Economics Letters 109 (2010) 38–41

Contents lists available at ScienceDirect

Economics Letters

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

Limitations of the relative standard deviation of win percentages for measuring competitive balance in sports leagues ☆

P. Dorian Owen *

Department of Economics, University of Otago, PO Box 56, Dunedin 9054, New Zealand

ARTICLE INFO

Article history:
Received 1 June 2009
Received in revised form 16 July 2010
Accepted 28 July 2010
Available online 6 August 2010

JEL classifications:
D63
L83

Keywords:
Competitive balance
Relative standard deviation
Idealized standard deviation
Sports economics

ABSTRACT

The relative standard deviation of win percentages, widely used to measure competitive balance, has an upper bound which varies with the numbers of teams and games played. Accounting for this upper bound provides additional insights into competitive balance comparisons.

1. Introduction

Competitive balance in sports leagues, i.e., how evenly teams are matched, is reflected in the degree of inequality in match and championship outcomes. Because of its pivotal role in the economic analysis of professional sport, considerable effort has gone into measuring competitive balance. By far the most commonly used measure is the relative standard deviation of win percentages. This compares the actual (ex post) standard deviation of win percentages with the standard deviation of win percentages in the 'idealized' case in which each team has an equal chance of winning each game.

The relative standard deviation of win percentages is widely regarded as the most useful measure of competitive balance "because it controls for both season length and the number of teams, facilitating a comparison of competitive balance over time and between leagues" (Fort, 2007, p. 643). Although it explicitly incorporates season length and the number of teams, it does not control for these variables in the sense of partialling out their effects. Moreover, the league's playing schedules impose an upper bound on the value of the relative standard deviation, which is also sensitive to season length and the number of teams. Ignoring its feasible range of outcomes limits the usefulness of the relative standard deviation for comparing within-season competitive balance across leagues or over time if the numbers of teams and/or games played are not constant, which in practice is usually the case. Additional insights can be gained by using a normalized standard deviation measure that takes into account variations in the relevant upper bound.

2. Measuring competitive balance with actual and relative standard deviations

Competitive balance in a sports league is a multi-faceted concept. The different dimensions include the distribution of wins across teams in the league within a single season, the persistence of the teams' record of wins across successive seasons over time, and the degree of concentration of overall championship wins reflected in the teams' shares of championship wins over a number of seasons (Kringstad and Gerrard, 2007).

The ex post or 'actual' standard deviation (*ASD*) of the teams' win ratios (or, equivalently, win percentages) in a single season is a natural measure for the first of these dimensions. This can be represented as

$$ASD = \sqrt{\sum_{i=1}^{N} [(w_i/G_i) - 0.5]^2 / N} \quad (1)$$

☆ I am grateful to an anonymous referee for very helpful comments and suggestions, and to the participants at the New Zealand Econometric Study Group meeting, Auckland (February 2010), the Australian Conference of Economists, Adelaide (September 2009), and a seminar in the Department of Economics and Finance, La Trobe University (November 2009).

* Tel.: +64 3 479 8655; fax: +64 3 479 8174.
E-mail address: Dorian.Owen@otago.ac.nz.

0165-1765/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.econlet.2010.07.012

in which N equals the number of teams in the league, and w_i and G_i are, respectively, the number of wins accumulated and the number of games played by team i in a season. A smaller standard deviation of win ratios across teams in a season indicates a more equal competition. However, when comparing values of *ASD*, either for the same league over time or across different leagues, N and/or G are typically not constant. Other things equal, *ASD* tends to decrease as G increases because there is likely to be less random noise in the final outcomes. Hence, it is common to compare *ASD* to a benchmark 'idealized standard deviation' corresponding to an ex ante representation of a perfectly balanced league in which each team has an equal probability of winning each game.[1] In the absence of ties (draws), the idealized standard deviation, $ISD = 0.5/G^{0.5}$ can be derived as the standard deviation of a binomially distributed random variable with a (constant) probability of success of 0.5 across independent trials (Fort and Quirk, 1995).[2] The relative standard deviation, *RSD* is expressed as *ASD*/*ISD*. As G increases, any reduction in *ASD* will be compared against the reduced value of the benchmark *ISD*.

RSD is a 'static' measure based on the variation of (final) win ratios across teams in a single season. Its evolution can be plotted over time, but it does not capture championship concentration or persistence of performance of individual teams over successive seasons. Given the multidimensional nature of competitive balance, it is generally considered unrealistic to expect any single measure to reflect all of its different dimensions. This apart, the *RSD* measure has met with widespread acceptance. It is the most widely used competitive balance measure in the sports economics literature; e.g., see Fort (2006a, Table 10.1).

However, despite its resounding endorsement as "the tried and true" measure of within-season competitive balance (Utt and Fort, 2002, p. 373), *RSD* has properties that limit its usefulness in comparisons of competitive balance involving different numbers of teams and/or games. Firstly, *RSD* has an upper bound. The league's playing schedules impose an upper limit on the variance of the distribution of wins; this has implications for interpreting *RSD* that have not been recognized.[3] A second distinctive feature of *RSD* is the different measures of 'sample size' that appear in its numerator (N, the number of teams) and denominator (G, the number of games played by each team). If each team plays the other teams several times in a season, then N and G can differ markedly from each other. These characteristics can complicate the interpretation of exactly the sorts of comparisons of competitive balance (involving scenarios with different N and/or G) for which *RSD* is usually advocated (e.g., Fort, 2006b, pp. 175–177; Leeds and von Allmen, 2008, pp. 156–157).

