Research article

Matthew Doria and Barry Nalebuff*

Measuring competitive balance in sports

https://doi.org/10.1515/jqas-2020-0006
Received January 21, 2020; accepted September 6, 2020;
published online September 29, 2020

Abstract: In order to make comparisons of competitive balance across sports leagues, we need to take into account how different season lengths influence observed measures of balance. We develop the first measures of competitive balance that are invariant to season length. The most commonly used measure, the ASD/ISD or Noll-Scully ratio, is biased. It artificially inflates the imbalance for leagues with long seasons (e.g., MLB) compared to those with short seasons (e.g., NFL). We provide a general model of competition that leads to unbiased variance estimates. The result is a new ordering across leagues: the NFL goes from having the most balance to being tied for the least, while MLB becomes the sport with the most balance. Our model also provides insight into competitive balance at the game level. We shift attention from team-level to game-level measures as these are more directly related to the predictability of a representative contest. Finally, we measure competitive balance at the season level. We do so by looking at the predictability of the final rankings as seen from the start of the season. Here the NBA stands out for having the most predictable results and hence the lowest full-season competitive balance.

Keywords: competitive balance; HHI; Noll-Scully index; sports analytics; uncertainty of outcome.

1 Introduction

The issue of competitive balance illustrates a fundamental difference between sports and business competition. No one buys tickets to watch Apple compete with Samsung. If Apple makes better phones and captures all the market, that's competition at its best. In contrast, if one sports team handily beats all its rivals then there is less excitement and suspense in watching the game (Ely, Frankel, and Kamenica 2015). The fans of the dominant team might be gratified with all the wins, but fans of the other teams may tune out (Rottenberg (1956), Neale (1964), Knowles, Sherony, and Haupert (1992)).

To promote competition and maintain excitement, most sports leagues take actions that limit the extent of inequality among teams. Examples of these actions include: giving better draft choices to teams that finish at the bottom of the standings; restricting the amount teams can spend to acquire talent; and capping the number of players on a team's roster.

To the extent that competitive balance is a goal, it needs to be measured. Existing measures of competitive balance reflect two distinct dimensions of inequality, one static and the other dynamic. Static measures look at balance within a given season. For example, how much better are the records of the best teams than the worst teams? Dynamic or dominance measures look at changes in team strengths across multiple seasons; see, for example, Eckard (1998), Humphreys (2002), and Ramchandani et al. (2018). Do the teams that do well in one season continue to dominate in the next season and beyond? A league could have a low static level of competitive balance, but a high level of dynamic balance if different teams were to dominate the league each year.

The focus of this paper is on static measures of competitive balance with application to the four major North American sports leagues. We are especially interested in being able to make comparisons across sports leagues with different season lengths. The four North American league season lengths vary by a factor of 10, from 16 games per team in the NFL to 162 in MLB. European football leagues season lengths are much more homogeneous—the season length is 34 in the Bundesliga, and 38 in the EPL, Serie A and La Liga. As a result, providing a proper adjustment for season length is more important in the four North American leagues, enough so that it can change the rankings.

We offer three innovations to static measures of competitive balance. Most importantly, we develop the first

We thank **Jason Abaluck**, **Florian Ederer**, **Ed Kaplan**, **Roger Noll**, **Nils Rudi**, and two anonymous referees for their perceptive feedback. Barry Nalebuff has previously done consulting work for the NBA. The views in this article entirely reflect those of the authors individually.

***Corresponding author: Prof. Barry Nalebuff**, Yale School of Management, New Haven, Connecticut, USA,
e-mail: barry.nalebuff@yale.edu. https://orcid.org/0000-0002-6290-7009
Matthew Doria, Salary Cap Management at NBA, New York, NY, USA,
e-mail: mdoria10@gmail.com

measures of competitive balance that are invariant to season length. In the process, we show why the commonly used Noll-Scully ("N-S") metric is biased and how to correct it. To do so, we build a general model of competition. This model leads to tight bounds on unbiased estimates of team-level variance, and, correspondingly, a new perspective on team-level competitive balance across the four major North American sports leagues. Based on the N-S metric, the NFL was thought to have the greatest level of balance, while the NBA was thought to have the lowest level. Using the unbiased measure, the NFL falls from the top of the list to tie with the NBA at the bottom. Similarly, while college basketball has greater balance than the NBA according to the N-S metric, after we correct for the relatively short season of college basketball, we find the NBA has greater competitive balance.

Team-level balance in European leagues tends to be measured using variations on the Herfindahl-Hirschman Index (HHI). We show how to adjust the HHI and related metrics for season length; here, the effect is small and the balance ranking across European leagues is unchanged.

Our second innovation is motivated by the observation that measures of competitive balance at the team-level (e.g., the N-S metric, HHI, and our variance estimate) do not provide direct insight into the predictability of a typical contest. This leads us to focus on and measure balance at the game level—to what extent is a typical game evenly matched and thus less predictable? As suggested by the title of Pat Toomay's novel (and the movie it inspired), if a league is well balanced at the game level then *On Any Given Sunday*, any team can win. As we show in Section 7, individual games are more evenly matched in MLB and the NHL than in the NBA and NFL. At the extreme, when the top team faces the bottom team, the bottom-ranked team has over a 30% chance of winning in MLB, but only a 3% chance in the NFL.

Our third innovation provides a measure of competitive balance based on each team's chance to end the regular season in first place, which we refer to as season-level competitive balance. For each league, we look at how many teams viably compete to end up in first place. One could think of this as an uncertainty-of-outcome measure taken at the start of the season. To calculate this measure, we start with each team's chance of winning a single game and then via simulation compound those probabilities over the length of the full season to determine each team's final rank. Unlike the prior measures, which seek to normalize for season length, season length is now an important factor. While our game-level metric reveals a similar level of balance between the NFL and the NBA for a typical game, because the NBA season is over five times as long, the final NBA rankings are much more predictable than those of the NFL (or the NHL and MLB). The chance that the best team (i.e., the team with the highest game-winning probability) will end the season with the best record is 42% in the NBA, 25% in MLB, 23% in the NHL and 22% in the NFL.

