



School of Business
Te Kura Pakihi

ISSN 1178-2293 (Online)

University of Otago
Economics Discussion Papers
No. 0908

Original version: July 2009
Revised version: March 2010

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

P. Dorian Owen

ABSTRACT

The relative standard deviation of win percentages, the most widely used measure of within-season competitive balance, has an upper bound which is very sensitive to variation in the numbers of teams and games played. Taking into account this upper bound provides additional insight into comparisons of competitive balance across leagues or over time.

Keywords:

Competitive balance; Relative standard deviation; Idealized standard deviation

JEL classification:

D63, L83

Contact details:

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

email: Dorian.Owen@otago.ac.nz

tel: (64) 3 479 8655

fax: (64) 3 479 8174

[☆] I am grateful to an anonymous referee for very helpful comments and suggestions, and to 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).

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 teams' record of wins across successive seasons over time, and the degree of concentration of overall championship wins reflected in teams' shares of championship wins over a number of seasons (Kringstad and Gerrard, 2007).

The ex post or 'actual' standard deviation (*ASD*) of 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

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

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

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).

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

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 can not win games in which they do not play.

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:

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

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,

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

$$= \frac{(2N-1)}{6(N-1)} - \frac{1}{4} = \frac{N+1}{12(N-1)}$$

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

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

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:

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

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

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).

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, $RSD^{ub} \rightarrow \infty$ as $N \rightarrow \infty$ and, for commonly observed values of N , the increase in RSD^{ub} is more dramatic than for ASD^{ub} , varying (for $K = 1$) from 1.732 for $N = 8$ to 3.215 for $N = 30$ (an increase of 85.6%).

Variation in the upper bounds can be explicitly incorporated in a normalized measure of competitive balance, such as

$$RSD^* = RSD/RSD^{ub} = ASD/ASD^{ub} = ASD^* \quad (4)$$

The normalized measure, ASD^* ($= RSD^*$), lies in the interval $[0, 1]$, with 0 representing perfect parity and 1 maximum imbalance.⁶ Goossens (2006) also suggests using the equivalent of ASD^* , although she calculates ASD^{ub} numerically for different values of N rather than deriving a general expression. Goossens's argument for preferring ASD^* to RSD is that RSD can be less than 1 (if $ASD < ISD$). This is not surprising given the ex ante probabilistic nature of ISD , in comparison to the ex post minimum (0) and maximum (ASD^{ub}) values for ASD . However, given the properties discussed above, the additional insights from comparing RSD or ASD relative to their upper bounds provide a more fundamental justification for the use of a normalized standard deviation measure, especially in cases involving different N and K values.

The limitations of RSD and the usefulness of also examining the normalized version of RSD or ASD are easiest to demonstrate in cases in which N or K is varied and the degree of imbalance is controlled, as in the case of the upper bounds. However, additional insights

$IGE(2) = CV^2/2$, 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 .

⁶ Trivially, this adjustment also removes the dependence of the upper bound of RSD on K .

are not confined to the case of perfect inequality. For example, consider increasing K while ‘scaling up’, i.e., reproducing exactly the corresponding set of results for the original league. As a concrete example, consider a league with balanced schedules ($N = 16$, $K = 2$, $G = 30$) in which ASD is 0.037 and RSD is 0.408.⁷ Assume K is increased to 4 (with each team playing every other team twice at home and twice away) and that the additional home and away results are identical to those in the actual two rounds played. This scaled-up set of results displays unchanged win ratios (and the same value for ASD); however, RSD increases to 0.577. Because the upper bound for RSD increases as K increases (as in Eq. (3)), the normalized measure, RSD^* ($= ASD^*$) remains unchanged, reflecting what is, in a relative sense, an unchanged competitive balance situation. By comparison, the increase in RSD , for an unchanged ASD , reflects the scale effect of more games decreasing ISD .

