



School of Business
Te Kura Pakihi

ISSN 1178-2293 (Online)

University of Otago
Economics Discussion Papers
No. 1005

July 2010

Measuring Parity in Sports Leagues with Draws: Further Comments

P. Dorian Owen

This paper re-examines the calculation of the relative standard deviation (RSD) measure of competitive balance in leagues in which draws are possible outcomes. Some key conclusions emerging from the exchange between Cain and Haddock (2006) and Fort (2007) are reversed. There is no difference, for any given points assignment scheme, between the RSD for absolute points compared to percentages of points. However, variations in the points assignment that change the ratio of points for a win compared to a draw do result in different RSD values, although the numerical differences are minor.

Keywords: sports economics, competitive balance, relative standard deviation, idealized standard deviation, draws/ties

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

Measurement of competitive balance, or ‘parity’, is an important consideration in the economic analysis of sports leagues, but this can be complicated by the possibility that some matches end as draws (or, equivalently, ties). This is especially the case for comparisons, either across different leagues or over time, involving different probabilities of a draw and/or different relative points assignments to wins and draws.

Cain and Haddock (2006, hereafter CH) consider the most widely used measure of within-season competitive balance, the relative standard deviation (RSD) of end-of-season match outcomes (win percentages, points or points percentages), and give examples of how to calculate the RSD for leagues in which draws are possible.

Fort (2007) criticizes several aspects of CH’s contribution, including the confusion caused by their deviation from standard terminology, inappropriate attribution of the development of the RSD, and lack of reference to other contributions on trinomial match outcomes. These are valid criticisms and will not be dwelt on here. However, Fort (2007, p. 642) also suggests that CH “miss the actual usefulness of their argument” that, if the RSD measures are based on *percentages* rather than absolute points, the different ways of assigning points to wins and draws makes no difference to the results. Consequently, he considers the merits of the absolute points versus percentage versions of the RSD in different contexts in which competitive balance is measured, such as tracking competitive balance over time or capturing the effects of balance on fan demand.

The aim of this comment is to show that these conclusions on percentage versus absolute points measures of the RSD do not follow from CH’s analysis. If the appropriate comparisons are made, there is no difference between the RSD for absolute points compared to percentages, for any particular points assignment involving draws. However,

varying the points assignment and/or the probability of a drawn match does affect the value of the RSD (whether applied to percentages or absolute points), although differences are quantitatively relatively minor for commonly used assignments.

The paper is structured as follows. First, the mechanics of calculating the RSD in leagues with draws are re-examined and anomalies in the CH-Fort exchange identified. The implications for the percentages versus absolute points result and the measurement differences due to adopting different points assignments are then examined, with illustrations from the English Football League examples considered by CH and Fort.

CALCULATING IDEALIZED STANDARD DEVIATIONS

The most commonly used measure of competitive balance in sports leagues is the RSD of win percentages, based on the distribution of wins across teams in the league within a single season (Fort, 2006). This compares the (ex post) actual standard deviation (ASD) of win percentages with the ‘idealized’ standard deviation (ISD) of win percentages, the latter taking its name from the idealized case in which each team has an equal chance of winning each game. The RSD measure, calculated by dividing the ASD by the ISD, can also be applied to other end-of-season outcomes, such as absolute points or percentages of possible points, and these are usually of more interest if points are allocated for outcomes other than straight wins.

CH’s discussion of the implications of different scoring systems is based on the example of the 1888-1889 (inaugural) season of the English Football League. In that season, the League consisted of twelve teams playing a balanced schedule: each team played every other team twice (once at home and once away). The end-of-season record of wins, draws, losses and implied points totals for each team is reported in CH’s (2006)

Table 1 and Fort's (2007) Table 1, so is not repeated here. Instead, Table 1 reports the overall league-wide values for ASD, ISD and RSD. Following CH and Fort, the notation (3,1,0), for example, represents three points for a win, one point for a draw, and zero points for a loss. Panel A combines the (1,0.5,0) results from Fort's Table 1, the (2,1,0) results from CH's Table 1 repeated (with the addition of the ISD) in Fort's Table 1, and the (3,1,0) results from CH's Table 1 (again adding the ISD values).¹