We begin in Section 2 with a history of and motivation for the N-S metric. Section 3 models a simplified two-team binomial league to illustrate the mistaken adjustment for season length employed using the N-S metric. Section 4 introduces a general model of competition and uses the model to develop tight bounds on an unbiased estimate of team-level variance. With these bounds, Section 5 presents an unbiased comparison of the underlying degree of team-level imbalance across the four North American leagues. Section 6 examines the use of variance as a measure of balance and argues for measuring variance at the game level, not at the team level. We do so in Section 7 where we employ rank-order model of competition. In this model, a team's chance of winning depends on its rank relative to that of its opponent. This model allows us to develop a closed-form solution for both team-level and game-level measures of competitive balance. Section 8 considers the predictability of end-of-season rankings and provides a metric for season-level balance. Section 9 offers a short conclusion. An Appendix provides details for several of the calculations in the paper.

2 Background

Rottenberg (1956) predicted that sports attendance would decline with greater dispersion in winning percentages. To measure dispersion, Roger Noll (1988, 1991) proposed looking at the difference between the win-loss percentage of the best team and the worst team in a season. There was no need to make any correction for season length as his analysis was focused on one league, the NBA, and how the balance changed over time. (Noll also looked at dynamic balance in terms of how many teams won championships and compared this to the theoretical ideal when all teams have equal ability.)

Following discussions with Noll, Gerald Scully (1989) proposed using the variance in win-loss records as measure of balance. If all teams had 0.500 records, there would be zero variance in the win-loss records and perfect balance. One could think of the zero-variance case as the ideal, but even in the case where all teams are equal we would not expect all teams to have 0.500 records. This is for the same reason that when one tosses a fair coin 16 times, the result will not always be eight heads and eight tails. Thus Quirk and Fort (1992) proposed a different ideal, the variance due to luck or sample size when each matchup is a 50:50 coin toss. This idealized variance can be used as the baseline to evaluate the observed variance.

A baseline is needed when making comparisons across leagues because, all else equal, a league with more games in its season will tend to have less sample size noise and thus less variance in team win percentages than a league with fewer games. Hence, in order to draw meaningful comparisons between leagues whose season lengths vary significantly, the variance in win percentages needs to be normalized to account for season length.

The idea of a variance ratio (or, more specifically, a standard deviation ratio) was formalized by Quirk and Fort (1992) in a measure that has come to be known as the Noll-Scully metric. The N-S metric for a league in a season is ASD/ISD, where ASD is the actual standard deviation in win percentages records and ISD is the ideal standard deviation in win percentages in the case where each game is a fair coin toss. While a ratio of 1 is the "ideal," it is theoretically possible that the N-S score is below 1, although this has not been observed in practice; see Goossens (2006).

A purported strength of the N-S measure is that one can use it to compare the competitive balance across different leagues; see, for example, Berri, Schmidt, and Brook (2007). The N-S metric is 1 for any season length in a perfectly balanced league. However, *in all other cases*, the ratio is not neutral to season length. In the next section, we show that the N-S metric is significantly biased against leagues with long seasons. The longer the season, the less competitively balanced the league will appear based on the N-S metric. Thus, the NFL with its short season will appear to be more competitively balanced than MLB with its long season.

The bias of the N-S measure has been well-established by Owen and King (2015) through extensive simulation. We develop a model of competition which generates an unbiased estimate of variance (as well as HHI) both at the team level and at the game level.

3 The mathematical problem with the N-S metric

To illustrate the problem with the N-S metric, we start with an overly simplified league, one with only two teams that play each other n times. Each time they meet, the better team has a chance $p \geq 0.5$ of winning. When we say the better team has a higher p, we are using p to denote a team's ability to win a game. In the literature, a team's p is sometimes referred to as its "ability." This shorthand use of ability can be misleading as a team's chance of winning depends on a constellation of factors that go beyond differences in raw player talent. A team's p also depends on how those differences play out based on the rules of the game. For example, a baseball team could have a star hitter, but the effect of that advantage is mitigated by the rule that requires all nine players to take their turn at bat; in contrast, there is no rule limiting how often a star player can take the shot in basketball. Similarly, the length of the sporting contest or number of scoring opportunities will influence how differences in talent translate into differences in winning probabilities (with shorter contests tending to dampen the effect of talent differences among teams). Teasing out the extent to which the various factors influence winning probabilities is outside the scope of this paper. Our measures of competitive balance depend on the underlying fundamentals, as captured by the distribution of winning probabilities across teams in the league.

By defining p for one team (and $1-p$ for the other), we establish the degree of imbalance in the league and it remains constant. This allows us to evaluate whether and how the expected N-S metric for the league changes with different season lengths.

What we observe are the win percentages which are not the same as the underlying probabilities, and that is the source of the problem. For example, if the season is short, say just one game in the extreme, we will observe one team with a 1.000 win percentage and the other with a 0.000 win percentage. Even if the teams are equally matched, the league will (artificially) look very imbalanced. The variance in win percentages will be 0.25 as each team is 0.5 away from a 0.500 win percentage.

As the season becomes infinitely long, the better team's winning percentage will converge to p while the worse team's winning percentage will converge to $1-p$, and the variance in win percentages will converge to $\frac{1}{2}(p - 0.500)^2 + \frac{1}{2}(1 - p - 0.500)^2 = (p - 0.500)^2$. For example, if p is 0.6, then the two teams' winning percentages will converge to 0.600 and 0.400, and the variance will converge to 0.01. The variance falls from 0.25 to 0.01 as the season goes from one game to infinitely long.

As this example illustrates, even when we hold the true imbalance fixed (at $p = 0.6$), the expectation of the observed variance falls as the season gets longer. To understand the size of this effect, we calculate the expected variance in the observed winning percentage for a season length of n games. Once we know the predicted effect of season length, we can correct for it and thereby achieve an unbiased estimate of the true level of dispersion for any season length.

If we knew the true p, the variance would be $\sigma^2 = (p-0.5)^2$ and this is unchanged with season length. The problem is that we don't know the p. Instead, we observe winning percentages w_1 and w_2 which leads to an observed variance $\hat{\sigma}^2$. As each game must have a winner and a loser, $w_1 = 1-w_2$, and the average team must win 50% of its games.[1] Thus

$$\hat{\sigma}^2 = 0.5*[w_1 - 0.5]^2 + 0.5*[w_2 - 0.5]^2 = [w_1 - 0.5]^2. \quad (1)$$

For notational simplicity, we refer to w_1 as w.

