Competitive Balance Measures in Sports Leagues:
The Effects of Variation in Season Length

P. Dorian Owen and Nicholas King
Department of Economics, University of Otago

Address for correspondence:

Professor Dorian Owen
Department of Economics
University of Otago
PO Box 56
Dunedin
NEW ZEALAND
Email: dorian.owen@otago.ac.nz
Telephone: 64 3 479 8655

Competitive Balance Measures in Sports Leagues:
The Effects of Variation in Season Length*

P. Dorian Owen and Nicholas King
Department of Economics, University of Otago

Abstract

Appropriate measurement of competitive balance is a cornerstone of the economic analysis of professional sports leagues. We examine the distributional properties of the ratio of standard deviations (RSD) of points percentages, the most widely used measure of competitive balance in the sports economics literature, in comparison with other standard-deviation-based measures. Simulation methods are used to evaluate the effects of changes in season length on the distributions of competitive balance measures for different distributions of the strengths of teams in a league. The popular RSD measure performs as expected only in cases of perfect balance; if there is imbalance in team strengths, its distribution is very sensitive to changes in season length. This has important implications for comparisons of competitive balance for different sports leagues with different numbers of teams and/or games played. (JEL L83, D63, C63)

Address for correspondence:
Dorian Owen, Department of Economics, University of Otago, PO Box 56, Dunedin 9054, New Zealand.
Phone 64 3 479 8655, E-mail: Dorian.Owen@otago.ac.nz

I. 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 regarded as 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, more lucrative sponsorship deals or larger shares of television broadcasting rights) can hire better players, improve team performance and increase their dominance. This undermines competitive balance and hence uncertainty of outcome, which, in turn, can threaten 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 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. These measures include the actual standard deviation, the ratio of the actual and idealized standard deviations (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).[1]

From this menu, the ratio of standard deviations (*RSD*), also known as the 'relative standard deviation', is the most commonly used measure of competitive balance in the sports economics literature. This is generally considered the most useful "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. A 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 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). Analysis of the implications of this feature has, so far, concentrated on deriving analytical expressions for lower and upper bounds of selected competitive balance measures. These lower and upper

---

[1] See Dobson and Goddard (2011, Ch. 3) for a succinct review of these and other measures.

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, when comparisons (e.g., across different leagues or for the same league over time) involve different numbers of teams or games played, which is often the case. Although these results are suggestive of potential problems, actual outcomes may be far from the extremes of the feasible ranges and, arguably, may be less affected by changes in competition design. For meaningful comparisons, it is important to have a clear idea of how the complete 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 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 tournament design or promotion/relegation systems 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. 2012). 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 points ratios. 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 II we outline the various standard deviation-based measures of competitive balance considered in our analysis. Section III contains the details of the simulation design. The distributional properties of the various competitive balance measures, for different numbers of teams and games played, are reported in Section IV. Concluding comments and implications for analysts of competitive balance, sports administrators and antitrust authorities are contained in Section V.

II. 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,

4

or ties) and different points assignments are possible for wins, draws and losses; therefore, examining points ratios is more general than considering win ratios. We emphasize standard-deviation-based measures because of their popularity in the analysis of competitive balance in practice; 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

$$ASD = \sqrt{\sum_{i=1}^{N}[p_i - \overline{p}]^2 / (N-1)} \, . \tag{1}$$

N equals the number of teams in the league, $p_i = P_i/T_i$, where P_i and T_i are, respectively, the actual number of points accumulated and the maximum possible points total attainable by team *i* in a season, and $\overline{p} = \sum_{i=1}^{N} p_i / N$ is the league's mean points ratio. 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). Consequently, following Noll (1988) and Scully (1989), sports economists commonly use *RSD*, which compares *ASD* to a benchmark 'idealized standard deviation' (*ISD*) (e.g., Schmidt and Berri 2003;

Fort and Lee 2007). *ISD* 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.[2]

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 that allows for home advantage (Trandel and Maxcy 2011).

RSD, calculated as *ASD/ISD*, is the most widely used competitive balance measure in the sports economics literature (Fort 2006, 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* 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 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 points ratio. To avoid values below unity, Goosens (2006) advocates using a normalized standard deviation measure, here denoted *ASD**, which

---

[2] '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 the factors involved.

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).[3] 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, *RSD* has an upper bound, RSD^{ub}, which is an increasing function of

the number of teams, *N*, and the number of rounds of matches, *K* (Owen 2010); indeed,

RSD^{ub} is much more sensitive to variation in the numbers of games played than ASD^{ub}.[4]

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 commonly advocated for comparisons

that involve 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 potentially more

fundamental justification for the use of a normalized standard deviation measure that lies in

the interval [0, 1].[5]

The coefficient of variation of points ratios, $CV = ASD/\bar{p}$ is another potential

standard-deviation–based competitive balance measure. If $\bar{p}$ equals 0.5, then variation in

CV is entirely due to variation in *ASD*, so nothing is gained by also considering *CV*.

---

[3] Goosens (2006) calls *ASD** the "National Measure of Seasonal Imbalance". Brizzi (2002) also considers a normalized *ASD* measure (applied 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.
[4] For the case of a balanced schedule with no draws (or with draws treated 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).
[5] Note that the normalized variant of *RSD*, defined as $RSD^* = RSD/RSD^{ub}$, is identical to $ASD/ASD^{ub} = ASD^*$.

However, if draws are possible and are not considered as worth half a win, then $\bar{p}$ can vary across seasons, and *CV* is a feasible alternative.[6]

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 strengths; this provides the motivation for the current study.

III. SIMULATION DESIGN

In order to examine how the distributional properties of standard-deviation-based measures of within-season competitive balance are affected by changes in the format of a sports league, specifically season length, 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 allow for the existence of draws, home advantage, and alternative points assignment schemes, all of which can affect match outcomes and hence, ultimately, teams' end-of-season points ratios and league rankings.

We consider leagues with balanced schedules, in which each team plays every other team in the league the same number of times. Season length, i.e., the number of games played by each team, *G*, is the same for all teams and equals $K(N-1)$. This format is common, 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.

---

[6] 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).

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 allowing 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 ratios 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 inequality in their strengths. One of the main advantages of using a simulation approach is that team strengths are pre-specified and hence known. 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 assumptions about the distribution of team strengths and for season length (which depends on N and K) in order to assess which measures provide the most reliable representation of the imbalance in team strengths across a range of different leagues in practice.