The ASD values are for the end-of-season outcomes for the variable of interest, either absolute points accumulated (*Points*) or points accumulated as percentages (or, strictly, proportions) of maximum possible points (*Points%*).² The ISD corresponds to an ex ante representation of a perfectly balanced league.³ In the absence of draws (ties), the ISD is derived as the standard deviation of a binomially distributed random variable with a constant probability of 'success' (winning) of 0.5 across independent trials. Applied to win proportions, $ISD = 0.5/N^{0.5}$, where N is the number of games played by each team in a season; see Fort and Quirk (1995) for a derivation. Fort (2007, p. 642) refers to this conventional binomial-based calculation as the (1,0.5,0) approximation "that assigns half a win to both teams in a tie game or match". With $N = 22$, this corresponds to the entry for the ISD (equal to 0.107 to three decimal places) in the *Points%* columns in Panel A in Table 1, as reproduced from CH and Fort. Note that the entries for the ISD for *Points%* in Panel A, calculated using the formula above, are treated as invariant to both the probability of a draw and the points allocation system under consideration; this is further discussed below. The entry for the ISD for *Points* in the (1,0.5,0) column in Panel A is the corresponding binomial standard deviation applied to absolute points, i.e., $0.5N^{0.5}$. The entries for the RSD are then calculated as ASD/ISD for each outcome of interest.

Fort's (2007, p. 642) most fundamental criticism of CH is that "they ignore the fact (after acknowledging it) that if one uses percentages rather than absolute points, their measures of competitive balance are invariant to the point system actually used". This is illustrated in Table 1, Panel A, for which the RSD values based on *Points%* are identical for the entries for the (1,0.5,0) and (2,1,0) points assignments. Note, however, that the corresponding RSD value for the (3,1,0) assignment (reported in CH's Table 1, but not in Fort's Table 1), although the same to one decimal place, is not strictly identical to the value for (1,0.5,0) and (2,1,0).

However, the use of the binomial '(1,0.5,0) approximation' in the RSD calculations, although apparently innocuous, is misleading. Several points are worth noting.

There is an internal contradiction between the use of the pure binomial approximation to evaluate the ISD and the recognition that draws have non-zero probabilities. The binomial approximation is based on the assumption that $P(\text{win}) = 0.5 = P(\text{loss})$. As wins, draws and losses are mutually exclusive and exhaustive events, this implies $P(\text{draw}) = 0$. Hence, the binomial-based ISD calculation is not relevant for outcomes in which $P(\text{draw}) \neq 0$, regardless of the weighting of wins and draws, in terms of points assigned. Although CH (p. 331) observe that conventional usage of the binomial approximation "follows from an assumption that ties are one half a win and one half a loss", they later (p. 322) explicitly and correctly argue that it is erroneous to use the conventional binomial calculation, even in cases in which a draw has half the points value of a win.

Following directly from this, it is potentially confusing to label the pure binomial ISD, for which draws are ignored, as the '(1,0.5,0) approximation'. Calculating the pure binomial ISD is useful in assessing the bias from inappropriately using this measure, but

the pure binomial ISD does not give the relevant value of the ISD for a (1,0.5,0) points allocation for cases in which the probability of a draw, $P(\text{draw}) = d \neq 0$. If X_i is the absolute number of points accruing to team i in any game, and $P(\text{win}) = P(\text{loss}) = (1 - d)/2$, then

$$E(X_i) = [((1 - d)/2) \times 1] + (d \times 0.5) + [((1 - d)/2) \times 0] = 0.5,$$

$$V(X_i) = [((1 - d)/2)(1 - 0.5)^2] + d(0.5 - 0.5)^2 + [((1 - d)/2)(0 - 0.5)^2] = (1 - d)/4,$$

where $E(\cdot)$ and $V(\cdot)$ denote the expected value and variance, respectively. Hence, if Y_i is the total number of points accumulated by team i in a season, then