The resulting estimation problem is similar to how one estimates true variance from observed variance. The expected value of the observed variance in win percentages is

$$E[\hat{\sigma}^2] = E[w - 0.5]^2 \quad (2)$$

$$= E[(w - p) + (p - 0.5)]^2 \quad (3)$$

$$= E(w - p)^2 + E(p - 0.5)^2 \quad (4)$$

$$= \frac{p(1-p)}{n} + E(p - 0.5)^2 \quad (5)$$

$$= \sigma^2_{sample\ size} + \sigma^2. \quad (6)$$

The key step in going from equation (3) to (4) is that $E(w-p) = 0$, so that $E[(w-p)(p-0.5)] = 0$. In moving from (4) to (5), the term $E(w-p)^2$ is the sample size variance in the frequency of observed heads from a weighted coin. From the binomial distribution, we know $\sigma^2_{sample\ size} = p(1-p)/n$.

We have distilled the expected value of observed variance into the sum of two components: the first is the variance attributable to finite sample size; the second is the component of variance attributable to the different true winning probabilities across the teams. This is similar to the standard formula used in measuring mean squared error, except in our case the bias term is not bias but the term of interest.[2]

As the formula makes clear, the expected observed variance decreases with the length of the season. The reason is that the longer the season, the lower the variation due to sample size. As the season converges to an infinite length, the observed win percentages converge to p and $1-p$, and the variance converges to σ^2. Since the expected observed variance changes with n, we can't directly compare the observed variances across leagues with different season lengths.

To solve this problem, Quirk and Fort (1992) use as a normalization the expected variance in the idealized case where all teams have equal win rates ($p = 0.5$). In that case, the predicted observed variance in a season of length n would be $\sigma^2 = 0.25/n$. The N-S metric (or more precisely the square of the N-S metric) is the ratio of the observed variance to the expected idealized variance, $\hat{\sigma}^2/(0.25/n)$. We use equation (5) to calculate the expected value of (N-S)^2:

$$E\left[(\text{N} - \text{S})^2\right] = \frac{\frac{p(1-p)}{n} + E(p - 0.5)^2}{\frac{0.25}{n}} \quad (7)$$

$$= 4p(1-p) + 4n\sigma^2 \quad (8)$$

The first term is independent of season length, but the second is not. Only in the special case where the teams are equal ($p = 0.5$ and $\sigma^2 = 0$) is this measure of balance independent of season length, and the expected value of (N-S)^2 is 1 for all n.

In the realistic case where $\sigma^2 > 0$, so that the teams are not equal, it follows that the expected value of the N-S metric rises with n, a point recognized by Owens and King (2015) and Birnbaum (2016). The greater the value of σ^2, the more the expected value of N-S increases with n.

Consider the expected value of the N-S metric for the case where $p = 0.6$ as n goes from 1 to 16 to 82 to 162. Not only does the (N-S)^2 metric and the N-S metric approach infinity as n gets large, but the bias is significant even for small n. For example, holding p at 0.6, as the season length increases from 16 games (e.g., the NFL) to 162 games (e.g., the MLB), the expected N-S metric more than doubles, increasing from 1.03 to 2.55; see Table 1. And yet the underlying difference between the two teams remains the same. This suggests that the N-S metric significantly overstates competitive balance in leagues that play fewer games relative to leagues that have longer seasons.

Table 1: N-S values for various n.

n	Expected (N-S)^2 metric	Expected N-S metric
1	0.96 + 0.04 = 1.00	1.00
16	0.96 + 16*0.04 = 1.60	1.03
82	0.96 + 82*0.04 = 4.24	1.84
162	0.96 + 162*0.04 = 7.44	2.55
large n	= 0.04n	$0.2\sqrt{n}$

1 For the NHL, where ties and overtime games are possible, we treat ties as equivalent to a half win and overtime/shoot-out wins and losses as equivalent to regulation-game wins and losses. Similarly, for the NFL, we treat ties as a half win. We do this to make an apple-to-apples comparison across leagues.

2 If we estimate each team's observed win rate by the constant 0.5, the mean squared error associated with that estimate is the variance $\sigma^2_{sample\ size}$ plus what is called the bias, which is our variable of interest, σ^2.

If we return to equations (7) and (8), we can identify the source of the problem. By implicitly assuming that the ideal case is equal win probability or $p = 0.5$, the denominator in the N-S score only considers variance due to the randomness in game outcomes, i.e., the first term in the numerator of (7). It ignores variance due to any underlying imbalance in the system, i.e., σ^2, the second term in (7). But no actual league is ideally balanced and the imbalance in the system is what we are looking to estimate and compare across leagues. The main source of variability in winning percentages as n gets large is the different win probabilities between the teams. Once we recognize this, we can correct for the measurement error and thereby calculate an unbiased estimate of the true variance.

4 The solution

We are looking for an unbiased estimate of the variance in teams' win probabilities, one that is invariant to season length. The challenge is we don't observe p or, with many teams, the p_i for each team. Thus, we want to find a way to estimate σ^2 from the observed variance, $\hat{\sigma}^2$. To do so, we build a general binomial model of competition.

In a season where each team plays n games, let team i's chance of winning game k be p_{ik}. Note that p_{ik} is not team i's chance against opponent k. This is team i's chance in game k of the season; it may depend on the point in the season, the opponent, and whether the game is home or away. This approach is quite general. Two games against the same rival could have different winning probabilities at different points in the season. The winning probabilities can depend on what has happened earlier in the season. There is no assumption that the season is balanced. The only constraint is that for the team that plays team i in game k, its chance of winning must be $1-p_{ik}$.