Modelling Individual Match Outcomes

To model 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 and the extent of home advantage (if any):

$$M_{ijm} = H + S_i - S_j \tag{2}$$

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.[7]

To provide plausible values of team strengths and, where considered, home advantage, these are calibrated against actual results from English Premier League (EPL) football. The model in equation (2) is 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 are adopted. For what we denote the 'linear model', we calculate the strength ratings (and, if included, H) that minimize the sum of squared errors, $\sum\limits_{m} E_{ijm}^2$, in

---

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

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 (1952) model for paired comparisons to estimate the strength ratings and, if included, H. The probability of a home team win is based on the relative strength ratings of the two opposing teams. The probability that home team i beats away team j is given by $P_{win,i,j} = \exp(r_i)/[\exp(r_i) + \exp(r_j)]$, where r_i is the strength rating for team i ($i = 1, ..., N$). $P_{win,i,j}$ is a positive function of r_i, a negative function of r_j, and exhibits diminishing returns to relative ability. With no draws, the probability that home team i loses to away team j is given by $P_{lose,i,j} = 1 - P_{win,i,j}$. For ease of notation, denoting $R_i = \exp(r_i)$, $P_{win,i,j} = R_i/(R_i + R_j)$ and $P_{lose,i,j} = R_j/(R_i + R_j)$. Maximum likelihood estimates of the R_i values ($i = 1, ..., N$) are obtained by maximizing the log-likelihood function for the set of match outcomes for the season, and the r_i ratings obtained as $\ln(R_i)$. The Bradley-Terry model design is flexible and can incorporate a general home advantage, team-specific home advantages, drawn matches, or combinations of these (Rao and Kupper 1967; Agresti 1990; King 2011). Draws are common outcomes in the EPL data, so the Bradley-Terry models are optimized with a draws parameter present. Because the linear and Bradley-Terry models use the numerical difference in the competing teams' strengths as the basis of an expected match outcome, we can normalize ability ratings, so that the S_i ratings in equation (2) have a mean of zero for both model types.[8]

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 EPL teams are approximately linear functions of the teams' rank, in each season. We therefore

---

[8] 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).

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.[9]

Two variants of the simulation model are considered to alleviate concerns that any results may be dependent on a specific simulation model. In the first model (denoted the 'linear simulation model'), simulating a match outcome involves adding a generated random error to the right-hand side of equation (2):

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

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. 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 incorporated in 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.

---

[9] In the simulations, team strengths are fixed throughout each simulated season. Updating of team strengths as the season progresses is possible in this setup (e.g., Clarke 1993; King et al. 2012) but, in order to focus on the properties of the competitive balance measures, we assume team strengths remain constant throughout the season.

In the second simulation model, the Bradley-Terry model is used to generate probabilities of a win, draw, or loss for the home team, based on the relevant teams' strength ratings. Simulating match outcomes simply involves generating a random number, x, from a uniform distribution on the interval (0, 1), which determines the category of match outcome. For example, if the probabilities of a home win, a draw and an away win for a particular match are, respectively, 0.4, 0.35 and 0.25, then $0 \leq x < 0.4$ would indicate a home team win, $0.4 \leq x < 0.75$ a draw, and $0.75 \leq x < 1.0$ 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 is also observed in the simulated data whenever home advantage is included in the match outcome generation process, whereas a value of approximately 0.5 is observed in the simulated data whenever home advantage is not included. This suggests inclusion of the home advantage parameter gives realistic results, whereas exclusion gives location-neutral results.

Specification of Distributions of Teams' Strength Ratings

Five strength rating distributions are used in the simulations. All have an average strength rating of zero and follow a linear pattern of ratings, with teams equally spaced, from the strongest to the weakest team. The five distributions take values for the range, $R =$ (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). For both the fitted Bradley-Terry and linear models, the optimized EPL strength ratings have a range that approximately matches the middle ($R = 2.5$) 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 strength ratings for $N = 20$ are illustrated diagrammatically in Figure 1. 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 decrease as N increases. Details of the five strength rating distributions considered, for different values of N, are reported in Tables A1 to A4 in Appendix A. Once we depart from a perfectly balanced league, many different patterns of inequality in team strengths are possible, but the pattern we adopt 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.[10]

Simulation Parameters

Simulations were run for both the linear and Bradley-Terry-type model for the following parameters and variations in specification: range (R): 0, 1.25, 2.5, 3.75, 5; number of teams (N): 10, 15, 20, 25; number of rounds per season (K): 2, 4, 6, 8, 10; draws possible (with the probability of a draw approximately matching that in the EPL) or no draws; home advantage: $H = 0$, $H \neq 0$; and points assignment schemes: (2,1,0), (3,1,0).[11]

---

[10] 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.
[11] 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. However, the key results relating to the effects of variation in season length are not affected in any important way by variations in the other characteristics; to save space we therefore report representative results.

For each set of simulated end-of-season points ratios, we calculate *ASD*, *RSD* = *ASD*/*ISD*, *ASD** = *ASD*/*ASD*^{ub}, and *CV* = $ASD / \bar{p}$ (if $\bar{p} \neq 0.5$). 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* ≠ 0, we use the *ISD* expression derived by Trandel and Maxcy (2011, p.10).[12]

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* is calculated with (*N* − 1) as the divisor (Owen 2010).[13]

---

[12] The home-advantage-corrected *ISD* calculated by Trandel and Maxcy (2011) treats a draw, if feasible, as half a win. We apply this form of *ISD* only to the case of home advantage with no draws.

[13] *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, the mean points proportion will usually not equal 0.5 in practice; however, by construction, it is 0.5 for the most unequal distribution.

IV. SIMULATION RESULTS

To compare the distributions of the various competitive balance 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). We focus on presenting results for the Bradley-Terry model allowing for draws and home advantage, with a (3,1,0) points allocation. These are representative of the results obtained with other variations in parameters and specifications, the effects of which are minor compared to the dominant patterns documented below.[14]

Figure 2 contains kernel densities for *ASD** and *RSD* for the case of $N = 20$ and $K = 2$, for varying degrees of imbalance in the strength of the teams (from $R = 0$ through to $R = 5$). These measures satisfy a basic minimum requirement for a reasonable indicator of competitive balance: increasing levels of imbalance in the distribution of team strengths are represented by rightward shifts in the densities. A similar pattern is obtained for *ASD* and *CV*.