$$E(Y_i) = NE(X_i) = 0.5N,$$

$$V(Y_i) = V(\Sigma(X_i)) = N.V(X_i) = N(1 - d)/4.$$

CH assume $P(\text{win}) = 0.375$, $P(\text{draw}) = 0.25$, $P(\text{loss}) = 0.375$, based on the sample proportions of match outcomes for the top two tiers of the English Football League between the 1888-1889 and 2003-2004 seasons. Hence, for absolute points, $V(Y_i) = 0.188N$ and $\text{ISD}(1,0.5,0) = \sqrt{0.188N}$.⁴ This is *not* the value labeled $\text{ISD}(1,0.5,0)$ in Fort's (2007, p. 646) Equation (1), which reports the pure binomial ISD, $\sqrt{0.25N}$. This is in contrast to the values of $\text{ISD}(2,1,0)$ ($= \sqrt{0.75N}$) and $\text{ISD}(3,1,0)$ ($= \sqrt{1.734N}$) derived by CH (on p.332 and p.337, fn. 4) and set out in Fort's Equations (2) and (3), respectively, which *do* explicitly and correctly take into account the non-zero probability of a draw in the calculation of the ISD for absolute points.⁵ The entry in the *Points* column for (1,0.5,0) in Panel B reports the correct value derived above.

More importantly, all the ISD entries for *Points%* in Table 1, Panel A are incorrect. The source of this error is CH's (2006, fn. 4) assertion that "[i]f the percentage of total possible points is used to calculate the result of each team, then $0.5/\sqrt{N}$ is the appropriate correction". Fort (2007, p. 646) uses this result, referring to CH's fn. 4 as justification. However, CH's assertion is not correct.

To calculate the ISD, the relevant variance of points as a proportion of total available points, $V(\text{Points}\%)$, is determined by the generic formula $V(Y_i/\alpha N) = V(Y_i)/(\alpha N)^2$, where α is the number of points for a win (i.e., 1, 2 or 3). The results for the ISD for $d \neq 0$ and evaluated for $d = 0.25$ are as follows:

$$\text{ISD}(\text{Points}\%: 1,0.5,0) = \sqrt{(1-d)/4N} = \sqrt{0.188/N}, \quad (1)$$

$$\text{ISD}(\text{Points}\%: 2,1,0) = \sqrt{(1-d)/4N} = \sqrt{0.188/N}, \quad (2)$$

$$\text{ISD}(\text{Points}\%: 3,1,0) = \sqrt{[(1-d)(d+9)/4]/9N} = \sqrt{0.193/N}. \quad (3)$$

These formulae are the ISD percentage of points expressions analogous to the expressions for the ISD for actual points in Fort's Equations (2) and (3) and the corrected version of his Equation (1) discussed above.

IMPLICATIONS

Putting this all together, these results imply that the combined CH/Fort results (reproduced in Panel A of Table 1) are not all calculated on a comparable basis. The shaded ISD entries are based on the pure binomial assumption, whereas the non-shaded entries explicitly take into account the non-zero probability of a draw. This difference in treatment carries over to the calculation of the RSD. Panel B reports the relevant results

when *all* the ISD values appropriately take into account the assumption $P(\text{draw}) = 0.25$ and $P(\text{win}) = P(\text{loss}) = 0.375$.

The results in Panel B show that the RSD measures applied to percentages of points are not identical for any choice of points assignment scheme. RSD is the same for the (1,0.5,0) and (2,1,0) points schemes, for which the relative weight of wins and draws is equal, but is different for the (3,1,0) scheme. As noted above, this is also true in Panel A, because, even if the ISD is inappropriately assumed the same across all the points schemes (as in Panel A), the ASD applied to percentage of points is different for the (3,1,0) scheme (and this difference is not just due to rounding); hence, the RSD is also different. That said, the differences in the RSD applied to percentages across the points assignment schemes considered are not quantitatively significant.

Panel B also demonstrates that, if comparing like with like, there is no difference between the RSD for absolute points compared to percentages, for any particular points assignment scheme. The differences apparent in Panel A are artifacts of the inappropriate calculations for the ISD of the percentage of points for the shaded entries. Consequently, when draws are taken into account, there are no difficult context-dependent choices to make between the RSD based on absolute points versus percentages. The absolute points and percentage measures tell exactly the same story, regardless of whether the purpose of the analysis is tracking competitive balance over time or examining the effect of competitive balance measures on fan demand.