Team i's average chance of winning, p_i, and its observed winning record, w_i, are

$$p_i = \frac{1}{n}\sum_{k=1}^{n} p_{ik} \tag{9}$$

$$w_i = \frac{1}{n}\sum_{k=1}^{n} w_{ik} \tag{10}$$

Since every game has a winner and a loser, the average win probability is always 0.5 and the observed variance is

$$\hat{\sigma}^2 = \frac{1}{T}\sum_{i=1}^{T} [w_i - 0.5]^2 \tag{11}$$

Similar to the results in equations (2) through (6), we show in Appendix calculation 1 that in a league with T teams:

$$E[\hat{\sigma}^2] = \frac{1}{T}\sum_{i=1}^{T} E[w_i - 0.5]^2 \tag{12}$$

$$= \sigma^2 + \frac{1}{n^2 T}\sum_{i=1}^{T}\sum_{k=1}^{n} p_{ik}(1 - p_{ik}) \tag{13}$$

where the true variance is

$$\sigma^2 = \frac{1}{T}\sum_{i=1}^{T} (p_i - 0.5)^2 \tag{14}$$

Note that this true variance measure is an average across the entire league and not for each division.

Calculation 1 in the Appendix goes on to show that

$$E[\hat{\sigma}^2] = \sigma^2\frac{n-1}{n} + \frac{0.25}{n} - \frac{1}{T}\sum_{i=1}^{T}\sum_{k=1}^{n}\frac{(p_{ik} - p_i)^2}{n^2} \tag{15}$$

Once $n > 1$, we can rearrange this formula and substitute the observed variance for its expectation. Since the expected value of the observed variance is by definition $E[\hat{\sigma}^2]$, this leads to an unbiased estimate of the true variance, represented by s^2:

$$s^2 = \frac{n}{n-1}\left[\hat{\sigma}^2 - \frac{0.25}{n} + \frac{1}{n^2\, T}\sum_{i=1}^{T}\sum_{k=1}^{n} (p_{ik} - p_i)^2\right] \tag{16}$$

This estimate is comprised of three adjustments. There is the standard Bessel correction for the number of observations. This is done after we subtract from the observed variance the expected variance in the idealized case where all teams are equal. In addition, we need to add a small correction factor for the fact that each team's winning probability p_{ik} is generally not constant across rivals and over the season.

Sports writers Tango (2006) and Mauboussin (2012) have suggested rearranging equation (6) to develop an estimate of the true variance: $s^2 = \hat{\sigma}^2 - \frac{0.25}{n}$. Subtracting the ideal variance is better than dividing by it. But as we can see from equation (16), this is not an unbiased formula. The issue is that the idealized variance is maximal and thus always larger than the true sample-size variance. While the difference is small for MLB with $n = 162$, it matters for the NFL where $n = 16$.

To explore our new estimate, we briefly return to the simple case with two teams. In that model, if the chance team 1 beats team 2 is constant over the season then $p_{1k} = p$ and our estimate in equation (16) becomes

$$s^2 = \frac{n}{n-1}\left(\widehat{\sigma}^2 - \frac{0.25}{n}\right). \tag{17}$$

Next we add a home-team advantage, a relevant factor that has been documented in most sports; see Trandel and Maxcy (2011) and Lopez, Matthews, and Baumer (2018). Continuing with the 2-team example, if we assume the chance team 1 beats team 2 is constant at p_h at home and p_a when away, then let $p = (p_h + p_a)/2$ and $h = (p_h – p_a)/2$, where h is the home-field advantage. If the schedule is balanced between home and away, $p_{1k} = p + h$ half the time, $p–h$ the other half, and the average value of p_1 is unchanged. Our estimate in equation (16) becomes

$$s^2 = \frac{n}{n-1}\left(\widehat{\sigma}^2 - \frac{0.25 - h^2}{n}\right). \tag{18}$$

This adjustment is smaller than prior adjustments to the N-S score as calculated by Trandel and Maxcy (2011). The size of their prior adjustment depends on the home advantage h^2/n compared to $0.25/n$, and thus the season length n factors out of the comparison. Here, h^2/n is being compared to $\widehat{\sigma}^2 - 0.25/n$, and so the adjustment depends on n and is much smaller when n is large.

We can apply this approach more generally to multi-team leagues. If we think of p_{ik} as team i's baseline chance of winning game k when played on neutral ground and this goes up or down by a constant h where there is a 50% chance any game is home or away, then team i's adjusted chance of winning game k is $p^*_{ik} = p_{ik} \pm h$.[3] In this case, $E(p^*_{ik} - p_i)^2 = (p_{ik} - p_i)^2 + h^2$, and our estimate becomes

$$s^2 = \frac{n}{n-1}\left[\widehat{\sigma}^2 - \frac{0.25 - h^2}{n} + \frac{1}{n^2\ T}\sum_{i=1}^{T}\sum_{k=1}^{n}(p_{ik} - p_i)^2\right] \tag{19}$$

To get a sense for the magnitude of the final term, consider the case where $p_{ik}–p_i$ ranges uniformly over $[–m, m]$. If $m = 0.2$, a top team with an average win record of 70% will have p_{ik} vary between 90% and 50% on neutral ground. The 90% win rate is against the lowest-ranked team and the 50% rate is when playing its closest rival. For a mid-ranked team with $p_i = 0.5$, it would have a 70% chance of beating the lowest-ranked team, but only a 30% chance of beating the top team (when playing on neutral ground). In this case, the value of the final term can be approximated by $m^2/3$

$$s^2 \approx \frac{n}{n-1}\left[\widehat{\sigma}^2 - \frac{0.25 - h^2 - m^2/3}{n}\right] \tag{20}$$

with $m = 0.20$ and $h = 0.10$, $s^2 \approx \frac{n}{n-1}\left[\widehat{\sigma}^2 - \frac{0.23}{n}\right]$.

In Section 7, we present a model that allows us to estimate our closed-form solution for s^2. In this model, each team's chance of winning (on neutral ground) is determined by its rank relative to the rank of its rival—the bigger the difference in ranking, the greater the chance the higher-ranked team wins. As we show, this model leads to something very close to the case of a uniform distribution of $p_{ik}–p_i$. We then use the model to estimate values of our analog of m for each of the leagues in each year.

At this point, even without solving the model, we can place realistic bounds on the estimate of s^2, bounds that allow us to make comparisons across leagues. We assume that $|p_{ik} - p_i| \le 0.2$. Data from the four major North American sports leagues shows that typically $h \le 0.10$. Putting this together leads to the bounds

$$\frac{n}{n-1}\left[\widehat{\sigma}^2 - \frac{0.25}{n}\right] \le s^2 \le \frac{n}{n-1}\left[\widehat{\sigma}^2 - \frac{0.20}{n}\right] \tag{21}$$

For purposes of comparison, the N-S measure of balance is based on the standard deviation rather than the variance. We provide a comparable score by taking the square root of our unbiased estimate. In a 30-team league, one can think of the estimate of s as being an indication of how much the 5$^{\text{th}}$-best team is above 0.500. In the next section, we present a graph of s across time and across the leagues. In this graph, the thickness of each line represents the range of possible values in equation (21).