Next, we examine the effects of changing the number of teams or the number of rounds for different levels of competitive balance.[15] Figure 3 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 that motivated the development of *RSD*; *ASD*, on average, takes higher values and its density has a larger variance if teams play fewer games (in this case

---

[14] To demonstrate the robustness of key results to variations in model and parameters, the supporting information contains simulated densities for the linear model, with draws, home advantage, and a (3,1,0) points allocation (Appendix B) and the linear model, with no draws, no home advantage, and a (2,1,0) points allocation (Appendix C). Summary statistics (minimum, mean, maximum, and the 5th, 25th, 50th, 75th and 95th percentiles) for the simulated distributions for *RSD* and *ASD** are reported in Appendices D, E and F.

[15] The results for varying *N* and *K* are examined separately, rather than focusing purely on *G*, the number of games each team plays, because the upper bound of *ASD* is a decreasing function of *N* and not *K*, but the upper bound for *RSD* is an increasing function of both (Owen 2010).

because there are fewer rounds of games). As K increases, the densities for ASD shift to the left and the variances of the observed ASD values decrease. This pattern, analogous to positive 'small-sample bias' in estimating the degree of inequality in teams' playing strengths using ASD, is also observable in its normalized variant, ASD^*, in panel (c) and CV in panel (d). Therefore, adjustments to allow for the feasible range of values of ASD (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, in this case, RSD does correct for the shift in the density of ASD as K varies, as illustrated in panel (b) of Figure 3, which are approximately centred on unity. As shown in Figure 4, similar patterns are obtained if N is allowed to vary for a constant value of K.[16] Therefore, in the case of perfect balance, whereas ASD, ASD^* and CV exhibit upward bias in estimating competitive imbalance for shorter seasons, 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 N or K.

The next set of comparisons looks at the case of varying N while keeping K fixed at 2, 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 5. Again, a leftward shift and reduced variance of the density functions are observed for ASD^* as N increases. However, the striking feature of Figure 5 is the behaviour of the densities for RSD in panel (b). As N increases, the density functions shift to the right.

This feature of the density for RSD is even more strikingly displayed in Figure 6. Here we keep $N = 10$ throughout and increase the number of rounds, K (from 2 up to 10). ASD^*

---

[16] In Figure 3 and following Figures, we focus on the densities for RSD and ASD^*, as the densities for ASD and CV behave in a similar fashion to ASD^*.

displays similar properties to Figure 5, 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.

Figures 7 to 9 compare the properties of 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 4, 5 and 6, reflecting the higher degree of imbalance in team strengths, but again display leftward drift and decreased variance as *K* or *N* increases. The densities for *RSD* show an even more marked rightward shift as *N* or *K* increases.[17]

For higher levels of imbalance, the larger the number of games played, the smaller is the overlap between the *RSD* densities for different values of *K*. In Figure 9, with *R* = 5 and *N* = 20, there is no overlap at all in the sets of simulated *RSD* values when comparing the densities for *K* = 2, 4, 6, 8 and 10; see the maximum and minimum simulated values in Appendix Table D9. The rightward shift in the *RSD* densities as more games are played is consistent with the upper bound of *RSD* being a strictly increasing function of *N* and *K*, as discussed in section II. However, what the simulations show is that the *whole* density function of *RSD* shifts to the right, for representative distributions of team strengths.

The notion that *RSD* controls for variation in season length (whether due to changes in the number of teams or rounds) therefore applies *only* in the polar case of perfect balance in team strengths, as illustrated by a comparison of Figures 3 and 4, panel (b), with the corresponding panels in Figures 5 to 9.

---

[17] The marked rightward shifts in Figures 5 and 7 are 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 *K* is increased for given *R* and *N*.

The sensitivity of *RSD* values to *N* and *K* complicates interpretation of comparisons of *RSD* values from different leagues (with different *N* and/or *K*) or from the same league if there are variations in its *N* or *K* over time, as is commonly the case. Direct comparisons of numerical values for *RSD* implicitly assume that *RSD* controls for season length, so that differences in *RSD* values are presumed to reflect primarily differences in the degree of imbalance in playing strengths.

To illustrate the problems with this approach, in Figure 10 we plot densities for five 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); *R* = 2.5, with *N* = 20 and *K* = 4 (76 games for each team); and *R* = 5, with *N* = 10 and *K* = 2 (18 games for each team). The densities for *RSD* in panel (b) overlap considerably, and do not even preserve the correspondence between the left-to-right order of the densities' location and the values of *R*, suggesting that *RSD* is unable to identify the underlying levels of competitive balance. In contrast, the densities for *ASD** in Figure 10, panel (a) do separate out the three different degrees of imbalance in playing strengths, at least for the limited comparisons in this example. This is consistent with the *ASD* being less dramatically affected by variations in the number of games played by each team.

The location shifts discussed above are also illustrated in Table 2, which summarizes the mean values of the density functions for *RSD* and *ASD** for different values of *N*, *K* and *R*. In particular, for each value of *R*, the variation in mean values of *ASD** across different value of *N* and *K* is relatively limited (despite the short-season upward bias). In contrast,

for each value of *R* (apart from *R* = 0), there is considerably more variation in the mean values of *RSD* across different *N* and *K* combinations.

Not only do the values of *RSD* diverge as more games are played, but the rate of divergence appears to depend on the degree of imbalance in playing strengths. The rightward shift in the *RSD* densities, as more games are played, appears to become more marked the more severe the departure from perfect balance (for example, compare Figures 5(b) and 7(b), and Figures 6(b) and 8(b). Table 2 shows that the greater the degree of imbalance, the wider is the range of possible mean values of the *RSD* densities for different *N* and *K* combinations. This produces considerable overlap in *RSD* values consistent with different levels of underlying imbalance. These characteristics make it unlikely that any correction could be devised for *RSD*, based on the number of games played, that would work for general departures from perfect balance, because in practice such departures can occur in many different ways and their extent is unknown.

In contrast to the divergent behaviour of *RSD*, the reduction in the upward bias in *ASD*, *ASD** and *CV*, as season length increases, reflects a form of convergent updating of the estimate of the degree of imbalance in team strengths as more information becomes available from the additional games played. Correction for the overestimation of imbalance when applying *ASD* and *ASD** to leagues with short seasons is therefore likely to be a more fruitful avenue for further investigation.

V. 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 ratio of standard deviations (*RSD*) applied to points (or win) proportions; its popularity is based on the widespread belief that it controls for season length when making comparisons of competitive balance.