Fort calculates values for the RSD for both points and points percentages using points allocation systems (1,0.5,0) and (3,1,0) applied to outcomes for the English Premier League for seasons 1995-1996 to 2004-2005 and the National Hockey League in the US

for seasons 1989-1990 to 1998-1999. The former is re-examined to illustrate the changes that arise from applying the above analysis. Table 2 reports the results (updated to 2008-2009). In contrast to Fort's Table 2, the RSD *Points* and *Points%* results are the same for any chosen points assignment scheme, so are not reported separately. As noted above, the RSD results for the (3,1,0) scheme are different to those for the (2,1,0) and (1,0,5,0) schemes, but, as Figure 1 illustrates, the differences between the time-series plot of the RSD values for (3,1,0) compared to the other schemes (represented by the plot for (2,1,0)) are trivial. Figure 1 also plots the RSD based on the pure binomial formula for the ISD (labeled RSD_Bin) to illustrate the extent of underestimation of the degree of imbalance from using the latter compared to any of the correct RSD values. The extent of this understatement can be quantified by using the method outlined by Fort (2007, p.646) applied to the corrected ISD values in Table 1. $RSD_Bin/RSD(2,1,0) = 0.866$ and $RSD_Bin/RSD(3,1,0) = 0.896$; i.e., using the pure binomial ISD value leads to the RSD being underestimated by about 13% in a (2,1,0) league and by about 10% in a (3,1,0) league.⁶ However, ignoring the bias in the level, RSD_Bin tracks the movement over time of the other measures closely, supporting Fort's (2007) contention that binomial-based RSD values are unlikely to have given misleading results in existing time-series analyses of the effects of variation in competitive balance.⁷

The time span covered by this illustrative example is relatively short. One implication is that the number of teams does not change for the English Premier League over the ten-year period considered. However, CH (2006, Table A1) calculate RSD values using (2,1,0) and (3,1,0) points allocations for the top two divisions of the English Football League for selected years (at five-year intervals, as far as possible) spanning the period

1888 to 2003. In addition to the problems identified above, another relevant issue in this application is the variation over time in the number of teams in each division. The RSD is widely used to measure competitive balance because it takes into account the number of teams and the number of games played. However, the RSD has an upper bound, implied by the league's playing schedules, which is an increasing function of the number of teams or games played (Owen, 2010). Considering variation in the RSD relative to its upper bound can provide additional insights in comparisons of within-season competitive balance over time if the numbers of teams and/or games played are not constant, as in this case.

Focusing on the top division (known as 'League Division 1' up to the 1991-1992 season and the 'Premier League' from 1992-1993 to date), Figure 2 plots the corrected values of the RSD (based on equations (1)-(3) above) corresponding to CH's (2006) Table A1 for (3,1,0) and (2,1,0) points allocations.⁸ In line with the arguments above, but in contrast to the results in CH's Table A1, the RSD values are the same for points or absolute values. The differences between the RSD values from applying the (3,1,0) and (2,1,0) points allocation schemes are again very minor. Figure 2 also plots the relevant upper bound for RSD, RSD^{ub} , which depends on the number of teams, and a normalized measure of competitive balance labeled ASD^* ($= ASD/ASD^{ub}$).⁹ ASD^* is identical to RSD^* ($= RSD/RSD^{ub}$) and lies in the interval $[0, 1]$, with 0 representing perfect parity and 1 maximum imbalance (Owen, 2010). According to the RSD values, the degree of imbalance in the last three seasons appears to have reached levels about 25% worse than in the inaugural 1888-1889 season. However, RSD^{ub} is also about 25% higher in recent seasons compared to 1888-1889, so that, in relative terms, the peak level of imbalance in

2007-2008 is at about the same level as in the League's inaugural season. However, both the RSD and the normalized ASD* (= RSD*) measures display an upward trend in competitive imbalance in the post-war period, as discussed by CH (2006).

