June 2012

Simulating distributions of competitive balance measures in sports leagues: The effects of variation in competition design

Nicholas King, P. Dorian Owen*

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

---

ABSTRACT
This paper examines the distributional properties of the relative standard deviation (RSD) of points percentages, the most common measure of competitive balance (CB) in the sports economics literature, in comparison with other standard-deviation-based CB measures. Simulation methods are used to evaluate the effects of changes in competition design on the distributions of CB measures for different distributions of the strengths of teams in a league. The popular RSD measure performs as expected in cases of perfect balance but, if there is imbalance in team strengths, its distribution is very sensitive to changes in competition design. This has important implications for comparisons of RSD values for different sports leagues with different numbers of teams and/or games played.

Keywords:
Measuring competitive balance
Relative standard deviation
Idealized standard deviation
Sports economics
Simulation

JEL classification:
C63, D63, L83

* Corresponding author. Tel.: +64 3 479 8655
Email address: Dorian.Owen@otago.ac.nz

1. Introduction

Professional sports leagues "are in the business of selling competition on the playing field" (Fort and Quirk, 1995, p.1265). An appropriate degree of competitive balance, how evenly teams are matched, is central to this endeavour, as this affects the degree of uncertainty over the outcomes of individual matches and overall championships. According to the 'uncertainty of outcome hypothesis' (Rottenberg, 1956), higher levels of competitive balance, reflected in more uncertain outcomes, increase match attendances, television audiences and overall interest (Forrest and Simmons, 2002; Borland and Macdonald, 2003; Dobson and Goddard, 2011). However, in a free market, teams with greater financial resources (e.g., due to location in larger population centres or more lucrative sponsorship deals) can hire better players, improve team performance and increase their dominance. This undermines competitive balance and, hence, uncertainty of outcome, which, in turn, threatens the sustainability of the league because of excessive predictability of outcomes. Consequently, in sports antitrust cases, a lack of competitive balance is a widely used justification for restrictive practices (such as salary caps, player drafts and revenue sharing) that would not be countenanced in other industries (Fort and Quirk, 1995; Szymanski, 2003).

An important strand of the competitive balance literature involves assessing the extent of competitive balance, tracking its movements over seasons, and examining the effects of regulatory, institutional and other changes in business practices (Fort and Maxcy, 2003). Appropriate measurement of competitive balance is an important prerequisite for such analyses. Consequently, considerable effort has gone into measuring competitive balance.

Standard measures of dispersion, inequality and concentration, applied to end-of-season league outcomes such as win percentages or points percentages, are commonly used to measure competitive balance. These include the actual standard deviation and the relative standard deviation (Noll, 1988; Scully, 1989; Quirk and Fort, 1992; Fort and Quirk, 1995), the Gini coefficient (Fort and Quirk, 1995; Schmidt and Berri, 2001), the Herfindahl–Hirschman index (Depken, 1999), concentration ratios (Koning, 2000) and relative entropy (Horowitz, 1997), as well as a variety of other measures (Humphreys, 2002; Dobson and Goddard, 2011, Ch. 3).

From this menu, the most commonly used measure of competitive balance in the sports economics literature is the 'relative standard deviation' (also known as the 'ratio of standard deviations'), *RSD*, which is generally considered to be 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). The aim of this paper is to examine, using simulation methods, the distributional properties of the widely used *RSD* measure in comparison with other standard-deviation-based measures.

Despite their widespread use, relatively little is known about the properties of the sampling distributions of such measures in the context of comparing competitive balance in different sports leagues with different design characteristics. One distinctive characteristic of sports leagues is that their playing schedules impose restrictions on the distribution of wins; for example, teams cannot win matches in which they do not play. Playing schedules therefore constrain the range of feasible values of measures of competitive balance (Horowitz, 1997; Utt and Fort, 2002; Owen et al., 2007; Owen, 2010; Manasis et al., 2011).

Analyses of the implications of this have, so far, concentrated on deriving analytical expressions for lower and upper bounds of selected competitive balance measures. These lower and upper bounds are found to depend on the number of teams and/or the number of games played by each team (Depken, 1999; Owen et al., 2007; Owen, 2010; Manasis et al., 2011). This complicates the interpretation of balance measures, especially when comparisons (e.g., across different leagues or for the same league over time) involve different numbers of teams or games played, a situation that is extremely common. However, although these results provide useful information in defining the ranges of feasible values of the competitive balance measures, actual outcomes may be far from the extreme maximum values and, arguably, may be less affected by changes in competition design. For meaningful comparisons, it is important to have a clear idea not only of the location and range of feasible values of the chosen balance measure, but also how the distributions vary in response to different aspects of league design (e.g., the number of teams, the number of games played by each team or other variations in the playing schedules). Simulation provides an ideal approach to evaluate the effects of such changes in competition design on the distributions of different competitive balance measures for known distributions of the strengths (abilities) of the teams in the league.

Simulation methods have been applied to several different aspects of the analysis of sports leagues, including predicting the outcomes of matches and tournaments (e.g., Clarke, 1993; Koning et al., 2003), examining the effects of league or tournament design on specific measures of competitiveness or outcome uncertainty (Scarf et al., 2009; Puterman and Wang, 2011), assessing the effects on match attendance of changes in league structure or of equalizing playing talent across teams (Dobson et al., 2001; Forrest et al., 2005), illustrating the properties of a theoretical model of strategic behaviour in football (Dobson and Goddard, 2010), computing measures of match importance (Scarf and Shi, 2008), and generating ex ante measures of uncertainty of outcome (King et al., 2011). However, there has been surprisingly little use of simulation methods to examine the properties of measures of competitive balance. As far as we are aware, the only other study to do so is by Brizzi (2002); he considers normalized measures based on the standard deviation, the Gini coefficient, the mean absolute deviation and the mean letter spread, all applied to *points totals* rather than point proportions. However, he does not consider the popular *RSD* measure and simulates sampling distributions only for the case of exactly equally matched teams, i.e., the polar case of perfect competitive balance.

In section 2 we outline the various standard deviation-based measures of competitive balance that we consider in our analysis. Section 3 contains the details of the simulation design and Section 4 the results obtained on the distributions of the various competitive balance measures for different numbers of teams and games played. Concluding comments and implications for analysts of competitive balance, sports administrators and antitrust authorities are contained in Section 5.

2. Standard-deviation-based measures of competitive balance

In practice, measurement of competitive balance is complicated by its multidimensional nature. The different dimensions include the evenness of teams in individual matches, the distribution of wins or points across teams at the end of a season, the persistence of teams' record of wins or points across successive seasons, and the degree of concentration of championship wins over a number of seasons (Kringstad and Gerrard, 2007).