If N varies across time for a given league, ASD^* can again offer useful additional insights. For example, in the First Division of New Zealand Rugby Union’s National Provincial Championship (NPC) $K = 1$ in round-robin play (prior to the semi-finals) but N has varied over time. The RSD of win ratios (with draws, which are relatively rare in rugby, counting as 0.5 of a win) was 1.763 in 1990 compared to 1.633 in 1994, suggesting a higher level of competitive imbalance in 1990. The upper bounds for RSD were, respectively, 2.000 ($N = 11$) and 1.826 ($N = 9$); RSD was therefore closer to the upper bound reflecting complete inequality in 1994 ($ASD^* = 0.894$) than in 1990 ($ASD^* = 0.882$). The improvement in competitive balance between 1990 and 1994 signaled by RSD reflects the effect of reduced N ; ASD^* reverses the ranking in terms of relative competitive balance, but, more importantly, shows how close to perfect imbalance the NPC was in both years.

⁷ These results apply to what is widely believed to be one of the most competitive league outcomes in association football: the 1983-84 season in the somewhat obscure Romanian Divizia C, Seria a VIII-a league (see <http://www.rssf.com/miscellaneous/even.html>).

When comparing competitive balance across different leagues, the differences in N and K are often considerably greater than for a single league over time, so the relevant upper bounds of RSD can differ markedly. Because of the sensitivity of RSD^{ub} to K and N , it is feasible for the observed values of RSD in one league (with $0 < RSD < RSD^{ub}$) to be greater than the value of RSD^{ub} in another. For example, the *upper bound* for RSD for the NPC noted above (1.826 to 2) is *less* than many of the calculated RSD values for the ‘big four’ US leagues reported in Fort’s (2006b) Table 6.3, especially basketball for which all reported values are greater than 2 and many are greater than 3. Therefore, the numerical value of RSD provides no guidance on how that outcome compares to the relevant completely unequal distribution; ASD^* ($= RSD^*$) provides this information. However, relying solely on the normalized measure may not identify cases of imbalance that are worse in an absolute sense. For example, consider a less than completely unequal distribution of wins in baseball (with 162 games per team) in which the weakest team loses, say, 140 games. Even though $ASD^* < 1$, this may reasonably be regarded as a more concerning case of imbalance than a completely unequal distribution of wins in the NPC (in which $ASD^* = 1$ but, with $N = 11$, the weakest team loses only 10 games).

4. Conclusion

In the sports economics literature there is a widespread belief that the idealized standard deviation provides a ‘common standard’ (Leeds and von Allmen, 2008, p. 156) against which to compare the actual standard deviation of win ratios. However, the resulting relative standard deviation measure has an upper bound, and hence a range of feasible values, which varies markedly in response to variation in the number of teams and/or number of games played. When making comparisons across leagues or over time, using

RSD, but ignoring its upper bound, provides at best a partial view of competitive balance that emphasizes the scale effects of different values of N and/or G . The added insights obtained by considering the upper bound of *RSD* provide a much more compelling reason than previously advanced for also using a relative measure, such as the normalized standard deviation, when evaluating within-season competitive balance.

References

- Bajo, O., Salas, R., 2002. Inequality foundations of concentration measures: An application to the Hannah-Kay indices. *Spanish Economic Review* 4, 311-316.
- Fort, R., 2006a. Competitive balance in North American professional sports. In: Fizek, J. (Ed.), *Handbook of Sports Economics Research* (M.E. Sharpe, Armonk, NY), pp. 190-206.
- Fort, R.D., 2006b. *Sports Economics (Second edition)* (Pearson Prentice Hall: Upper Saddle River, NJ).
- Fort, R., 2007. Comments on “measuring parity”. *Journal of Sports Economics* 8, 642-651.
- Fort, R., Quirk, J., 1995. Cross-subsidization, incentives, and outcomes in professional team sports leagues. *Journal of Economic Literature* 33, 1265-1299.
- Fort, R., Quirk, J., 1997. Introducing a competitive economic environment into professional sports. In: Hendricks, W. (Ed.), *Advances in the Economics of Sport, Volume 2* (JAI Press, Greenwich, CT), pp. 3-26.
- Goossens, K., 2006. Competitive balance in European football: comparison by adapting measures: national measure of seasonal imbalance and top3. *Rivista di Diritto ed Economia dello Sport* 2, 77-122.
- Horowitz, I., 1997. The increasing competitive balance in Major League Baseball. *Review of Industrial Organization* 12, 373-387.
- Kringstad, M., Gerrard, B., 2007. Beyond competitive balance. In: Parent, M.M., Slack, T. (Eds), *International Perspectives on the Management of Sport* (Butterworth-Heinemann, Burlington, MA), pp. 149-172.
- Leeds, M., von Allmen, P., 2008. *The Economics of Sport (Third edition)*. (Pearson Addison Wesley: Boston, MA).