We recognize that the expected value of the standard deviation is not the square root of the expected value of the variance (as the square root is not a linear function). While this is a standard issue with the commonly used estimates of the standard deviation, simulation shows the expected bias is below 1%.[4]

There is one remaining potential issue with using s^2 (or its square root) as a way to compare balance across sports leagues. We have corrected for the length of the

3 As a referee pointed out, we need to be careful that $p_{ik} \pm h$ is not above 1 or below 0. Given the realistic range of values for p_{ik} and the size of h, this is nearly always true. Even when the best team plays the worst team on a neutral ground, the chance of winning is rarely above 90% and the home team advantage is generally below 10%. In that case, adding a home-team advantage doesn't change the average value of p_i. We return to this issue when estimating our rank-order model in Section 7.

4 We ran simulations for virtual leagues where the distribution of winning probabilities, and hence σ^2, was known. In these simulations, the schedules were perfectly balanced and so season length did not line up perfectly with actual season lengths. The simulated chance team i beat team j was $\frac{1}{2} + (i - j)\Delta$, where i and j were the ranks of the two teams (as discussed in the rank-order model in Section 7), and Δ was estimated for each league. With one million simulations, the estimates s^2 were within 0.02% of the actual value. The s estimates for σ were slightly biased downward (as expected) but were still within 0.8% of the true value for the NFL and within 0.3% for the other three leagues (due to their longer seasons).

season, but we have not adjusted for the number of teams in a league.

The issue is the maximum possible value of σ^2 is not 0.25—as would be the case if it were possible for half the teams to have perfect records and the other half to win no games. In sports leagues, the bottom teams end up playing each other and so one team must win. The most unequal level of balance possible arises when teams can be ranked from 1 to T and the higher-ranked team always beats the lower-ranked team; see Horowitz (1997) and Fort and Quirk (1997). The resulting distribution of team win probabilities will be $\left[0, \frac{1}{T-1}, \frac{2}{T-1}, \ \ldots 1\right]$ which leads to the maximum possible value of σ^2:

$$\text{Max} \quad \sigma^2 = \frac{1}{T}\sum_{i=1}^{T}\frac{(i-1)^2}{(T-1)^2} - \frac{1}{4} = \frac{T+1}{12(T-1)} \quad . \tag{22}$$

The maximum possible variance is not independent of the number of teams. (As the number of teams becomes large, the distribution of win probabilities becomes uniform over [0, 1] and the variance converges to 1/12.) Since the number of teams differs across the leagues (and over time), the question arises whether we should compare the variances directly or the variances relative to their possible maximum; see Owen, Ryan, and Weatherston (2007) for a discussion of this issue in the context of the Herfindahl index and Owen (2010) in the context of Noll-Scully.

In practice, this complication may not matter. The estimated variances are much smaller than their possible maximum. The typical values are 25% of the maximum in the NBA and NFL, 7% in the NHL, and 4% in MLB. This suggests that the maximum is far from binding. The maximum possible values of σ^2 are also tightly clustered across the leagues. From low to high, they differ by less than 0.5%. When looking at the maximum possible standard deviations, the range is only 0.2%. Given the much wider differences across leagues—typically over 100% between the NHL and MLB compared to the NBA and NFL—a 0.2% potential correction due to league size will not change any of our conclusions.

We make one final detour before turning to empirical results. We have so far emphasized variance as a measure of competitive balance. In European football, it is more common to consider variations on the Herfindahl-Hirschman Index (HHI) as a measure of balance; see Depken (1999), Michie and Oughton (2004), and Plumley and Flint (2015).

HHI is closely connected to variance—knowing one metric directly leads to the other. The HHI is defined as the sum of squared market shares across firms in an industry. To translate this to sports, we look at "market share" of wins. If team i's win percentage is w_i, its share of wins is $nw_i/(nT/2) = 2w_i/T$. Thus

$$\text{HHI} = \left(\frac{2}{T}\right)^2 \sum_{i=1}^{T} w_i^2. \tag{23}$$

To compare variance with HHI,

$$\text{HHI} = \left(\frac{2}{T}\right)^2 \sum_{i=1}^{T}\left[\left(w_i^2 - \frac{1}{4}\right) + \frac{1}{4}\right] = \frac{4}{T}\left[\widehat{\sigma}^2 + \frac{1}{4}\right] \tag{24}$$

The two measures are linearly related, although the index varies inversely with the number of teams. Michie and Oughton (2004) control for the number of teams by looking at what they call the Herfindahl Index of Competitive Balance (HICB), where $\text{HICB} = 100 * \text{HHI}/(1/T) = 100 * [4\widehat{\sigma}^2 + 1]$. Effectively, HICB brings the HHI back to a variance measure. The true HHI would be based on the actual, not the observed, variance. To provide an unbiased estimate of the true HHI or HICB from the observed data, we use the same tools to estimate s^2 based on $\widehat{\sigma}^2$ and then apply this to the formula in (24).

This relationship is not exact when one adds the potential for ties to the mix and a scoring system where wins are awarded three points and ties lead to one point for each team. In Section 7, we show how to adjust HHI-related measures for season length. The impact is relatively small as the season lengths are much more homogeneous across different European football leagues than across different North American sports leagues.

5 Empirical observations

We have established that the N-S score fails to make the appropriate correction for season length. Here we look at our s estimates across leagues. We compare these results to the earlier conclusions based on the N-S metric. Figure 1 shows the year-by-year breakdown of Noll-Scully across the NBA, NFL, MLB, and NHL since 1991.

From this chart, it appears that the NBA is an outlier in terms of static competitive balance and that the NFL is the league with the most balance. Berri, Schmidt, and Brook (2007) in *The Wages of Wins* use the N-S score to suggest that the NBA has significantly less competitive balance than the three other leagues.