Simulation methods are used to examine the effects of changes in season length on the distributional properties of *RSD*, in comparison with other standard-deviation-based measures of competitive balance, for different distributions of the strengths of teams in a league. The popular *RSD* measure performs as expected in the polar case of perfect balance, which is rarely, if ever, observed in practice. However, if there is imbalance in team strengths, its distribution is very sensitive to changes in season length; for unchanged levels of imbalance in team strengths, *RSD* values essentially blow out as the number of games played by each team increases, and this divergence is more marked the further the departure of the underlying inequality of strengths from perfect balance.

Comparison of *RSD* for sports leagues with different numbers of teams and/or games played can therefore lead to misleading conclusions about the differences in the underlying degrees of competitive balance. Other standard-deviation-based measures, such as *ASD*, *ASD** and *CV*, although overestimating imbalance for shorter seasons, are much less sensitive to variations in season length and appear to offer a more useful basis for cross-league comparisons in competitive balance.

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

Means of the Density Functions for *RSD* and *ASD** for Varying Degrees of Imbalance

			RSD					ASD*				
			Degree of imbalance, R					Degree of imbalance, R				
N	K	G	0	1.25	2.5	3.75	5	0	1.25	2.5	3.75	5
10	2	18	0.96	1.29	1.84	2.22	2.47	0.29	0.39	0.57	0.70	0.80
10	4	36	0.96	1.57	2.45	3.04	3.18	0.20	0.33	0.53	0.68	0.78
10	6	54	0.96	1.84	2.94	3.69	4.16	0.16	0.32	0.52	0.67	0.77
10	8	72	0.95	2.03	3.35	4.25	4.78	0.14	0.30	0.52	0.67	0.77
10	10	90	0.95	2.24	3.73	4.73	5.35	0.13	0.30	0.51	0.67	0.77
15	2	28	0.95	1.40	2.05	2.55	2.85	0.24	0.35	0.53	0.68	0.77
15	4	56	0.96	1.74	2.78	3.52	3.98	0.17	0.31	0.51	0.66	0.76
15	6	84	0.95	2.01	3.34	4.27	4.85	0.14	0.29	0.50	0.65	0.76
15	8	112	0.95	2.29	3.83	4.91	5.58	0.12	0.29	0.50	0.65	0.76
15	10	140	0.95	2.51	4.28	5.49	6.22	0.11	0.28	0.50	0.65	0.75
20	2	38	0.95	1.49	2.26	2.83	3.20	0.21	0.33	0.52	0.66	0.76
20	4	76	0.95	1.90	3.11	3.95	4.48	0.15	0.30	0.50	0.65	0.75
20	6	114	0.95	2.21	3.74	4.79	5.45	0.12	0.28	0.49	0.64	0.75
20	8	152	0.95	2.51	4.29	5.52	6.29	0.11	0.28	0.49	0.64	0.75
20	10	190	0.94	2.79	4.78	6.15	7.02	0.09	0.28	0.49	0.64	0.75
25	2	48	0.95	1.59	2.46	3.10	3.51	0.19	0.32	0.51	0.65	0.75
25	4	96	0.96	2.03	3.37	4.32	4.93	0.14	0.29	0.49	0.64	0.75
25	6	144	0.94	2.39	4.10	5.26	6.00	0.11	0.28	0.49	0.64	0.74
25	8	192	0.95	2.72	4.70	6.07	6.93	0.10	0.27	0.48	0.64	0.74
25	10	240	0.95	3.01	5.24	6.77	7.73	0.08	0.27	0.48	0.64	0.74

Note: Means of simulated density functions are for the Bradley-Terry model, with draws, home advantage, and a (3,1,0) points allocation. N is the number of teams, K the number of rounds of games, and G the number of games played by each team. R is the range of playing strengths of teams (maximum – minimum).

FIGURE 1
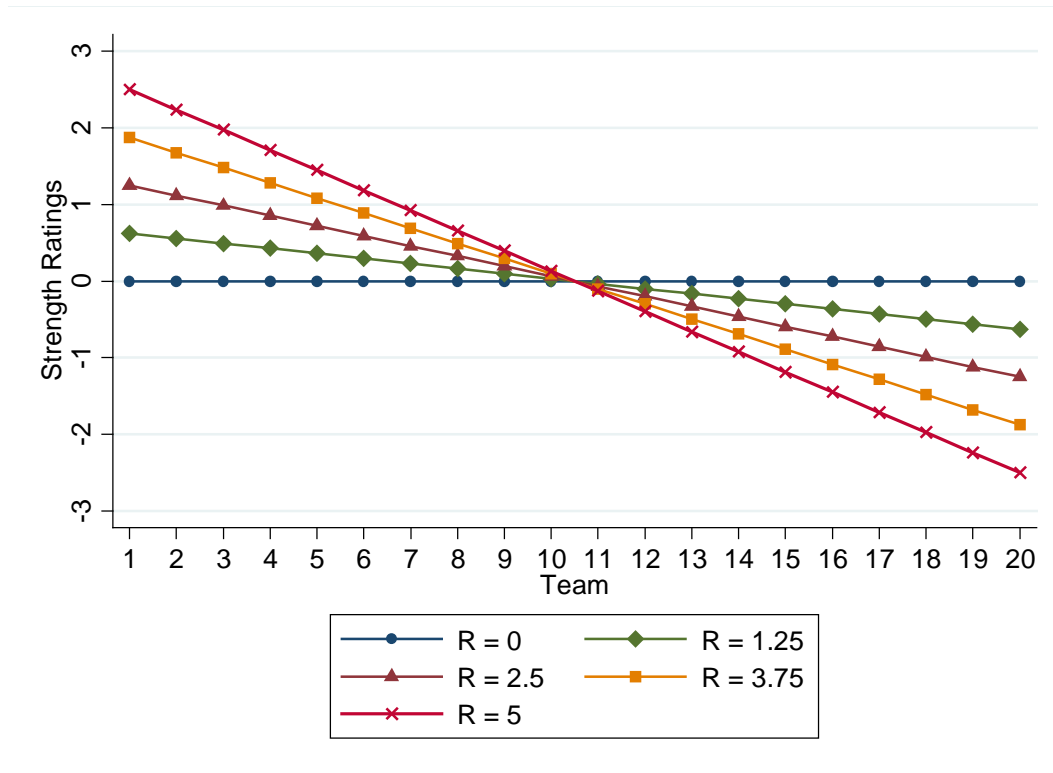Strength rating distributions used for simulations, $N = 20$