In this study we focus on different variants of the standard deviation of points ratios (or, equivalently, points percentages) based on end-of-season standings. In many sports, including association football, points are allocated for results other than wins (e.g., draws, or ties) and different points assignments are possible for wins, draws and losses; therefore, examining points ratios is more general than considering win proportions. We emphasize standard-deviation-based measures because of their popularity in the analysis of competitive balance in practice; however, in principle, the same approach can be used for any competitive balance measures.

The 'actual' standard deviation of points ratios, *ASD*, provides a simple, natural measure of the ex post variation in end-of-season points ratios. This can be calculated as

$$ASD = \sqrt{\sum_{i=1}^{N} [(P_i / T_i) - \overline{p}]^2 / (N-1)} \qquad (1)$$

where N equals the number of teams in the league, P_i and T_i are, respectively, the actual number of points accumulated and the maximum possible points attainable by team i in a season, $p_i = P_i / T_i$, and $\overline{p} = \sum_{i=1}^{N} p_i$ is the league's mean points ratio. Note that for scenarios in which draws are not possible or are worth half a win, then $\overline{p}$ always equals 0.5, so the mean points ratio does not need to be estimated and N can be used, instead of (N – 1), as the divisor in calculating the standard deviation.

Other things equal, the larger the dispersion of points ratios around the league mean in any season, the more unequal is the competition. However, although the mathematical expression for *ASD* does not depend explicitly on the number of games played by each team, *ASD* tends to decrease if teams play more games because the extent of random noise in the final outcomes is reduced (Leeds and von Allmen, 2008, p.156). Consequently, following Noll (1988) and Scully (1989), sports economists commonly use *RSD*, which compares *ASD* to a benchmark 'idealized standard deviation', *ISD*. The latter corresponds to the standard deviation of the outcome variable in a perfectly balanced league in which each team has an equal probability of winning each game.[1]

If draws are not possible, then *ISD* can be derived as the standard deviation of a binomially distributed random variable with a probability of success of 0.5 across independent trials; hence, $ISD = 0.5/G^{0.5}$, where G is the number of games played by each team (Fort and Quirk, 1995). If draws are possible, analogous expressions for *ISD* can be derived, allowing for different possible points assignments for wins, draws and losses (Cain and Haddock, 2006; Fort, 2007; Owen, 2012). A variant of *ISD* has also been proposed to allow for home advantage (Trandel and Maxcy, 2011).

RSD is the most widely used competitive balance measure in the sports economics literature (Fort, 2006a, Table 10.1); indeed, it has been described as "the tried and true" measure of within-season competitive balance (Utt and Fort, 2002, p.373). *RSD*, expressed as *ASD*/*ISD*, takes the value of unity if the league is perfectly balanced, with higher values representing greater levels of imbalance. However, Goosens (2006, p.87) criticises *RSD* for sometimes taking values *below* unity (i.e., $ASD < ISD$), implying "a competition that is more equal than when the league is perfectly balanced". Such an apparently contradictory

[1] 'Idealized', in this context, does not necessarily imply 'ideal', in the sense of an optimal value for *ASD*. Perfect balance and complete imbalance are polar cases, but the former is almost certainly not the optimal level of competitive balance. What constitutes an optimal level is an open question, and the answer may vary from one league to another; see Fort and Quirk's (2010, 2011) formalization of some of the factors involved.

interpretation is feasible because *ISD* represents an ex ante *probabilistic* benchmark rather than the *actual* ex post minimum for ASD, i.e., zero, corresponding to a situation in which all teams end up with the same number of points or points ratio. To avoid values below unity, Goosens (2006) advocates using a normalized standard deviation measure, here denoted *ASD**, which compares *ASD* with its maximum feasible value; i.e., $ASD^* = ASD/ASD^{ub}$, where ASD^{ub} is the upper bound of *ASD* corresponding to the ex post 'most unequal distribution' (Fort and Quirk, 1997; Horowitz, 1997; Utt and Fort, 2002).[2] 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.

Of more concern, Owen (2010) shows that *RSD* has an upper bound, RSD^{ub}, which is an increasing function of *N* and *K*; indeed, RSD^{ub} is much more sensitive to variation in the numbers of teams and games played than ASD^{ub}.[3] Consequently, comparing competitive balance across different leagues using *RSD* is likely to be more problematical than if *ASD* is used, especially if the relevant upper bounds of *RSD* (and hence the feasible range of outcomes) differ markedly due to differences in these parameters across leagues. Paradoxically, *RSD* is usually advocated for just such comparisons involving scenarios with different numbers of teams and/or games played (e.g., Leeds and von Allmen, 2008, pp.156-157; Fort, 2011, pp.167-169; Blair, 2012, pp.67-68). Variation in the feasible range of values for *RSD* for different *N* and *K* therefore represents a more fundamental

---

[2] Goosens (2006) calls *ASD** the "National Measure of Seasonal Imbalance". Brizzi (2002) also considers a normalized ASD measure, although he applies it to total points and inverts the measure to give an index of equality, $EQ = 1 - (ASD/ASD^{max})$, where ASD^{max} is the standard deviation of total points in the most unequal distribution.

[3] For the case of a balanced schedule with no draws (or with draws treated as half a win), $RSD^{ub} = 2[K(N + 1)/12]^{0.5}$, which is an increasing function of both *K* and *N*, compared to $ASD^{ub} = [(N + 1)/\{12(N - 1)\}]^{0.5}$, which, in the limit, as *N* increases, tends to $(1/12)^{0.5} = 0.289$ (Owen, 2010).

justification for the use of a normalized standard deviation measure, as *ASD** lies in the interval [0, 1].[4]

The coefficient of variation of points proportions, $CV = ASD/\overline{p}$ is another potential standard-deviation–based competitive balance measure. If $\overline{p}$ equals 0.5, then variation in *CV* is entirely due to variation in *ASD*, so nothing is gained by also considering *CV*. However, if draws are possible and are not considered as worth half a win, then $\overline{p}$ can vary across seasons, and *CV* is a feasible alternative.[5]

Overall, however, other than results on the upper bounds of *ASD* and *RSD*, little is known about the distributional properties of the various standard-deviation–based measures of competitive balance under different degrees of inequality in team abilities.

