFL	NHL	England	Italy	Belgium
51–00	0.16	0.04	0.18	0.54	0.78	0.64
61–00	0.18	0.08	0.20	0.60	0.77	0.59
71–00	0.17	0.17	0.23	0.58	0.70	0.59
81–00	0.09	0.15	0.17	0.53	0.67	0.53
91–00	0.11	0.07	0.14	0.35	0.51	0.37

Looking at the individual leagues, there is some indication that baseball has become more balanced over the last 20 years while there is no obvious trend in the NFL or NHL. For the open leagues Belgium was generally furthest away from the theoretical distribution under equal win probabilities, but for the individual leagues there was no clear trend over time.

As far as the top 5 are concerned, the *G*-index scores for the CLs are all very similar and close to 0. Comparing Tables 7 and 8 this suggests that the CLs have been more successful at creating contenders rather than sharing out the most successful slot. However, since the ultimate Championship winners have been determined by play-offs that are more random than the regular season, this is probably not a problem. By contrast, the *G*-index for the OL suggest that entry into the top 5 has been more or less as difficult as into the top rank, and without a system of play-offs this suggests both little mobility at the top and considerably less competitive balance than in the closed leagues.

6. Conclusions and policy implications

This paper has proposed a measure of competitive balance that is dynamic, taking into account the turnover of teams at the top over time, rather than conventional measures that tend to emphasize *within*- but not *between*-season competitive balance. We have shown that by a conventional measure the open soccer leagues of Europe are, if anything, more balanced than the North American closed leagues. However, by the dynamic measure of competitive balance the OLs appear significantly less balanced than the closed leagues. We believe that the dynamic measure presents a better picture of competitive balance than the static measure.

One reason for believing that this is a better picture is that we have calculated the theoretical distribution of winning teams under the null hypothesis of equal winning probabilities for the teams, and shown that the open leagues deviate far more from the theoretical distribution than the OLs. The hypothesis that the Europe's open leagues are competitively balanced is far harder to support than the hypothesis that North America's CLs are balanced.

It is important to ask why the European pattern should be so different from the North American one. In general, successful teams draw more support, so that the greater concentration of success in Europe suggests more concentrated support. For example, clubs from Milan (AC and Inter) and Turin (Juventus) have dominated the championship with relatively few other population centers. For example, Rome (population 3.5 m) can boast only four championship titles since the start of the Serie A in 1929, compared to Turin's 28 (population 1.5 m).[20] However, it is by no means clear that fans of Roma and Lazio are less interested or less willing to pay for success than their more successful northern rivals. Moreover, while in many countries the most successful teams have been located in the largest cities (most notably in the case of Spain with Real Madrid and Barcelona) in many other countries this is not clear. In England, for example, league competition has for the last 30 years been dominated by teams from Liverpool (3rd largest population,[21] 13 titles) and Manchester (7th largest, 7 titles) while London based teams have won only 4 titles and Birmingham (2nd largest) have won only a single title.

If concentration is not a product of population endowments, it may be a product of hysteresis. Dynasties are not unknown in North American sports, most notably the Yankees in baseball.[22] But dynasties are usually seen as a significant source of competitive imbalance needing to be counteracted with redistributive measures either in the labour market (*e.g.* salary caps, roster limits) or in the product market (revenue sharing).[23] Such measures are much less widely adopted in Europe. Labor markets operate almost entirely free of any constraints while revenue sharing is rare and limited in extent. But if competitive balance promotes interest in the sport why wouldn't European leagues take on the kind of extensive redistribution seen in North America?

Two possible answers suggest themselves. One is that competitive balance in the sense described here does not matter to the fans. There certainly does not appear to be less interest in European soccer than in North American sports. But as long as the contest within each season is close (*e.g.* measured by standard deviation of win percent) then fans may be indifferent to dominance by a small number of teams over many seasons. This suggests one way in which the study of competitive balance should develop, *i.e.* to focus on the number of teams that need to be in contention to make the contest interesting.

20 Tommasi (2000).

21 Source: www.citypopulation.de, Thomas Brinkhoff.

22 See Levi et al. (2000) for a discussion of baseball's competitive balance problem.

23 These measures are reviewed in detail in Szymanski (2003).

A second possible explanation is that the promotion and relegation system itself mitigates against welfare enhancing redistribution schemes. Szymanski and Valletti (2003) compare the incentive to redistribute income in closed and open league systems (*i.e.* with promotion and relegation). They argue that the cost of revenue sharing to large drawing teams is the foregone income from current success, while the benefit is their share in a more valuable (because more balanced) contest. In a closed league every team is guaranteed to participate in that contest, while in an open league any team might be relegated in the future. Thus a strong team has a weaker incentive to share its revenues in an open system.[24] From a policy point of view, perhaps the most interesting topics for further debate are (a) whether the alternative sources of interest in European soccer compensate fans for the relative lack of competitive balance and (b) returning to the Rosen and Sanderson quotation at the beginning of the paper, whether the choice of balance enhancing measures are hindered by the promotion and relegation system itself.

Annex

In this annex, we propose an indicator alternative to the *G*-index. As a first step, we calculate—as with the *G*-index—$y^L(k,T)$, *i.e.* the theoretical number of teams appearing in rank k over a period of T years. This number takes into account all the precise details of a certain league that may have changed over time. Then, we calculate $n(k, T)$, *i.e.* the equivalent dimension of a closed league with a constant structure that would have generated then same number of teams appearing in rank k over the same period. In the third step, we consider the actual number $y_a^L(k,T)$ and then construct $n_e(k,T)$, *i.e.* the theoretical dimension of a closed league with a constant structure that would have generated the same number. Finally, our indicator is given by the ratio $n_e(k,T)/n(k,T)$. Results are reported in Table A1. A league is balanced the closer is the *E*-index to 1, *i.e.* to the equivalent theoretical benchmark. Notice that this exercise allows to construct an indicator that is homogeneous both for open and for closed leagues. This indicator gives a snapshot of competitive balance at the end of a given period, without concentrating on how a particular configuration is reached over time—contrary to the *G*-index. Results illustrate once again the sharp contrast between open and closed leagues.