CONCLUSIONS

With some fine-tuning, the analysis in CH (2006) and Fort (2007) can be applied to calculate RSD measures of within-season competitive balance in leagues in which draws are possible outcomes. However, if the underlying ISD values are calculated appropriately, properly taking into account the non-zero probability of a draw, some of the conclusions emerging from the CH/Fort exchange are reversed. There is no difference, for any given points assignment scheme, between the RSD for end-of-season absolute points compared to percentages of points. Whether concerned with tracking competitive balance over time or examining the effects of balance on fan demand, there is consequently no substantive choice to be made between the absolute and percentage measures. In contrast, variations in the points assignment do give numerically different results for RSD (whether applied to percentages or absolute points) if the ratio of points for a win compared to a draw differs across the points assignments under comparison. However, in practice, for commonly used assignments, the numerical differences that occur are quantitatively relatively minor compared to the season-to-season variation in RSD and/or changes in the feasible range of values for RSD due to changes in the numbers of teams.

TABLE 1: ASD, ISD and RSD Comparisons, The Football League, 1888-1889

	(1,0.5,0)		(2,1,0)		(3,1,0)	
	<i>Points</i>	<i>Points%</i>	<i>Points</i>	<i>Points%</i>	<i>Points</i>	<i>Points%</i>
<i>A: CH/Fort combined Table 1</i>						
ASD	3.971	0.181	7.943	0.181	11.676	0.177
ISD	2.345	0.107	4.062	0.107	6.177	0.107
RSD	1.693	1.693	1.955	1.693	1.890	1.660
<i>B: ISD calculated using $P(win) = P(loss) = 0.375$ and $P(draw) = 0.25$</i>						
ASD	3.971	0.181	7.943	0.181	11.676	0.177
ISD	2.031	0.092	4.062	0.092	6.177	0.094
RSD	1.955	1.955	1.955	1.955	1.890	1.890

NOTE: In Panel A, results for (1,0.5,0) are from Fort (2007), Table 1; results for (2,1,0) are from CH (2006), Table 1 and Fort (2007), Table 1; results for (3,1,0) are from CH (2006), Table 1, with ISD values added.

TABLE 2: RSD Comparisons, English Premier League, 1995-1996 to 2008-2009

	<i>RSD(1,0.5,0)</i>	<i>RSD(2,1,0)</i>	<i>RSD(3,1,0)</i>
<i>Season</i>	<i>Points or Points%</i>	<i>Points or Points%</i>	<i>Points or Points%</i>
1995-1996	1.895	1.895	1.868
1996-1997	1.456	1.456	1.466
1997-1998	1.522	1.522	1.531
1998-1999	1.803	1.803	1.724
1999-2000	1.999	1.999	1.979
2000-2001	1.696	1.696	1.721
2001-2002	2.041	2.041	2.056
2002-2003	1.922	1.922	1.887
2003-2004	1.864	1.864	1.870
2004-2005	2.045	2.045	2.104
2005-2006	2.300	2.300	2.292
2006-2007	1.947	1.947	1.953
2007-2008	2.481	2.481	2.431
2008-2009	2.260	2.260	2.246

NOTE: Points data are from the Barclays Premier League, Official Website (<http://www.premierleague.com/page/Statistics/>)