3. Simulation design

In order to examine the sensitivity of competitive balance measures to variation in basic parameters reflecting the format of a sports league, we consider simulated results for a wide range of different scenarios corresponding to different values of *N* (the number of teams), *K* (the number of rounds of matches), and different distributions of team strengths. We also allow for the effects of other features such as whether draws are feasible, existence of home advantage, and alternative points assignment schemes, all of which can affect match outcomes and hence, ultimately, end-of-season teams' points proportions and league rankings. Our focus is on examining how the distributions of standard-deviation-based measures (*ASD*, *RSD*, *ASD**, *CV*) of within-season competitive balance applied to end-of-

[4] Note that the normalized variant of *RSD*, defined as $RSD^* = RSD/RSD^{ub}$, is identical to $ASD/ASD^{ub} = ASD^*$.
[5] Note that $CV = [2\ IGE(2)]^{0.5}$, where *IGE*(2) is a member of the family of generalized entropy measures of inequality (Bajo and Salas, 2002).

season points proportions are affected by common changes in the design of the league, particularly variation in N and K.

Note that, in this paper, we consider only leagues with balanced schedules, in which each team plays every other team in the league the same number of times. Consequently, the number of games played by each team, G, is the same for all teams and equals $K(N - 1)$. This format is common in sports leagues, especially in European football (typically with $K = 2$). However, the simulation methods used can be adjusted to reflect the details of any unbalanced schedule of matches, in which a team may play some teams more frequently than others.

The simulation design includes the following components:

(i) An explicit characterization of the strengths of the teams in the league;

(ii) A model that generates match outcomes that allows for the effects of relative team strengths, home advantage (if included) and stochastic factors;

(iii) Combining generated outcomes for individual matches in the playing schedule of matches for the season, using a given points assignment scheme, to arrive at end-of-season points totals for each team in the league and hence values of the various competitive balance measures;

(iv) Repeating the generation of individual match outcomes in the playing schedule a large number of times to generate a distribution of values for each end-of-season competitive balance measure for given values of team strengths and competition design parameters;

(v) Rerunning the simulations for different assumptions about the distribution of teams' strengths and different N and K.

In reality, teams' strengths are unobservable, and so, therefore, is the degree of evenness in the teams' strengths. One of the main advantages of using a simulation approach is that team strengths (or abilities) are pre-specified and hence known. The predefined distribution of team strengths will be reflected in match outcomes and hence the final points proportions at the end of the season. Simulating a large number of seasons and calculating the end-of-season competitive balance measures allows us to examine the distributions of these measures for different underlying assumptions about the distribution of team strengths and for different competition design characteristics (especially N and K) in order to assess which measures provide the most reliable representation of the spread of team strengths across a range of different leagues in practice.

3.1 The simulation model for individual matches

To simulate match outcomes we use a framework similar to that of Stefani and Clarke (1992) and Clarke (1993, 2005). The outcome of each match is characterized by the home team's winning margin (points or goals scored by the home team less points or goals scored by the away team). The winning margin depends on the teams' relative playing strengths (or abilities) and the extent of home advantage (if any):

$$M_{ijm} = H + S_i - S_j \qquad (2)$$

where M_{ijm} is home team i's expected winning margin against away team j in match m; H is home advantage, and S_i is the strength rating for team i.[6]

[6] $M_{ijm} < 0$ corresponds to an expected win for the away team with a points margin of $|M_{ijm}|$.

To provide plausible values of team strengths and, where considered, home advantage, these were calibrated against actual results from English Premier League (EPL) football. The model in equation (2) was fitted to match results, season by season, for 10 seasons of the EPL (2001/2 to 2010/11), both allowing for a constant home advantage and with H omitted. The error in prediction is calculated as:

$$E_{ijm} = A_{ijm} - M_{ijm} \tag{3}$$

where A_{ijm} and E_{ijm} are, respectively, the actual match winning margin and the error in the prediction of the match outcome of home team i against away team j in match m.

Two approaches were adopted. For what we will denote the 'linear model', we calculated the strength ratings (and, if included, H) that minimize the sum of squared errors, $\sum E_{ijm}^2$, in match outcome predictions for each season (regardless of whether or not draws can occur, as there is no specific draws parameter in this method). We also fit a Bradley-Terry-type model (Bradley and Terry, 1952), calculating the strength ratings (and, if included, H) that maximize the likelihood of observing that season's results. Because the EPL data incorporate draws as possible match outcomes, the Bradley-Terry-type models are optimized with a draws parameter present. Because both models use the numerical difference in the competing teams' abilities as the basis of an expected match outcome, we can normalize ability ratings to have a mean of zero for both model types.[7]

Table 1 reports the mean, maximum and minimum ranges of strength ratings for ten EPL seasons under each model type. Plots of the ranked strength ratings for the 20 teams in the EPL were found to be approximately linear functions of the teams' rank, in each

[7] A team's strength rating can be interpreted as the expected points margin resulting from a match against an average team at a neutral venue (with no home advantage for either team).

season. We therefore maintain this observed linear pattern and range of fitted strength ratings in the EPL as a benchmark for formulating the different predetermined strength ratings in our simulations.

Two variants of the simulation model were considered to alleviate concerns that any results may be dependent on a specific simulation model. Simulating a match outcome involves adding a generated random error to the right-hand side of equation (2):

$$\mathrm{S}M_{ijm} = (H + S_i - S_j) + GE_m \tag{4}$$

where SM_{ijm} and GE_m are, respectively, the simulated winning margin and the generated error for home team *i*'s match against away team *j* in match *m*.

In the simulations, team strengths are fixed throughout each simulated season.[8] In the first simulation model (denoted the 'linear simulation model'), GE_m is drawn from a normally distributed random variable with a zero mean and standard deviation, $\sigma = 1.5$. The properties of *GE* are consistent with those of the actual errors obtained if the model in equations (2) and (3) is fitted, by minimizing the sum of squared errors in prediction, to the EPL data (2001/2 to 2010/11). Hence, the distribution of generated errors is approximately equal to the distribution of observed errors.

Positive values for simulated winning margins indicate home-team wins and negative values indicate away-team wins. The linear simulation model does not explicitly account for the possibility of draws, but they can be added to the model by finding the value of *d* (where $-d$ < match outcome < d) such that the proportion of match outcomes classified as draws is the same as observed in the EPL data.

[8] Updating of team strengths as the season progresses is possible (e.g., Clarke, 1993; King et al., 2011) but, in order to focus on the properties of the competitive balance measures, it makes more sense to work with team strengths that remain constant throughout the season.