24 A third reason, suggested by a referee, is that the wider range of competitions on offer in Europe may offset the lack of dynamic competitive balance. For example, in one season the top teams compete in a domestic league, a domestic (knock-out) Cup competition, the European Champions' League, and in addition the top players are selected to play for their national team in international competition. There is certainly a greater array of competition in Europe, but this does not explain why the fans would not find a more balanced league competition more interesting or why the league authorities would not choose to try to make it so. Moreover, Szymanski (2001) provides evidence that lack of competitive balance in the English FA Cup is leading to declining interest in that competition.

Table A1. *E*-index for teams with highest seasonal winning percentage.

Period	MLB	NFL	NHL	England	Italy	Belgium
51–00	0.65	0.83	0.67	0.21	0.12	0.11
61–00	0.67	0.75	0.65	0.15	0.12	0.09
71–00	0.61	0.58	0.58	0.10	0.12	0.13
81–00	0.62	0.36	0.49	0.07	0.11	0.15
91–00	0.29	0.41	0.36	0.13	0.09	0.13

Table A2. *E*-index for teams with top 5 seasonal winning percentage.

Period	MLB	NFL	NHL	England	Italy	Belgium
51–00	1.00	1.06	0.84	0.39	0.16	0.29
61–00	1.00	1.02	0.84	0.31	0.16	0.29
71–00	1.00	0.99	0.89	0.29	0.18	0.31
81–00	1.01	0.97	0.83	0.28	0.18	0.37
91–00	0.91	1.06	0.70	0.34	0.23	0.51

References

Andreff, W. and Staudohar, P., ''The evolving model of european sports finance,'' Journal of Sports Economics, vol. 1 no. 3, pp. 257–276, 2000.

European Commission, The European Model of Sport. Consultation paper of DGX, Brussels, 1998.

Fort, R. and Quirk, J., ''Cross subsidization, incentives and outcomes in professional team sports leagues,'' Journal of Economic Literature, vol. XXXIII no. 3, pp. 1265–1299, 1995.

Hall, S., Szymanski, S., and Zimbalist, A., ''Testing causality between team performance and payroll: the case of major league baseball and english soccer,'' Journal of Sports Economics, vol. 3 no. 2, pp. 149–168, 2002.

Hoehn, T. and Szymanski, S., ''The Americanization of European football,'' Economic Policy, vol. 28, pp. 205–240, 1999.

Horowitz, I. ''The increasing competitive balance in major league baseball,'' Review of Industrial Organization, vol. 12, pp. 373–387, 1997.

Humphreys, B., ''Alternative measures of competitive balance in sports leagues,'' Journal of Sports Economics, vol. 3 no. 2, pp. 133–148, 2002.

Kesenne, S., ''Revenue sharing and competitive balance in professional team sports,'' Journal of Sports Economics, vol. 1 no. 1, pp. 56–65, 2000.

Kipker, I., ''Determinanten der zuschauernachfrage im professionellen teamsport: Wie wichtig ist die sportliche ausgeglichenheit?'' Unpublished chapter of Ph.D. dissertation, 2000.

Koop, G., ''Modelling the evolution of distributions: an application to Major League Baseball,'' University of Glasgow Discussion Paper, 2002.

Levi, R., Mitchell, G., Volcker, P., and Will, G., The report of the independent members of the commissioner's blue ribbon panel on baseball economics. Major League Baseball, NY, 2000.

Noll, R., ''The economics of promotion and relegation in sports leagues; the case of English football,'' Journal of Sports Economics, vol. 3 no. 2, pp. 169–203, 2002.

Quirk, J. and Fort, R., Pay Dirt: The Business of Professional Team Sports. Princeton University Press: New Jersey, 1992.

Rosen, S. and Sanderson, A., ''Labour markets in professional sports,'' Economic Journal, vol. 111 no. 469, pp. F47–F68, 2001.

Ross, S. and Lucke, R., ''Why highly paid athletes deserve more antitrust protection than unionized factory workers,'' Antitrust Bulletin, vol. 42 no. 3, pp. 641–679, 1997.

Ross, S. and Szymanski, S., ''Promotion and relegation,'' World Economics, vol. 2 no. 2, pp. 179–190, 2001.

Ross, S. and Szymanski, S. ''Open competition in league sports,'' Wisconsin Law Review, vol. 2002 no. 3, pp. 625–656, 2002.

Rottenberg, S., ''The baseball players' labor market,'' Journal of Political Economy, vol. 64, pp. 242–258, 1956.

Sloane, P.J., ''The economics of professional football: the football club as a utility maximizer,'' Scottish Journal of Political Economy, vol. 17 no. 2, pp. 121–146, 1971.

Szymanski, S., ''Income inequality, competitive balance and the attractiveness of team sports: Some evidence and a natural experiment from English soccer,'' Economic Journal, vol. 111 no. 469, pp. F69–F84, 2001.

Szymanski, S., ''The economic design of sporting contests,'' Journal of Economic Literature, forthcoming, 2003.

Szymanski, S. and Kuypers, T., Winners and losers: the business strategy of football. Penguin Book: London, 1999.

Szymanski, S. and Valletti, T., ''Promotion and relegation in sporting contests,'' Imperial College Business School Discussion paper, 2003.

Tommasi, R., Storia della Serie A. Edizioni Marchesi Grafiche, Rome, 2000.