3. The upper bound of the relative standard deviation

The upper bound of *RSD* can be derived by considering the ex post 'most unequal distribution' of win ratios (Fort and Quirk, 1997; Horowitz, 1997; Utt and Fort, 2002). This involves one team winning all its games, the second team winning all except its game(s) against the first team, and so on down to the last team, which wins none of its games. For ease of exposition, consider balanced schedules of games in which each of the N teams plays every other team the same number of times, K, with no ties (draws) or with ties (draws) treated as half a win. Each team plays $G_i = G = K(N-1)$ games.

The actual (ex post) variance of win ratios (*AVAR*) across the N teams in a season (with the mean win ratio equal to 0.5 for any degree of competitive balance) is given by:

$$AVAR = \left[\sum_{i=1}^{N} (w_i/G_i)^2/N\right] - (0.5)^2 = \left[\sum_{i=1}^{N} (w_i/K(N-1))^2/N\right] - (0.5)^2.$$

In a perfectly unbalanced league, its upper bound, $AVAR^{ub}$, is given by:

$$AVAR^{ub} = \frac{1}{N}\left[\frac{K^2(N-1)^2}{K^2(N-1)^2} + \frac{K^2(N-2)^2}{K^2(N-1)^2} + \ldots + \frac{K^2(N-N)^2}{K^2(N-1)^2}\right] - (0.5)^2.$$

Note that the K^2 terms cancel, implying that $AVAR^{ub}$, and hence the corresponding upper bound for *ASD*, are invariant to the number of rounds played if schedules are balanced. Simplifying this expression,

$$\begin{aligned} AVAR^{ub} &= \frac{1}{N(N-1)^2}\left[(N-1)^2 + (N-2)^2 + \ldots + (N-N)^2\right] - (0.5)^2 \\ &= \frac{1}{N(N-1)^2}\left[\frac{N(2N-1)(N-1)}{6}\right] - (0.5)^2 \quad \text{(Owen et al., 2007, p. 301)} \\ &= \frac{(2N-1)}{6(N-1)} - \frac{1}{4} = \frac{N+1}{12(N-1)}. \end{aligned}$$

Taking the square root, the upper bound for *ASD*, denoted ASD^{ub}, is given by:

$$ASD^{ub} = [(N+1)/\{12(N-1)\}]^{0.5}. \tag{2}$$

Substituting $G = K(N-1)$ into the expression for *ISD*, and noting that the ex ante *ISD* measure is unaffected by the actual outcome for *ASD*, gives:

$$RSD^{ub} = \frac{ASD^{ub}}{ISD} = \frac{[(N+1)/\{12(N-1)\}]^{0.5}}{0.5/[K(N-1)]^{0.5}} = 2[K(N+1)/12]^{0.5}. \tag{3}$$

The upper bound of *RSD* in Eq. (3) depends not only on the number of teams in the league, N, but also on the number of times they play against each other, K. Increases in N and/or K lead to increases in RSD^{ub}, conventionally interpreted as implying a decrease in competitive balance. However, given we are considering the upper bound, wins are initially as unequally distributed as they can be and remain that way. Thus, *RSD* captures the scale effect arising from the dependence of *ISD* on the number of games played but, by ignoring its upper bound, *RSD* does not reflect competitive balance relative to its feasible maximum.

The upper bound of *ASD* in Eq. (2) also depends on N, with expansions in N leading to a decrease in ASD^{ub}.[4] However, RSD^{ub} is much more sensitive than ASD^{ub} to variations in N. This is illustrated in Fig. 1, and is apparent from a comparison of Eqs. (2) and (3). For large N, the $(N+1)$ and $(N-1)$ terms approximately cancel out, so that, in the limit, ASD^{ub} tends to $(1/12)^{0.5} = 0.289$.[5] For smaller values of N, as in most sports leagues, the dependence on N is not removed entirely, but is relatively modest, with, for example, ASD^{ub} varying from 0.327 for $N=8$ to 0.298 for $N=30$ (a decrease of 8.8%). In contrast,

[1] The use of a relative measure involving a benchmark standard deviation corresponding to an ex ante perfectly balanced league is attributable to Noll (1988) and Scully (1989), but became popular following its use by Quirk and Fort (1992) and Fort and Quirk (1995).

[2] If ties are possible, *ISD* can be applied to absolute total points or the percentage of points, with amendments to account for different possible points assignments for wins, ties and losses (e.g., Fort, 2007).

[3] The implications of the league's schedule of matches for interpretation of the Gini coefficient and the Herfindahl–Hirschman index applied to wins are examined by Utt and Fort (2002) and Owen et al. (2007) respectively. Given these measures' emphasis on teams' shares of wins, their focus is primarily on the fact that teams cannot win games in which they do not play.

[4] It is straightforward to show that $\partial ASD^{ub}/\partial N < 0$ if $N \geq 2$.

[5] This asymptotic result is consistent with *ASD* corresponding more closely to a pure inequality measure, such as *IGE*(2), a member of the family of generalized entropy measures of inequality (Bajo and Salas, 2002). $IGE(2) = CV^2/2$, where *CV* is the coefficient of variation. If the mean of the win ratios in a season equals 0.5, variation in *CV* applied to win ratios corresponds to variation in *ASD*.