In the second simulation model, we use a framework similar to that of Bradley and Terry (1952) in which match outcomes are obtained by generating probabilities of wins, draws, or losses occurring for the home team. Simulating match outcomes involves generating a random number uniformly distributed on the interval (0, 1); this number will fall into one of the three match outcome categories. For example, if the probabilities of a home win, a draw and an away win are, respectively, 0.4, 0.35 and 0.25, then a random number between 0 and 0.4 would indicate a home team win, between 0.4 and 0.75 a draw, and between 0.75 and 1 an away team win.

The probability of the home team winning in the EPL is approximately 0.6 (assuming draws are considered as half a win). This probability was also observed in the simulated data whenever home advantage was included in the match outcome generation process, whereas a value of approximately 0.5 was observed in the simulated data whenever home advantage was not included. This suggests that inclusion of the home advantage parameter gives realistic results while exclusion gives location-neutral results.

3.2 Specification of distributions of teams' strength ratings

Five strength rating distributions were used in the simulation. All have an average strength rating of zero and follow a linear pattern of ratings from the strongest to the weakest team. The five distributions have ranges (maximum strength – minimum strength) of 0, 1.25, 2.5, 3.75 and 5, i.e., covering a spectrum from teams of equal strength (perfect balance) through to very unequally distributed team strengths (a high degree of imbalance). Under both the Bradley-Terry-type and linear models, the optimized EPL strength ratings sit approximately near the middle of the five distributions specified. Hence, the results of the simulation cover both perfect equality and a high degree of inequality, and centre roughly where we would expect to see actual EPL results taking place.

The benchmark strength ratings generated from the EPL data are for N = 20. To construct strength rating distributions for different values of N but with the same level of 'strength inequality', we maintain a constant *range* of strength ratings but allow the slope of the plot of strength ratings against team number to change as we adjust the number of teams in the competition. This approach has two advantages. Firstly, maintaining a constant range of strength ratings means the probability of the strongest team beating the weakest team remains constant with changes in N. Secondly, an 'average'-strength team will have unchanged probabilities of beating both the strongest and weakest teams.[9]

3.3 Simulation parameters

Simulations were run for both the linear and Bradley-Terry-type model for the following parameters and variations in specification:

Range (maximum strength – minimum strength): 0, 1.25, 2.5, 3.75, 5

N (number of teams): 10, 15, 20, 25

K (number of rounds per season): 2, 4, 6, 8, 10

Draws possible or No draws

Home advantage: $H = 0$, $H \neq 0$

Points assignment: (2,1,0), (3,1,0)[10]

[9] Although not a specific design feature, the standard deviation of strength ratings is approximately preserved as N varies for each range of strength ratings considered.

[10] The notation (3,1,0), for example, represents three points for a win, one point for a draw, and zero points for a loss.

A thousand simulations were run for each combination of parameters and specification choices for each of the two simulation models. This gives a large number of different combinations of model type and parameter selection, so we report representative results and indicate any important deviations.

For each set of simulated end-of-season points proportions, we calculate *ASD*, *RSD* = *ASD*/*ISD*, $ASD^* = ASD/ASD^{ub}$, and $CV = ASD/\bar{p}$. For the calculation of *ASD*, we use *N* as the divisor in cases in which $\bar{p} = 0.5$, i.e., if draws are not allowed or if a (2,1,0) points allocation is used; otherwise we use $(N - 1)$, as in equation (1).

ISD is calculated as $0.5/G^{0.5}$, where $G = K(N - 1)$ is the number of games each team plays per season, in cases in which draws are not feasible and $H = 0$. If draws are feasible, then we use the corresponding $ISD = \sqrt{(1-d)/4G}$ for a (2,1,0) points assignment scheme, or $\sqrt{[(1-d)(d+9)/4]/9G}$ for a (3,1,0) points assignment, where *d* is the simulated probability of a draw in that season (Owen, 2012, equations (2′) and (3′) respectively). If $H \neq 0$, we use the *ISD* expression derived by Trandel and Maxcy (2011, p.10).[11]

ASD^{ub}, the upper bound of *ASD*, corresponding to the most unequal distribution of results possible given the (balanced) schedule of games played, is evaluated as $[(N + 1)/\{12(N - 1)\}]^{0.5}$, if *ASD* is calculated with *N* as the divisor, and $[N(N + 1)/(12(N - 1)^2)]^{0.5}$, if *ASD* calculated with $(N - 1)$ as the divisor (Owen, 2010).[12]

CV is considered only for cases where $\bar{p} \neq 0.5$.

[11] The home-advantage-corrected *ISD* calculated by Trandel and Maxcy (2011) treats a draw, if feasible, as half a win. However, we apply this form of *ISD* only to the case of home advantage with no draws.
[12] ASD^* is invariant to whether the divisor in *ASD* is *N* or $(N - 1)$, as long as *ASD* and ASD^{ub} are defined consistently; this applies regardless of the details of the points assignment. For the non-symmetric (3,1,0) assignment, even though the mean points proportion will, in practice, not equal 0.5, it is 0.5 for the most unequal distribution.

4. Simulation results

To compare the distributions of the various CB measures visually for different distributions of strength ratings, model types and competition specifications, we use kernel density estimates (using the Epanechnikov kernel function in Stata, version 11). These have the advantage of reflecting the continuous nature of the CB measures (unlike discrete representations such as histograms or box plots), allowing easily tuned smoothing of minor irregularities due to sampling variation (unlike box plots), and providing a clear representation of the tails of the distributions (unlike cumulative curves) (Cox, 2007).

We focus on illustrative results for the Linear model allowing for draws and home advantage, with a (3,1,0) points allocation. Figure 1 contains kernel densities for the four standard-deviation-based measures for the case of $N = 10$ and $K = 2$, for varying degrees of imbalance in the strength of the teams (from $R = 0$ through to $R = 5$). Each measure satisfies a basic minimum requirement for a reasonable indicator of competitive balance in the sense that increasing levels of imbalance in the distribution of team strengths are represented by rightward shifts in the densities.

Next, we examine the effects of changing the number of teams or the number of rounds for different levels of competitive balance. Figure 2 shows the effects of varying K, the number of rounds, keeping the number of teams, N, fixed at 10, for the case of perfect balance in team strengths, i.e. $S_i = 0$ for all i. The densities for *ASD* in panel (a) illustrate the concern expressed in section 2 that motivated the development of the *RSD*. *ASD* tends to decrease if teams play more games (in this case because there are more rounds of games against the other nine teams). This is reflected in the leftward shift in the densities as K increases; there is also a reduction in the variance of the observed *ASD* values,. This pattern is observable not just in *ASD* but also its normalized variant, *ASD**, in panel (c) and *CV* in panel (d). Therefore, adjustments to allow for the possible range of values *ASD* can take (as with *ASD**) or variations in the league's mean points ratio (as with *CV*) do not correct for these properties of *ASD*. In contrast, *RSD* does, in this case, correct for the shift in the density of *ASD* as *K* varies, as illustrated by the densities in panel (b). A similar pattern is obtained if *N* is allowed to vary for a constant value of *K*. In the case of perfect balance, *RSD* does exactly what it was designed to do by controlling for the effects of variations in the number of games played, either due to variation in the number of teams or the number of games each plays..