FIGURE 2
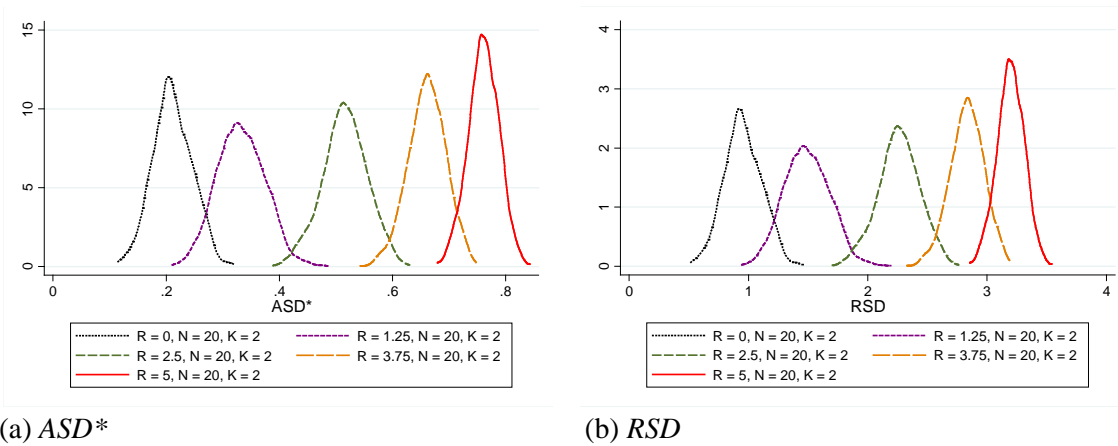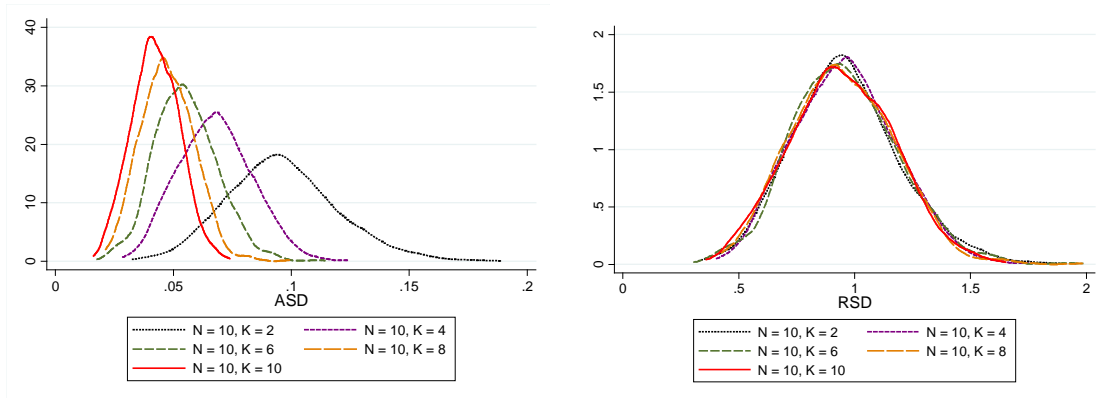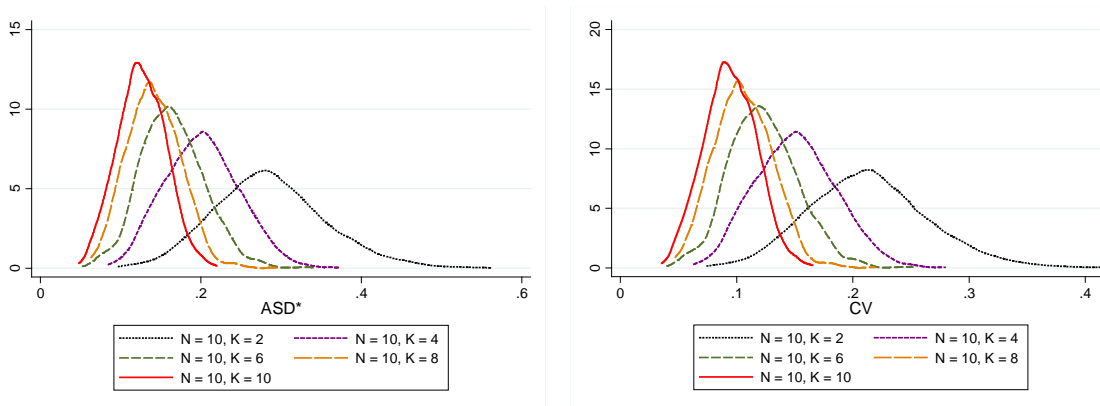Density functions of balance measures for varying degrees of imbalance

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

FIGURE 3

Density functions of balance measures for $R = 0$ (perfect balance), $N = 10$, varying K