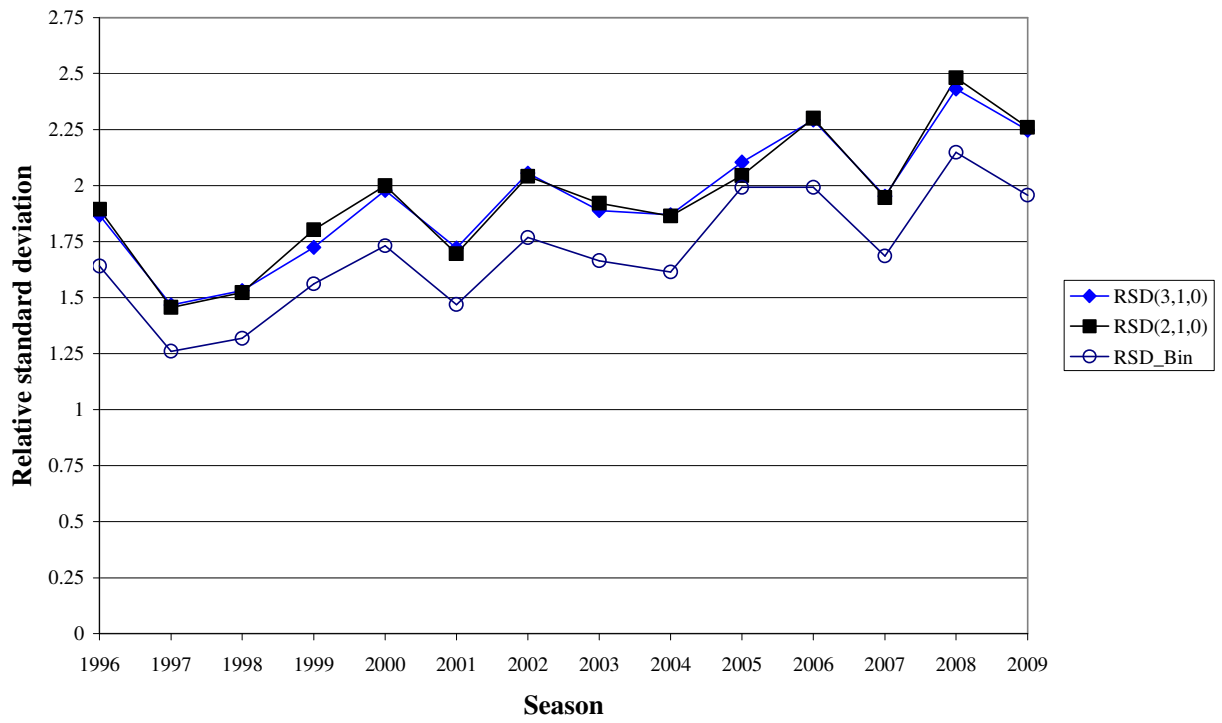


Figure 1: RSD for different points assignments for the English Premier League

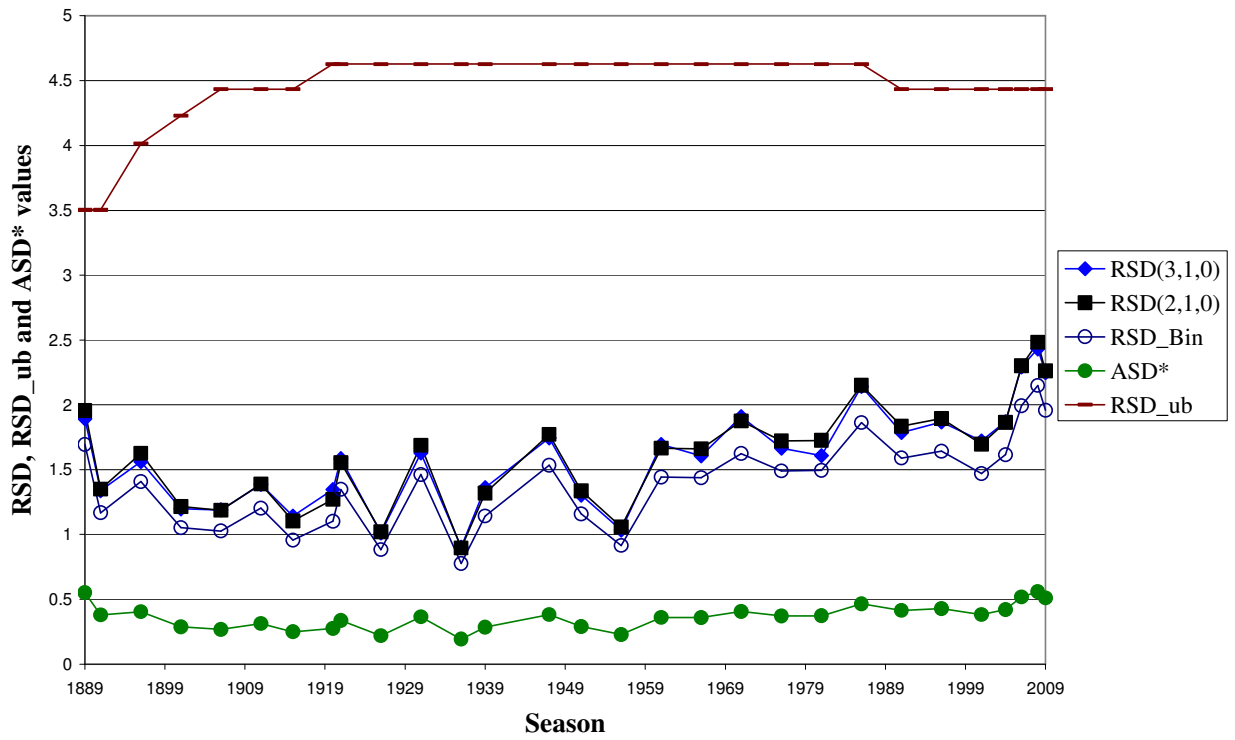


Figure 2: Long-run Movements in RSD and ASD* in the English Football League
(Selected Seasons)

REFERENCES

- Brandes, L., & Franck, E. (2007). Who made who? An empirical analysis of competitive balance in European soccer leagues. *Eastern Economic Journal*, 33, 379-403.
- Cain, L. P., & Haddock, D. D. (2006). Measuring parity: Tying into the idealized standard deviation. *Journal of Sports Economics*, 5, 169-185.
- Fort, R. (2006). Competitive balance in North American professional sports. In J. Fizez (Ed.), *Handbook of sports economics research* (pp. 190-206). Armonk, NY: M. E. Sharpe.
- 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.
- Noll, R. G. (1988). Professional basketball. Studies in Industrial Economics Paper No. 144. Stanford, CA: Stanford University.
- Owen, P. D. (2010). Limitations of the relative standard deviation of win percentages for measuring competitive balance in sports leagues. *Economics Letters*, forthcoming.
- Quirk, J., & Fort, R. D. (1992). *Pay dirt: The business of professional team sports*. Princeton, NJ: Princeton University Press.
- Scully, G. W. (1989). *The business of Major League Baseball*. Chicago, IL: University of Chicago Press.

¹ Fort (2007) criticizes CH for their confusing use of ‘ISD’ when referring to the relative standard deviation, RSD, throughout their paper. The current paper follows conventional usage, relabeling CH’s ‘ISD’ as RSD wherever appropriate.

² Note that, if M is the number of teams, the divisor used in calculating the ASD is $(M - 1)$ rather than M . The latter is often used in this context given that the sample mean is always 0.5, so it does not need to be estimated, for the (1,0.5,0) and (2,1,0) schemes. Whether M or $(M - 1)$ is used as the divisor does not affect any of the substantive arguments to follow.