The next set of comparisons looks at the case of varying *K* while keeping *N* fixed at 10, but this time for the case of moderate imbalance in team strengths ($R = 2.5$), as approximately relevant for the EPL. The densities are shown in Figure 3. Again, the leftward shift, and reduced variance of the measures, is observed for *ASD*, *ASD** and *CV* as *N* increases. However, the striking feature of Figure 3 is the behaviour of the densities for *RSD* in panel (b). As *N* increases the density shifts rightward. The upward shift in simulated RSD values is consistent with the result that the upper bound of *RSD* is an increasing function of *N*, as discussed in section 2.

This feature of the density for *RSD* is even more strikingly displayed in Figure 4. Here we keep $N = 10$ throughout and increase the number of rounds, *K* (from 2 up to 10). The *ASD*, *ASD** and *CV* measures display similar properties to those in Figure 3, as discussed above. The density for *RSD* shifts even more markedly to the right,, reflecting the more dramatic increase in the number of games played as *K* increases.

Figure 5 compares the shift in the densities of *ASD** and *RSD* for the highest level of imbalance in the simulations (R = 5). For *ASD**, the densities are all now centred at higher values of *ASD** compared to the locations in Figures 3 and 4 (panel (c)), reflecting the higher degree of imbalance in team strengths, but still display some leftward drift and decreased variance as K increases. The densities for *RSD* show an even more marked rightward shift as K increases; between them these densities span values of *RSD* in the range (2, 6) but with very little overlap between the densities for different values of K.[13] This result is consistent with the result on the upper bound of *RSD*. As noted in section 2, *RSD*'s upper bound is strictly increasing in N and K. What these simulations illustrate is that the upward shift in the upper bound as N or K increases is not purely of hypothetical interest; the whole density shifts upwards for representative distributions of team strengths.

The notion that *RSD* controls for variation in season length and the number of teams therefore applies only in the polar case of perfect balance in team strengths, as illustrated in a comparison of Figure 2, panel (b) with the corresponding panels in Figures 3, 4 and 5. The sensitivity of *RSD* values to N and K also complicates interpretation of comparisons of *RSD* values from different leagues (with different N and/or K) or from the same league over time if there are variations in N and K. Direct comparisons of numerical values for *RSD* implicitly assume that *RSD* controls for season length so that differences in *RSD* values are supposed to reflect primarily differences in the degree of imbalance in playing strengths. However, in Figure 6 we plot densities for four different combinations of N, K and R: R = 1.25, with N = 10 and K = 8 (72 games for each team); R = 2.5, with N = 20 and K = 2 (38 games for each team), R = 2.5, with N = 10 and K = 4 (36 games for each team); and R = 5,

---

[13] This marked rightward shift in Figure 3 is therefore not an artefact of the way in which the strength distributions for a given R are adjusted as N is varied, because similar rightward shifts in the *RSD* densities are obtained when N is increased for given R and K, and when K is increased for given R and N.

with N = 10 and K = 2 (18 games for each team). The densities for *RSD* in panel (b) overlap considerably, suggesting that *RSD* is unable to distinguish the variation in competitive balance. It is possible that an additional 'correction' could be devised to allow for the variation in the number of games played for general departures from perfect balance. However, it is not immediately obvious what form this would take, as the problem appears to become more severe the more marked the departure from perfect balance, and such departures can occur in many different ways. In contrast, the densities for *ASD** in panel (a) do separate out the three different degrees of imbalance in playing strengths, at least for the limited comparisons involved. This is consistent with the *ASD* being less dramatically affected by variations in the number of games played by each team. Correction for the leftward shift of the *ASD* and *ASD** measures may therefore be a more fruitful avenue for further investigation.

5. Conclusion

Appropriate measurement of competitive balance is a cornerstone of the economic analysis of professional sports leagues, not least because of the importance of arguments about the extent of competitive imbalance as a justification for radical restrictive practices in sports antitrust cases. The most common measure of within-season competitive balance in the sports economics literature is the relative standard deviation (*RSD*) of points (or win) proportions; its popularity is based on the widespread belief that it controls for season length when making CB comparisons. Simulation methods are used to examine the effects of changes in competition design on the distributional properties of *RSD*, in comparison with other standard-deviation-based CB measures, for different distributions of the strengths of teams in a league. The popular *RSD* measure performs as expected in cases of perfect balance, but if there is imbalance in team strengths its distribution is very sensitive to changes in competition design. Without further correction, comparison of *RSD* for different sports leagues with different numbers of teams and/or games played can lead to misleading conclusions about the underlying degree of competitive balance. Other standard-deviation-based measures, although subject to upward bias if the number of games played is not large, are less sensitive to variations in competition design and appear to offer a more useful basis for cross-league comparisons in competitive balance.

Acknowledgements

We are grateful to Caitlin Owen for excellent research assistance. We also gratefully acknowledge financial support from a University of Otago Research Grant. An earlier version of this paper was presented in the University of Otago, Economics Research Seminar Series (June 2012).

Table 1

Range of fitted strength ratings from the English Premier League, seasons 2001/02 – 2010/11

Model	Mean range	Maximum range	Minimum range
Bradley-Terry ($H \neq 0$)	2.812	3.711	1.989
Bradley-Terry ($H = 0$)	2.682	3.633	1.839
Linear ($H \neq 0$)	2.275	3.174	1.700
Linear ($H = 0$)	2.275	3.175	1.700

Table 2

Strength rating distributions used for simulations, $N = 10$

Team	Distribution 1	Distribution 2	Distribution 3	Distribution 4	Distribution 5
1	0	0.625	1.25	1.875	2.5
2	0	0.486111	0.972222	1.458333	1.944444
3	0	0.347222	0.694444	1.041667	1.388889
4	0	0.208333	0.416667	0.625	0.833333
5	0	0.069444	0.138889	0.208333	0.277778
6	0	–0.06944	–0.13889	–0.20833	–0.27778
7	0	–0.20833	–0.41667	–0.625	–0.83333
8	0	–0.34722	–0.69444	–1.04167	–1.38889
9	0	–0.48611	–0.97222	–1.45833	–1.94444
10	0	–0.625	–1.25	–1.875	–2.5