But, as we now know, the N-S scores are biased. In Figure 2, we show our unbiased estimates of the standard deviation in winning probabilities. The thickness of the lines represents the range of possible values based on the bounds in equation (21).

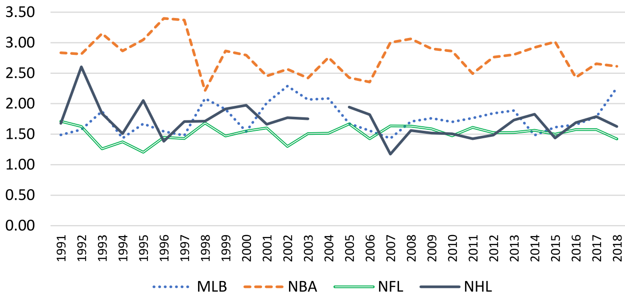

Figure 1: N-S metric in North American sports leagues.

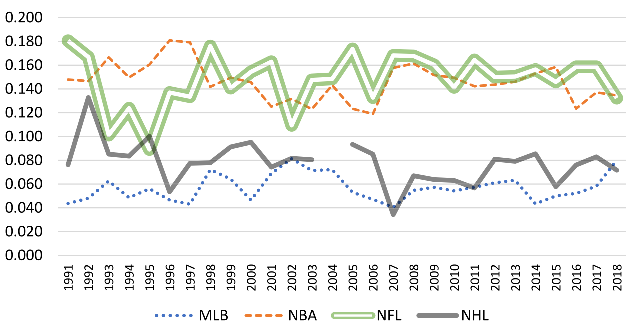

Figure 2: s estimates in North American sports leagues.

This chart shows that the NFL and the NBA are about even in terms of static competitive balance. The league with the most balance is MLB, followed by the NHL. Since the N-S score penalizes leagues with longer seasons, the new score shows that the MLB has greater competitive balance than previously thought, moving from third to first place. Meanwhile, the NFL, which was previously winning the N-S balance race, was doing so only due to its short season.

From 2010 through 2018, the estimates of s for the NBA and the NFL have been near 0.15. This means that the typical 5th-best team and typical 5th-lowest team are about 0.150 points away from 0.500. Thus, the typical 5th-best team wins 65% of its games and the typical 5th-lowest team wins 35% of its games. In contrast, for MLB and the NHL, the 2010 through 2018 estimates of s are roughly 0.06 and 0.07, respectively. For these two leagues, the typical 5th-best team wins 56–57% of its games and the typical 5th-lowest team wins 44–43% of its games.

This new ranking is in line with the results of Tango (2006). Tango looked at the observed variance minus the estimated randomness assuming all teams are equal to calculate an estimate of the true variance in winning probabilities. While this is not an unbiased estimate, the small level of bias is not large enough to mask the differences in the four leagues.

When a league has a higher σ^2, the win probabilities are farther apart and thus it should be easier to determine which teams have the highest and lowest "abilities." This idea is discussed in Mauboussin (2012, pp. 78–80) who uses as a metric the ratio of variance from luck (the ideal variance when all teams are equal) over the total observed variance—his ratio is precisely the inverse of the N-S metric squared and thus suffers from all the same issues. Mauboussin compares the 12% contribution of luck in the NBA to the 34% contribution of luck in MLB and concludes: "10 basketball games, for instance, reveal much more about ability than 10 baseball games do." Since he calculates an even larger contribution of luck in the NFL at 38% compared to MLB, he would presumably make the conclusion that 10 basketball games reveal much more about differential "ability" or what we call differential winning probabilities than 10 football games. But this latter conclusion would be wrong.

The variance due to different winning probabilities, when properly measured, is nearly identical in the NBA and NFL. The only reason that luck plays a bigger role in the NFL is that its season is 16 games long compared to 82 in the NBA. Because the NFL season is much shorter than the NBA season, luck plays a bigger role in the NFL in terms of *final* standings. One learns no more after 10 NFL games than 10 NBA games about which teams are better. It is true that one learns more after 10 NBA games than 10 MLB games, but this is because σ^2 is bigger in the NBA than in the MLB. The best teams win a much higher percent of their games in the NBA compared to the MLB, and thus it takes fewer games to determine which NBA teams are best.

6 A better variance measure

Having corrected the way to calculate a variance measure with respect to season length, we now address the more fundamental question of whether an unbiased measure of variance at the team-level is a good metric for competitive balance. The metric does not provide information about the predictability of end-of-season standings. It also doesn't directly provide information on the predictability of an individual representative game. The problem is not with the use of variance, but that both the N-S metric and our unbiased estimate s^2 measure variance across teams' season-average win probabilities rather than the variance that arises at the game level.

To highlight the issue, consider again the example of a league with the most unequal level of balance. In this league, the higher-ranked team always defeats a lower-ranked rival. With a balanced schedule, the distribution of win records will be close to uniform and $\sigma^2 \approx 1/12 = 0.083$.

The variance in team win probabilities may not fully reveal the level of imbalance in a representative game.

This leads us to consider variance in the winning probabilities, p_{ik}, at the game level:

$$\sigma_{game}^2 = \frac{1}{Tn}\sum_{i=1}^{T}\sum_{k=1}^{n}(p_{ik}-1/2)^2 \tag{25}$$

In this maximally unequal league, p_{ik} is either 0 or 1, depending on whether i is the lower- or higher-ranked team. Across all teams half the values are 0 and the other half are 1. Thus $\sigma_{game}^2 = 0.5^2 = 0.25$.

In this example, the p_{ik} reveal that all games are entirely predictable, a fact potentially hidden when looking just at the p_i. It isn't just that the variance is bigger at the game level, but that p_{ik} are directly connected to predictability at the game level while the p_i are not. To the extent we are using competitive balance as a proxy for unpredictability in a representative game, the variance in p_{ik} is more relevant than the variance in p_i.[5]

For this reason, we look at p_{ik}, not p_i, as the fundamental driver of competitive balance. While it might seem that this would lead to a much more complicated estimation problem, under our proposed model of competition in Section 7, σ_{game}^2 is almost exactly twice σ^2. Thus we can use the unbiased estimate s^2 to also estimate s_{game}^2. The ordering of competitive balance across leagues remains the same. Using this relationship also means that we can estimate s_{game}^2 using just 30 win records (in the NBA or MLB), rather than the 435 possible pairwise team records.