³ As Fort (2007) emphasizes, 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), although its popularity was enhanced following its use in influential studies by Quirk and Fort (1992) and Fort and Quirk (1995).

⁴ Throughout, unless exact, the leading terms are expressed to three decimal places.

⁵ CH’s ISD calculation for absolute points for the (2,1,0) system is used by Brandes and Franck (2007) in deriving competitive balance measures for four European football leagues.

⁶ These results apply equally to the RSD values calculated for absolute points or percentages. The (2,1,0) result is a general result because $ASD(2,1,0)$ always equals $2 \times ASD(1,0.5,0)$ for absolute points. The (3,1,0) result depends on the proportions of wins and draws in the actual sample, as these affect the calculated ASD values; the result reported relates to the ASD values in Table 1 for the 1888-1889 season.

⁷ For the sample of seasons in Table 2, the correlation between RSD_Bin and RSD(3,1,0) is 0.974 and between RSD_Bin and RSD(2,1,0) is 0.983. The correlation between RSD(2,1,0) and RSD(3,1,0) is 0.993.

⁸ Data are from the RSSSF archive (<http://www.rsssf.com/engpaul/FLA/league.html>).

⁹ The upper bound of the RSD shown in Figure 2 is for the (2,1,0) scheme, obtained by dividing the upper bound of the ASD, ASD^{ub} , by the relevant value of the ISD. ASD^{ub} is calculated as $[M(M + 1)/(12(M - 1)^2)]^{0.5}$, derived as in Owen (2010) but with ASD calculated with $(M - 1)$ as the divisor. Note that ASD^* is invariant to whether the divisor is M or $(M - 1)$, as long as the ASD and ASD^{ub} are defined consistently.