Table 3

Strength rating distributions used for simulations, $N = 15$

Team	Distribution 1	Distribution 2	Distribution 3	Distribution 4	Distribution 5
1	0	0.625	1.25	1.875	2.5
2	0	0.535714	1.071429	1.607143	2.142857
3	0	0.446429	0.892857	1.339286	1.785714
4	0	0.357143	0.714286	1.071429	1.428571
5	0	0.267857	0.535714	0.803571	1.071429
6	0	0.178571	0.357143	0.535714	0.714286
7	0	0.089286	0.178571	0.267857	0.357143
8	0	0	0	0	0
9	0	–0.08929	–0.17857	–0.26786	–0.35714
10	0	–0.17857	–0.35714	–0.53571	–0.71429
11	0	–0.26786	–0.53571	–0.80357	–1.07143
12	0	–0.35714	–0.71429	–1.07143	–1.42857
13	0	–0.44643	–0.89286	–1.33929	–1.78571
14	0	–0.53571	–1.07143	–1.60714	–2.14286
15	0	–0.625	–1.25	–1.875	–2.5

Table 4

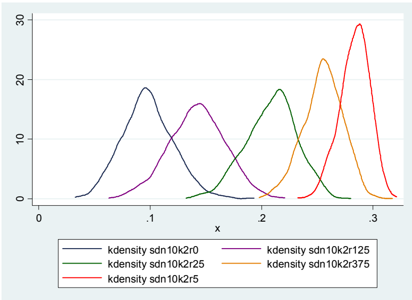
Strength rating distributions used for simulations, $N = 20$

Team	Distribution 1	Distribution 2	Distribution 3	Distribution 4	Distribution 5
1	0	0.625	1.25	1.875	2.5
2	0	0.559211	1.118421	1.677632	2.236842
3	0	0.493421	0.986842	1.480263	1.973684
4	0	0.427632	0.855263	1.282895	1.710526
5	0	0.361842	0.723684	1.085526	1.447368
6	0	0.296053	0.592105	0.888158	1.184211
7	0	0.230263	0.460526	0.690789	0.921053
8	0	0.164474	0.328947	0.493421	0.657895
9	0	0.098684	0.197368	0.296053	0.394737
10	0	0.032895	0.065789	0.098684	0.131579
11	0	−0.03289	−0.06579	−0.09868	−0.13158
12	0	−0.09868	−0.19737	−0.29605	−0.39474
13	0	−0.16447	−0.32895	−0.49342	−0.65789
14	0	−0.23026	−0.46053	−0.69079	−0.92105
15	0	−0.29605	−0.59211	−0.88816	−1.18421
16	0	−0.36184	−0.72368	−1.08553	−1.44737
17	0	−0.42763	−0.85526	−1.28289	−1.71053
18	0	−0.49342	−0.98684	−1.48026	−1.97368
19	0	−0.55921	−1.11842	−1.67763	−2.23684
20	0	−0.625	−1.25	−1.875	−2.5

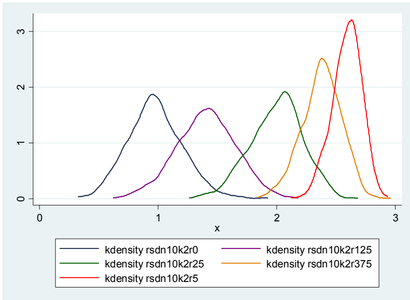
Table 5

Strength rating distributions used for simulations, $N = 25$

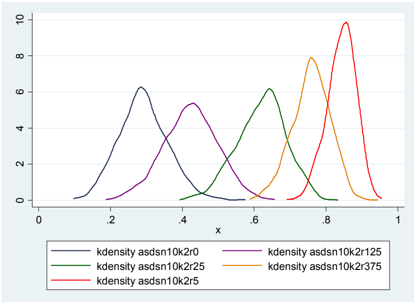
Team	Distribution 1	Distribution 2	Distribution 3	Distribution 4	Distribution 5
1	0	0.625	1.25	1.875	2.5
2	0	0.572917	1.145833	1.71875	2.291667
3	0	0.520833	1.041667	1.5625	2.083333
4	0	0.46875	0.9375	1.40625	1.875
5	0	0.416667	0.833333	1.25	1.666667
6	0	0.364583	0.729167	1.09375	1.458333
7	0	0.3125	0.625	0.9375	1.25
8	0	0.260417	0.520833	0.78125	1.041667
9	0	0.208333	0.416667	0.625	0.833333
10	0	0.15625	0.3125	0.46875	0.625
11	0	0.104167	0.208333	0.3125	0.416667
12	0	0.052083	0.104167	0.15625	0.208333
13	0	0	0	0	0
14	0	–0.05208	–0.10417	–0.15625	–0.20833
15	0	–0.10417	–0.20833	–0.3125	–0.41667
16	0	–0.15625	–0.3125	–0.46875	–0.625
17	0	–0.20833	–0.41667	–0.625	–0.83333
18	0	–0.26042	–0.52083	–0.78125	–1.04167
19	0	–0.3125	–0.625	–0.9375	–1.25
20	0	–0.36458	–0.72917	–1.09375	–1.45833
21	0	–0.41667	–0.83333	–1.25	–1.66667
22	0	–0.46875	–0.9375	–1.40625	–1.875
23	0	–0.52083	–1.04167	–1.5625	–2.08333
24	0	–0.57292	–1.14583	–1.71875	–2.29167
25	0	–0.625	–1.25	–1.875	–2.5



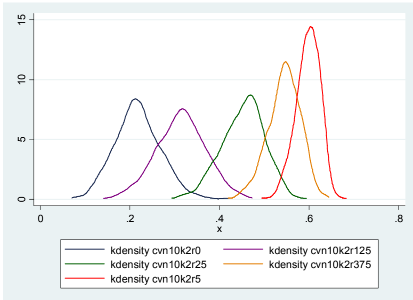
(a) *ASD*



(b) *RSD*

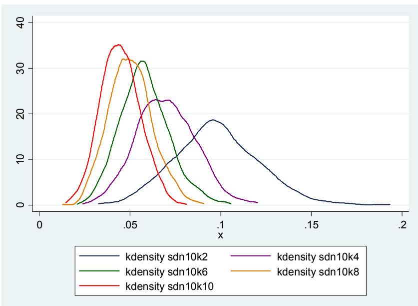


(c) *ASD**

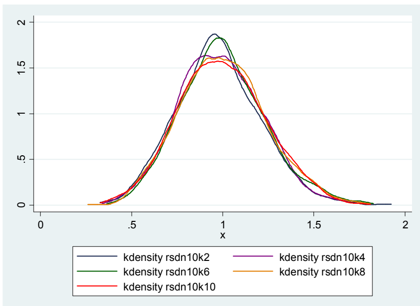


(d) *CV*

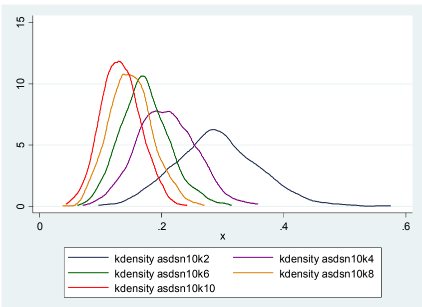
Fig. 1. Density functions of CB measures for different distributions of team strengths from $R = 0$ (perfect balance) to $R = 5$ (high degree of imbalance)