There remains the question of using variance as the statistic that best captures the competitive balance implied by the p_{ik}s. As we've seen, variance is closely connected to the Herfindahl-Hirschman Index of concentration. Indeed, we can think of variance as being an approximation to any natural competitive balance metric based on the p_{ik}. Consider, for example, the Shannon measure of entropy associated with the average game:

$$E = -\frac{2}{Tn}\sum_{i=1}^{T}\sum_{k=1}^{n}p_{ij}\,\mathrm{Log}_2(p_{ik}). \tag{26}$$

Using a Taylor expansion around $p_{ik} = 1/2$, Calculation #2 in the Appendix shows that

$$E \approx 1 - \frac{2}{Ln(2)}\;\sigma_{game}^2 \tag{27}$$

More generally, in any Taylor approximation around $p_{ik} = \frac{1}{2}$, the linear term will be zero, as $E[p_{ik}] = \frac{1}{2}$ across all teams in the league. Thus, to the extent that the first three terms of the Taylor expansion are a good approximation, variance at the game level will give us an indication of any measure of competitive balance based on the p_{ik}.[6]

In the next section, we consider two game-level measures of competitive balance that are functions of variance in our model. One measure looks at the chance the worst team has against the best. A second metric, P_{HRW}, looks at a random matchup and calculates the chance that the higher-ranked team wins. This metric captures the level of unpredictability in an average game. In a perfectly balanced league, P_{HRW} would be ½ and in a maximally unbalanced league P_{HRW} would be 1. Whether using variance or some other score, we are led to derive competitive balance metrics using dispersion in p_{ik} rather than p_i.

7 A rank-order model of power

We employ a model where the chance team i beats j depends on the power of i relative to j. We use this model to provide unbiased estimates for competitive balance across leagues. Our first calculation is an exact estimate of team-level variance in equation (16). Our second calculation provides a measure of competitive balance estimate based on the variance of each individual game result rather than each team's average winning probability.

We find that variance at the game level is almost exactly twice the variance at the team level. Since the new variance measure is almost exactly double the prior one, the relative rankings of the different leagues are unchanged. Indeed, there is little need to present new charts of competitive balance across leagues as the graphs are nearly identical except for the scale of the vertical axis (and the thinness of the lines).

The model also allows us to present competitive balance measures that are connected to uncertainty of outcome, both at the game level and at the season level.

5 Measuring variance at the game level solves the problem that the maximal possible variance depends on the number of teams in the league. When variance is measured at the game level, the maximum pairwise-variance is 0.25 no matter how many teams are in the league. This independence across league size eliminates any league-size source of bias in a cross-league comparisons of balance or a time series comparison when the number of teams changes over time.

6 This assumes the derivatives are bounded. If, for example, the measure of competitive balance is the standard deviation, then $\sigma = 0$ at the approximation $p_{ik} = \frac{1}{2}$ and the first derivative $(1/(2\sigma))$ is not defined.

7.1 Building a model

A realistic model of competition has team i's chance of beating team j depend on the relative game-winning power of the two teams. Since game-winning power no longer directly corresponds with the winning probability, we need some new notation. We denote the power score of team i by μ_i. Similar to our earlier caution about not equating winning probability with ability, the translation of relative power to win probability depends not just on player talents, but also on other factors such as the rules of sport.

There are several ways of translating power scores into win probabilities. The Bradley-Terry Model (Bradley and Terry (1952)) or power-score ratio approach assumes the chance team i beats j is

$$q_{ij} = \frac{\mu_i}{\mu_i + \mu_j} \quad . \tag{28}$$

Note that we have changed notation. Here we are assuming that the chance team i wins only depends on the opponent j and not when the game is played. In our original notation, $q_{ij} = p_{ik}$ for all the games k in which team i plays j. Similar to our earlier specification, these are winning probabilities on neutral ground. Later we add a home-team advantage.

When two teams have equal power, the chance of winning is 0.5 for each. The bigger the difference in power, the more predictable is the outcome. One can also use power measures to predict the score differential (and by implication the better team); see Glickman and Stern (1998) and Boulier and Stekler (2003).

A variant on this approach is the "Log5" model applied to sports by Bill James (1981). If team i's game-winning power is μ_i and it faces a team j with power μ_j, the chance i wins is

$$q_{ij} = \frac{\mu_i(1-\mu_j)}{\mu_i(1-\mu_j) + \mu_j(1-\mu_i)} \quad . \tag{29}$$

Similar to the ratio model, a team with no power has no chance of winning (unless it plays another team with no power). The team with more power has a greater chance of winning, and two teams with equal power have an equal chance of winning. In the ratio model, the maximum power is infinity, while in the Log5 model the maximum power is 1. In fact, the Log5 model is identical to the ratio model if we rescale the game-winning powers to be $\frac{\mu_i}{1-\mu_i}$.

A challenge with both these approaches is that we have to estimate a game-winning power for each team and then use those estimates to derive the implied probabilities and competitive balance. Lopez, Matthews, and Baumer (2018) undertake this process using betting data, allowing for team strengths that vary across seasons and even week to week.

We present a simpler model where a team's winning probability is determined by its rank relative to that of its rival.[7] Here, we assume that when the team with rank i plays the team with rank j, team i's chance of winning (on neutral ground) is

$$q_{ij} = \frac{1}{2} + (i-j)\Delta\,. \tag{30}$$

The greater the difference in rank, the greater the chance the team with the higher rank will win, and the single parameter Δ determines the size of the effect.[8] An advantage of this model is that it can be estimated using season-level rather than game-level data.

To calculate each team's overall or average chance of winning, we assume the schedule for each team is perfectly balanced. For the ith-ranked team,

$$q_i = \frac{1}{T-1}\sum_{j\neq i}\left[\frac{1}{2} + (i-j)\Delta\right] = \frac{1}{2} + \frac{T}{T-1}\left(i - \frac{T+1}{2}\right)\Delta\,. \tag{31}$$

For the top-ranked team in a 30-team league, the average chance of winning is $q_{30} = \frac{1}{2} + 15\Delta$, while for the team just above the middle, $q_{16} = \frac{1}{2} + \frac{15}{29}\Delta$.