(a) *ASD*

(b) *RSD*

(c) *ASD**

(d) *CV*

FIGURE 4

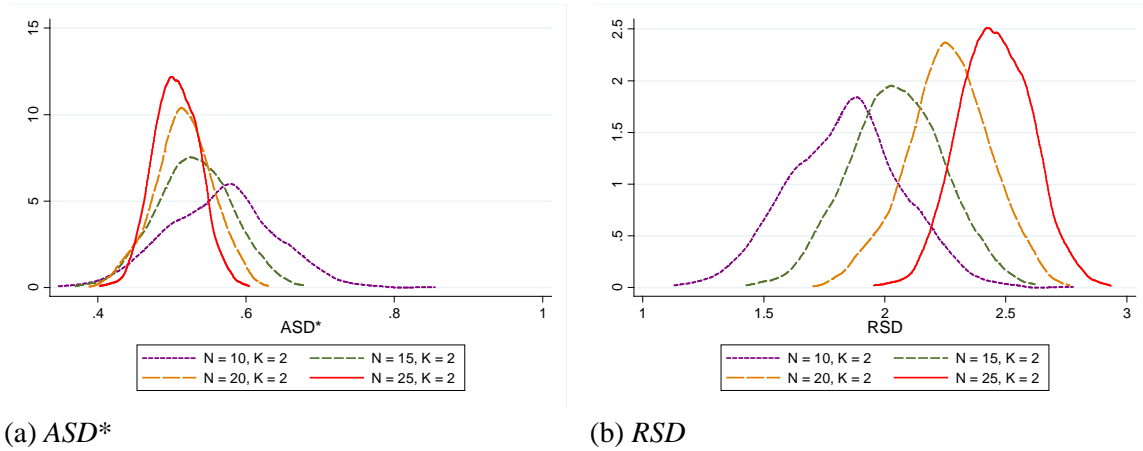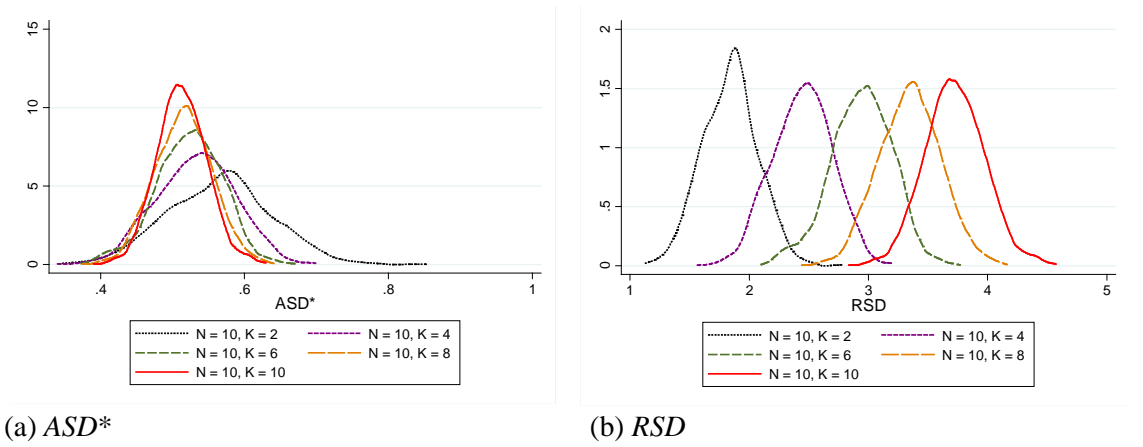Density functions of balance measures for $R = 0$ (perfect balance), $K = 2$, varying N
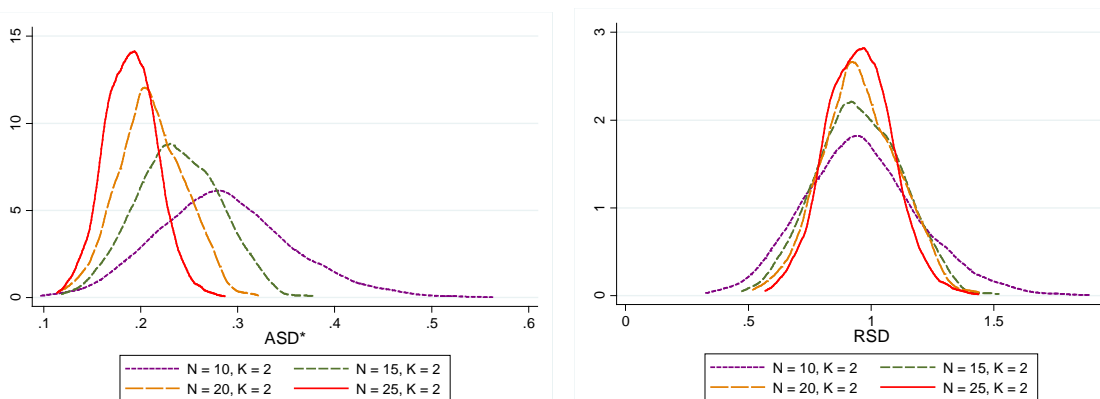
(a) *ASD**

(b) *RSD*

FIGURE 5
Density functions of balance measures for $R = 2.5$ (moderate imbalance), $K = 2$, varying N

(a) *ASD**

(b) *RSD*

FIGURE 6
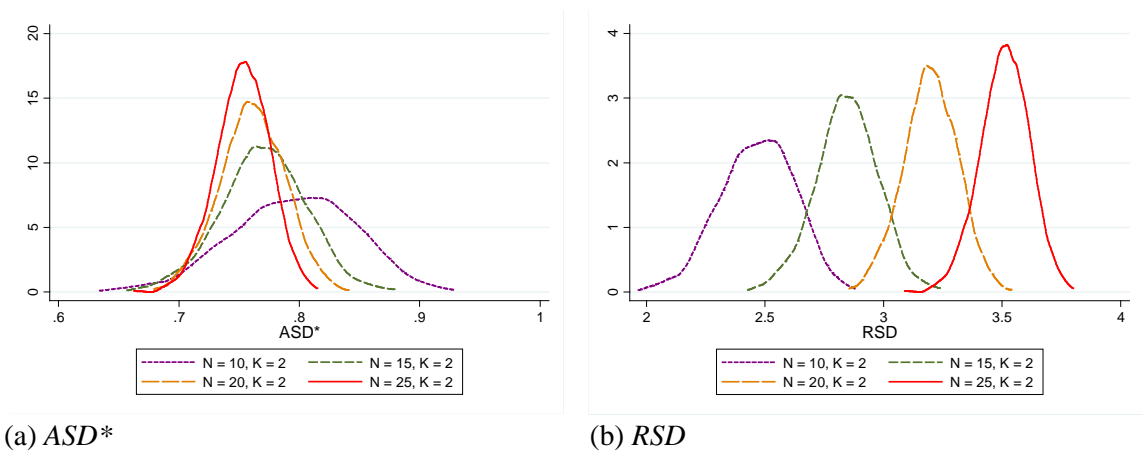Density functions of balance measures for $R = 2.5$ (moderate imbalance), $N = 10$, varying K

(a) *ASD**

(b) *RSD*

FIGURE 7
Density functions of balance measures for $R = 5$ (severe imbalance), $K = 2$, varying N

(a) *ASD**

(b) *RSD*

FIGURE 8

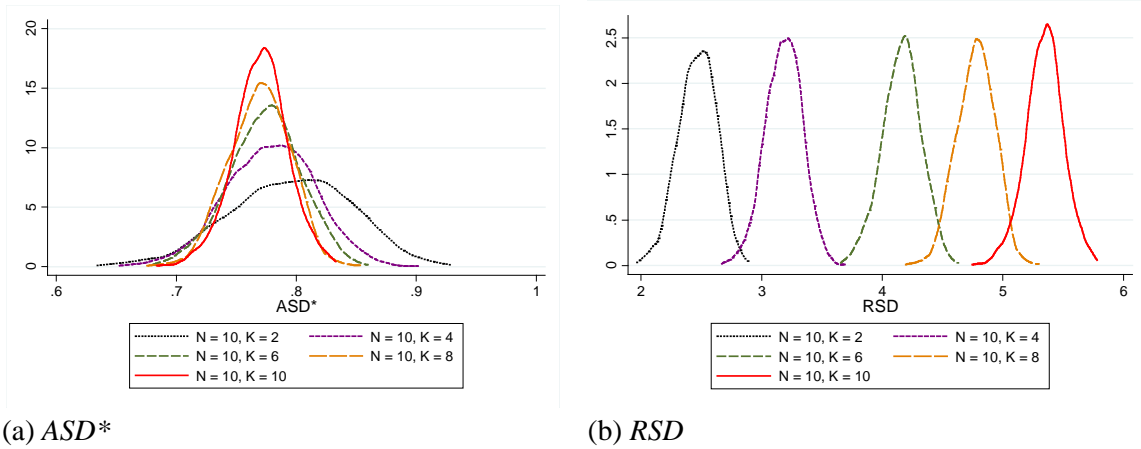Density functions of balance measures for $R = 5$ (severe imbalance), $N = 10$, varying K