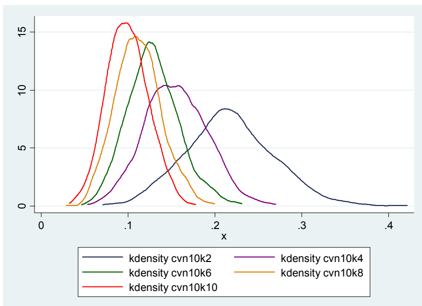
(a) *ASD*



(b) *RSD*

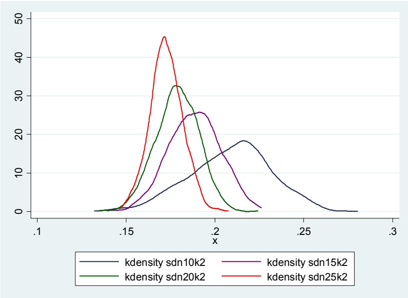


(c) *ASD**

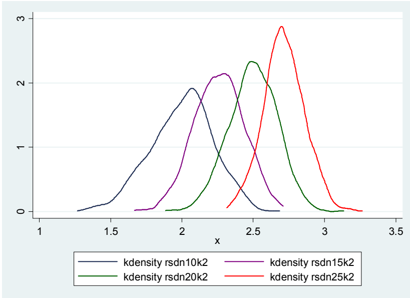


(d) *CV*

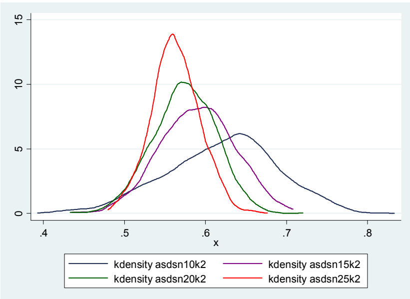
Fig. 2. Density functions of CB measures for $R = 0$ (perfect balance), $N = 10$, and different values of K



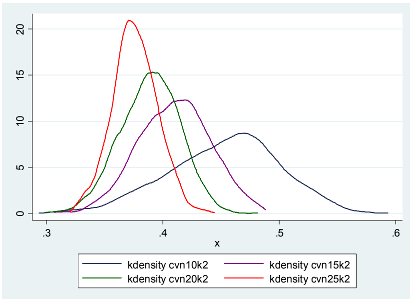
(a) *ASD*



(b) *RSD*

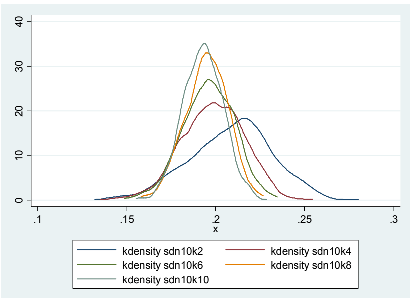


(c) *ASD**

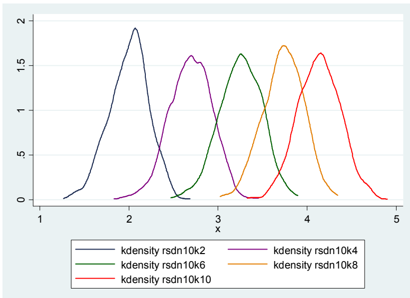


(d) *CV*

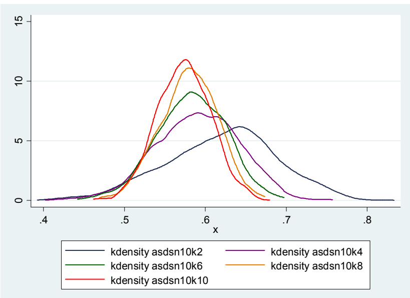
Fig. 3. Density functions of CB measures for $R = 2.5$ (moderate imbalance), $K = 2$, and different values of N



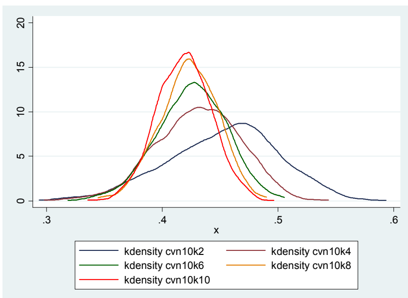
(a) *ASD*



(b) *RSD*

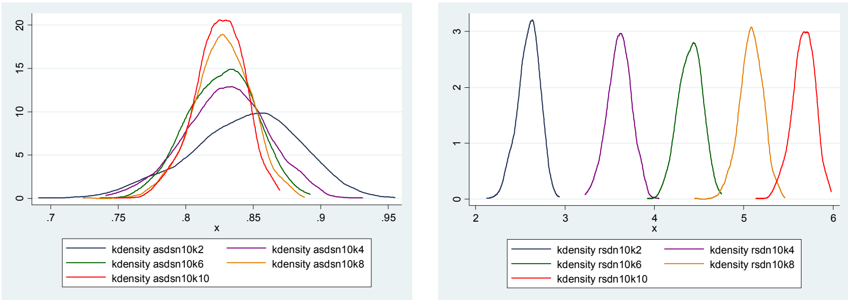


(c) *ASD**



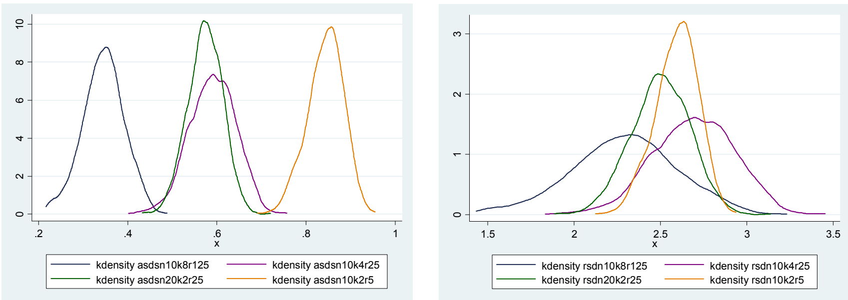
(d) *CV*

Fig. 4. Density functions of CB measures for $R = 2.5$ (moderate imbalance), $N = 10$, and different values of K



(a) *ASD* (b) *RSD*

Fig. 5. Density functions of CB measures for R = 5 (severe imbalance), N = 10, and different values of K



(a) *ASD* (b) *RSD*

Fig. 6. Density functions of CB measures for different combinations of R, N and K

References

Bajo, O., Salas, R., 2002. Inequality foundations of concentration measures: An application to the Hannah-Kay indices. Spanish Economic Review 4, 311-316.