Noll, R.G., 1988. Professional Basketball, Studies in Industrial Economics Paper No. 144.
Stanford University, Stanford, CA.

Owen, P.D. et al., 2007. Measuring competitive balance in professional sports using the Herfindahl-Hirschman index. *Review of Industrial Organization* 31, 289-302.

Quirk, J., Fort, R.D., 1992. *Pay Dirt: The Business of Professional Team Sports*. (Princeton University Press, Princeton, NJ).

Scully, G.W., 1989. *The Business of Major League Baseball*. (University of Chicago Press, Chicago, IL).

Utt, J., Fort, R., 2002. Pitfalls to measuring competitive balance with Gini coefficients. *Journal of Sports Economics* 3, 367-373.

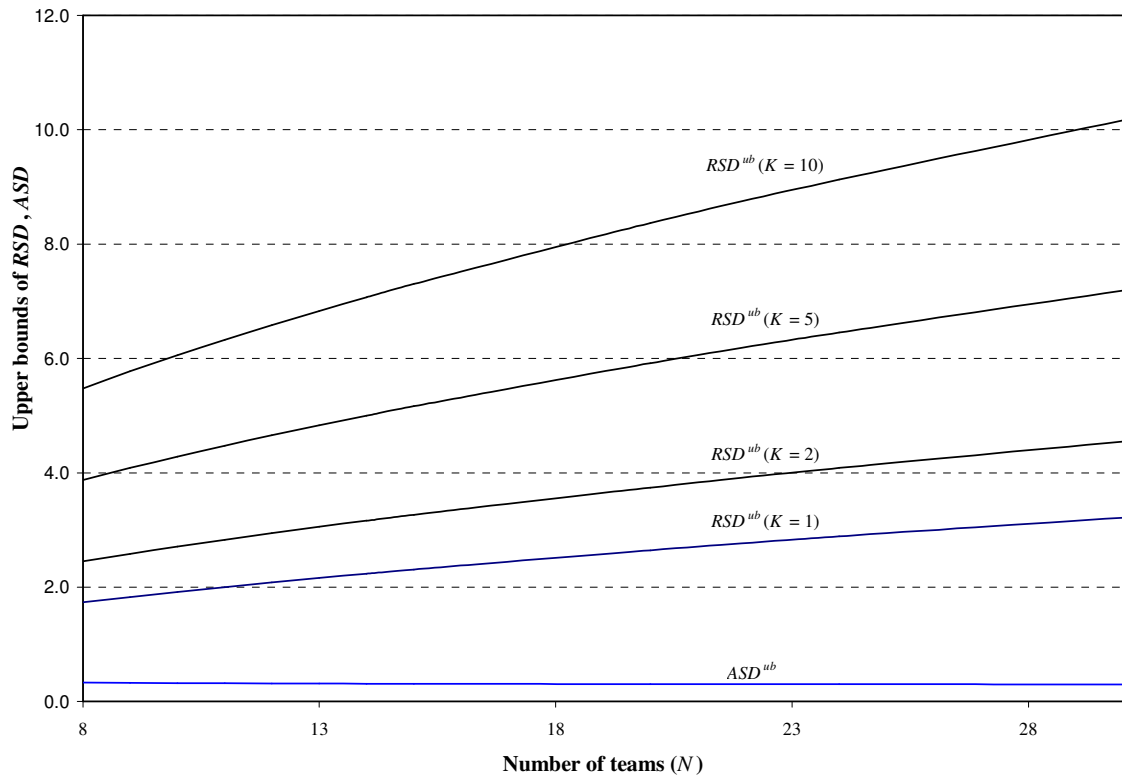


Fig.1. Variation in the upper bounds of RSD and ASD with N and K