(a) *ASD**

(b) *RSD*

FIGURE 9

Density functions of balance measures for $R = 5$ (severe imbalance), $N = 20$, varying K

(a) *ASD**

(b) *RSD*
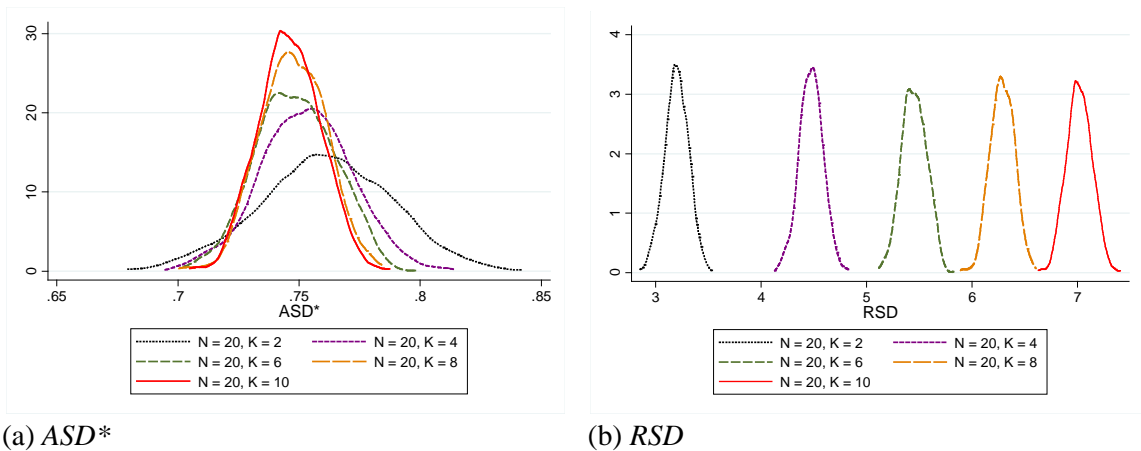
FIGURE 10

Density functions of balance measures for different combinations of R, N and K

(a) *ASD**

(b) *RSD*
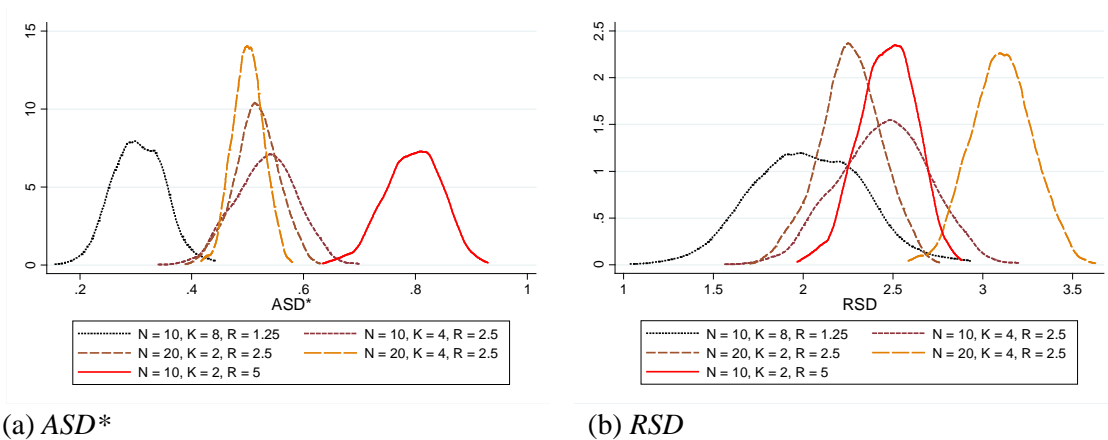
REFERENCES

Agresti, A. *Categorical Data Analysis*. New York: Wiley, 1990.

Bajo, O., and R. Salas. "Inequality Foundations of Concentration Measures: An Application to the Hannah-Kay Indices." *Spanish Economic Review*, 4, 2002, 311-16.

Blair, R. D. *Sports Economics*. Cambridge: Cambridge University Press, 2012.

Borland, J., and R. Macdonald. "Demand for Sport." *Oxford Review of Economic Policy*, 19, 2003, 478-502.

Bradley, R. A., and M. E. Terry. "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons." *Biometrika*, 39, 1952, 324-45.

Brizzi, M. "A Class of Indices of Equality of a Sport Championship: Definition, Properties and Inference," in *Developments in Statistics*, edited by A. Mrvar and A. Ferligoj. Ljubljana, Slovenia: Faculty of Social Sciences (FDV), University of Ljubljana, 2002, 175-95.

Cain, L. P., and D. D. Haddock. "Measuring Parity: Tying Into the Idealized Standard Deviation." *Journal of Sports Economics*, 5, 2006, 169-85.

Clarke, S. R. "Computer Forecasting of Australian Rules Football for a Daily Newspaper." *Journal of the Operational Research Society*, 44, 1993, 753-59.

Clarke, S. R. "Home Advantage in the Australian Football League." *Journal of Sports Sciences*, 23, 2005, 375-85.

Depken, C. A., II. "Free-Agency and the Competitiveness of Major League Baseball." *Review of Industrial Organization*, 14, 1999, 205-17.

Dobson. S., and J. Goddard. "Optimizing Strategic Behaviour in a Dynamic Setting in Professional Team Sports." *European Journal of Operational Research*, 205, 2010, 661-69.

Dobson, S., and J. Goddard. *The Economics of Football*. 2nd ed. Cambridge: Cambridge University Press, 2011.