Blair, R.D., 2012. Sports Economics. (Cambridge University Press, Cambridge).

Borland, J., MacDonald, R., 2003. Demand for sport. Oxford Review of Economic Policy 19, 478-502.

Bradley, R.A., Terry, M.E., 1952. Rank analysis of incomplete block designs: I. The method of paired comparisons. Biometrika 39, 324-345.

Brizzi, M., 2002. A class of indices of equality of a sport championship: Definition, properties and inference. In: Mrvar, A., Ferligoj, A. (Eds), Developments in Statistics (Faculty of Social Sciences (FDV), University of Ljubljana, Ljubljana, Slovenia), pp. 175-195.

Cain, L.P., Haddock, D.D., 2006. Measuring parity: Tying into the idealized standard deviation. Journal of Sports Economics 5, 169-185.

Clarke, S.R., 1993. Computer forecasting of Australian rules football for a daily newspaper. Journal of the Operational Research Society 44, 753-759.

Clarke, S.R., 2005. Home advantage in the Australian Football League. Journal of Sports Sciences 23, 375-385.

Cox, N.J., 2007. Kernel estimation as a basic tool for geomorphological data analysis. Earth Surface Processes and Landforms 32, 1902-1912.

Depken, C.A., II, 1999. Free-agency and the competitiveness of Major League Baseball. Review of Industrial Organization 14, 205-217.

Dobson, S., Goddard, J., 2011. The Economics of Football, 2nd Edition. (Cambridge University Press, Cambridge).

Dobson. S., Goddard, J., 2010. Optimizing strategic behaviour in a dynamic setting in professional team sports. European Journal of Operational Research 205, 661-669.

Dobson, S., Goddard, J., Wilson, J.O.S., 2001. League structure and match attendances in English rugby league. International Review of Applied Economics 15, 335-351.

Forrest, D., Simmons, R., 2002. Outcome uncertainty and attendance demand in sport: The case of English soccer. Journal of the Royal Statistical Society, Series D (The Statistician) 51, 229-241.

Forrest, D., Beaumont, J., Goddard, J., Simmons, R., 2005. Home advantage and the debate about competitive balance in professional sports leagues. Journal of Sports Sciences 23, 439-445.

Fort, R., 2006. Competitive balance in North American professional sports. In: Fizel, J. (Ed.), Handbook of Sports Economics Research (M.E. Sharpe, Armonk, NY), pp. 190-206.

Fort, R., 2007. Comments on "measuring parity". Journal of Sports Economics 8, 642-651.

Fort, R.D., 2011. Sports Economics (Third edition) (Pearson Prentice Hall: Upper Saddle River, NJ).

Fort, R., Maxcy, J., 2003. Comment on "Competitive balance in sports leagues: An introduction". Journal of Sports Economics 4, 154-160.

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.

Fort, R., Quirk, J., 2010. Optimal competitive balance in single-game ticket sports leagues. Journal of Sports Economics 11, 587-601.

Fort, R., Quirk, J., 2011. Optimal competitive balance in a season ticket league. Economic Inquiry 49, 464-473.

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.

Humphreys, B.R., 2002. Alternative measures of competitive balance in sports leagues. Journal of Sports Economics 3, 133-148.

Koning, R.H., 2000. Balance in competition in Dutch soccer. Journal of the Royal Statistical Society, Series D (The Statistician) 49, 419-431.

Koning, R.H., Koolhaas, M., Renes, G., Ridder, G., 2003. A simulation model for football championships. European Journal of Operational Research 148, 268-276.

King, N., Owen, P.D., Audas, R., 2011. Playoff uncertainty, match uncertainty and attendance at Australian National Rugby League matches. Economic Record (doi: 10.1111/j.1475-4932.2011.00778.x), Early View, 16p.

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).

Manasis, V., Avgerinou, V., Ntzoufras, I., Reade, J.J., 2011. Measurement of competitive balance in professional team sports using the normalized concentration ratio. Economics Bulletin 31, 2529-2540.

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

Owen, P.D., 2010. Limitations of the relative standard deviation of win percentages for measuring competitive balance in sports leagues. Economics Letters 109, 38-41.

Owen, P.D., 2012. Measuring parity in sports leagues with draws: Further comments. Journal of Sports Economics 13, 85-95.

Owen, P.D., Ryan, M., Weatherston, C.R., 2007. Measuring competitive balance in professional sports using the Herfindahl-Hirschman index. Review of Industrial Organization 31, 289-302.

Puterman, M.L., Wang, Q., 2011. Optimal dynamic clustering through relegation and promotion: How to design a competitive sports league. Journal of Quantitative Analysis in Sports 7(2), Article 7, http://www.bepress.com/jqas/vol7/iss2/7

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

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

Scarf, P.A., Shi, X., 2008. The importance of a match in a tournament. Computers and Operations Research 35, 2406-2418.

Scarf, P., Yusof, M.M., Bilbao, M., 2009. A numerical study of designs for sporting contests. European Journal of Operational Research 198, 190-198.

Schmidt, M.B., Berri, D.J., 2001. Competitive balance and attendance: The case of Major League Baseball. Journal of Sports Economics 2, 145-167.

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

Stefani, R., Clarke, S., 1992. Predictions and home advantage for Australian Rules Football. Journal of Applied Statistics 19, 251-261.

Szymanski, S., 2003. The economic design of sporting contests. Journal of Economic Literature 41, 1137-1187.

Trandel, G.A., Maxcy, J.G., 2011. Adjusting winning-percentage standard deviations and a measure of competitive balance for home advantage. Journal of Quantitative Analysis in Sports 7(1), Article 1, http://www.bepress.com/jqas/vol7/iss1/1

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