Dobson, S., J. Goddard, and J.O.S. Wilson. "League Structure and Match Attendances in English Rugby League." *International Review of Applied Economics*, 15, 2001, 335-51.

Forrest, D., and R. Simmons. "Outcome Uncertainty and Attendance Demand in Sport: The Case of English Soccer." *Journal of the Royal Statistical Society*, *Series D* (*The Statistician*), 51, 2002, 229-41.

Forrest, D., J. Beaumont, J. Goddard, and R. Simmons. "Home Advantage and the Debate about Competitive Balance in Professional Sports Leagues." *Journal of Sports Sciences*, 23, 2005, 439-45.

Fort, R.. "Competitive Balance in North American Professional Sports," in *Handbook of Sports Economics Research*, edited by J. Fizel. Armonk, NY: M.E. Sharpe, 2006, 190-206.

Fort, R. "Comments on "Measuring Parity"." *Journal of Sports Economics*, 8, 2007, 642-51.

Fort, R. D. *Sports Economics*. 3rd ed. Upper Saddle River, NJ: Pearson Prentice Hall, 2011.

Fort, R. and Y. H. Lee. "Structural Change, Competitive Balance, and the Rest of the Major Leagues. *Economic Inquiry*, 45, 2007, 519-32.

Fort, R., and J. Maxcy. "Comment on "Competitive Balance in Sports Leagues: An Introduction"." *Journal of Sports Economics*, 4, 2003, 154-60.

Fort, R., and J. Quirk. "Cross-Subsidization, Incentives, and Outcomes in Professional Team Sports Leagues." *Journal of Economic Literature*, 33, 1995, 1265-99.

Fort, R., and J. Quirk. "Introducing a Competitive Economic Environment into Professional Sports," in *Advances in the Economics of Sport*, *Volume 2*, edited by W. Hendricks. Greenwich, CT: JAI Press, 1997, 3-26.

Fort, R., and J. Quirk. "Optimal Competitive Balance in Single-Game Ticket Sports Leagues." *Journal of Sports Economics*, 11, 2010, 587-601.

Fort, R., and J. Quirk. "Optimal Competitive Balance in a Season Ticket League." *Economic Inquiry*, 49, 2011, 464-73.

Goossens, K. "Competitive Balance in European Football: Comparison by Adapting Measures: National Measure of Seasonal Imbalance and Top3." *Rivista di Diritto ed Economia dello Sport*, 2, 2006, 77-122.

Horowitz, I. "The Increasing Competitive Balance in Major League Baseball." *Review of Industrial Organization*, 12, 1997, 373-87.

Koning, R. H. "Balance in Competition in Dutch Soccer." *Journal of the Royal Statistical Society*, *Series D* (*The Statistician*), 49, 2000, 419-31.

Koning, R. H., M. Koolhaas, G. Renes, and G. Ridder. "A Simulation Model for Football Championships." *European Journal of Operational Research*, 148, 2003, 268-76.

King, N. "The Use of Win Percentages for Competitive Balance Measures: An Investigation of How Well Win Percentages Measure Team Ability and the

Implications for Competitive Balance Analysis." MCom thesis, University of Otago, 2011.

King, N., P. D. Owen, and R. Audas. "Playoff Uncertainty, Match Uncertainty and Attendance at Australian National Rugby League Matches." *Economic Record*, 88, 2012, 262-77.

Kringstad, M., and B. Gerrard. "Beyond Competitive Balance," in *International Perspectives on the Management of Sport*, edited by M. M. Parent and T. Slack. Burlington, MA: Butterworth-Heinemann, 2007, 149-72.

Leeds, M., and P. von Allmen. *The Economics of Sport*. 3rd ed. Boston, MA: Pearson Addison Wesley, 2008.

Manasis, V., V. Avgerinou, I. Ntzoufras, and J. J. Reade. "Measurement of Competitive Balance in Professional Team Sports Using the Normalized Concentration Ratio." *Economics Bulletin*, 31, 2011, 2529-40.

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

Owen, P. D. "Limitations of the Relative Standard Deviation of Win Percentages for Measuring Competitive Balance in Sports Leagues." *Economics Letters*, 109, 2010, 38-41.

Owen, P. D. "Measuring Parity in Sports Leagues with Draws: Further Comments." *Journal of Sports Economics*, 13, 2012, 85-95.

Owen, P. D. et al. "Measuring Competitive Balance in Professional Sports Using the Herfindahl-Hirschman Index." *Review of Industrial Organization*, 31, 2007, 289-302.

Puterman, M. L., and Q. Wang. "Optimal Dynamic Clustering Through Relegation and Promotion: How to Design a Competitive Sports League." *Journal of Quantitative Analysis in Sports*, 7(2), 2011, Article 7, http://www.bepress.com/jqas/vol7/iss2/7

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

Rao, P. V., and L. L. Kupper. "Ties in Paired-Comparison Experiments: A Generalization of the Bradley-Terry Model." *Journal of the American Statistical Association*, 62, 1967, 194-204.

Rottenberg, S. "The Baseball Players' Labor Market." *Journal of Political Economy*, 64, 1956, 242-58.

Scarf, P. A., and X. Shi. "The Importance of a Match in a Tournament." *Computers and Operations Research*, 35, 2008, 2406-18.

Scarf, P., M. M. Yusof, and M. Bilbao. "A Numerical Study of Designs for Sporting Contests." *European Journal of Operational Research*, 198, 2009, 190-98.

Schmidt, M. B., and D. J. Berri. "Competitive Balance and Attendance: The Case of Major League Baseball." *Journal of Sports Economics*, 2, 2001, 145-67.

Schmidt, M. B., and D. J. Berri. "On the Evolution of Competitive Balance: The Impact of an Increasing Global Search." *Economic Inquiry*, 41, 2003, 692-704.

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

Stefani, R., and S. Clarke. "Predictions and Home Advantage for Australian Rules Football." *Journal of Applied Statistics*, 19, 1992, 251-61.

Szymanski, S. "The Economic Design of Sporting Contests." *Journal of Economic Literature*, 41, 2003, 1137-87.

Trandel, G. A., and J. G. Maxcy. "Adjusting Winning-Percentage Standard Deviations and a Measure of Competitive Balance for Home Advantage." *Journal of Quantitative Analysis in Sports*, 7(1), 2011, Article 1, http://www.bepress.com/jqas/vol7/iss1/1

Utt, J., and R. Fort. "Pitfalls to Measuring Competitive Balance with Gini Coefficients." *Journal of Sports Economics*, 3, 2002, 367-73.