If we could observe the winning probabilities or Δ directly, we could calculate the true variance using team-average probabilities. Calculation #3 in the Appendix provides the details:

$$\sigma^2 = \frac{1}{T}\sum_{i=1}^{T}(q_i - 0.5)^2 = \Delta^2 T^2 \frac{T+1}{12(T-1)} \quad . \tag{32}$$

Similarly, if we knew all the q_{ij}, we could calculate the true variance at the game level. As shown in Calculation #4 of the Appendix,

$$\sigma^2_{game} = \frac{1}{T(T-1)}\sum_{i=1}^{T}\sum_{j\neq i}\left(q_{ij} - 0.5\right)^2 = \Delta^2\frac{T(T+1)}{6}. \tag{33}$$

Comparing equation (32) to (33), variance at the game level is almost exactly twice variance at the team level. The

7 Similar models have been developed in politics to predict the result of the electoral college. States are ranked according to their chance of favoring one party. A party wins all states up to some rank and the other party wins all higher-ranked states; see Chen, Ingersoll, and Kaplan (2008) and Fair (2009).

8 This is one possible formulation. An alternative is $q_{ij} = \frac{1}{2} + (i-j)^3\Delta$ so that difference in rank matters more as the rank difference gets larger. It is also possible that the difference between rank 30 (best) and 29 (next best) has a bigger effect than the difference between rank 16 and 15, as the teams in the middle are less differentiated in terms of winning probabilities. While more complicated models may produce better fits of win-loss records, we know the potential gain in estimating variance is small from the bounds in equation (21), and the linear model provides a simple closed-form solution for the variance in q_{ij}.

ratio is $2\frac{T-1}{T}$, which translates to 1.93 for MLB and the NBA, 1.94 for the NHL and NFL.

The practical question is how to develop an unbiased estimate of these two variances based on observed data. While it is certainly possible to estimate the relative win probability for each team in each potential matching, that would be a complicated undertaking. Fortunately, our model result says that we can calculate the game-level variance using only season-level data. We estimate s^2 from the observed variance in team records and the predicted sample variance in the rank-order model. We then employ the ratio $2\frac{T-1}{T}$ to translate s^2 into s^2_{game}.

Just as we observed in equation (13), the expected value of the observed variance is the true variance plus sample size noise:

$$E[\hat{\sigma}^2] = \frac{1}{T}\sum_{i=1}^{T}(w_i - 0.5)^2 = \sigma^2 + \frac{1}{n^2 T}\sum_{i=1}^{T}\frac{n}{T-1}\sum_{j\neq i} q_{ij}\left(1-q_{ij}\right) \tag{34}$$

The $\frac{n}{T-1}$ factor reflects the fact that in a perfectly balanced season of length n, team i plays j a total of $\frac{n}{T-1}$ times. We use the rank-order model to simplify the game-level variance:

$$q_{ij}\left(1-q_{ij}\right) = \left[\frac{1}{2} + (i-j)\Delta\right]\left[\frac{1}{2} - (i-j)\Delta\right] = 0.25 - (i-j)^2\Delta^2 \tag{35}$$

Substituting the value in (35) into (34) and summing, Appendix Calculation #5 shows

$$E[\hat{\sigma}^2] = \sigma^2 + \frac{0.25}{n} - \frac{\Delta^2 T(T+1)}{6n} \tag{36}$$

$$= \left[1 - \frac{2}{n}\frac{(T-1)}{T}\right]\sigma^2 + \frac{0.25}{n}, \tag{37}$$

where we use (32) for the value of σ^2 in going from (36) to (37).

There are two ways of representing this solution. In the context of equation (15), we can break the first term in (37) into $\frac{n-1}{n}$ plus a remainder and then substitute the value of σ^2 from (32) to yield:

$$E[\hat{\sigma}^2] = \sigma^2\frac{n-1}{n} + \frac{0.25}{n} - \frac{(T-2)}{nT}\frac{T^2(T+1)}{12(T-1)}\Delta^2 \tag{38}$$

$$\approx \sigma^2\frac{n-1}{n} + \frac{0.25}{n} - \frac{1}{3n}\left(\frac{(T-1)\Delta}{2}\right)^2. \tag{39}$$

Rearranging terms in (39) leads to

$$s^2 \approx \frac{n}{n-1}\left[\hat{\sigma}^2 - \frac{0.25 - m^2/3}{n}\right], \tag{40}$$

where $m = (T-1)\Delta/2$. This makes intuitive sense as q_{ij} differs from its mean q_i in equal steps going up by roughly $\frac{(T-1)\Delta}{2}$ when playing the lowest team and down by $-\frac{(T-1)\Delta}{2}$ when playing the top team.

We can also rearrange equation (37) to yield

$$s^2 = \left[\frac{n}{n-2+\frac{2}{T}}\right]\left[\hat{\sigma}^2 - \frac{0.25}{n}\right] \tag{41}$$

This formulation is similar to our lower bound estimate in equation (21); the only difference is that the denominator is $n-2+\frac{2}{T}$ rather than $n-1$.

With an 82-game season and a 30-team league in the NBA, the estimate of s^2 is just 1.2% larger (as a ratio) than the lower bound and the estimate of s is only 0.6% larger. The ratio is similarly close for the NHL and MLB. With a 16-game season in the NFL and 32 teams, the estimate s^2 is 6.7% larger (as a ratio) than the lower bound and the estimate of s is only 3.3% larger; see Table 2. The parameter estimates provided in Table 2 and all subsequent tables are averages of the relevant year-by-year parameter estimates.

Similarly, we can employ equation (41) to provide a season-length adjustment for HHI-based scores. Let HICB be the observed value of the Hirfindahl Index of Competitive Balance from Michie and Oughton (2004). An unbiased estimate for the true value of HICB, what we'll call UHICB, is based on an unbiased estimate for σ^2 rather than the observed $\hat{\sigma}^2$:

$$\text{UHICB} - 100 = 100 * 4s^2 \tag{42}$$

$$= \left[\frac{n}{n-2+\frac{2}{T}}\right]\left[\text{HICB} - 100 - \frac{100}{n}\right] \tag{43}$$

In European leagues, schedules are perfectly balanced and thus $n = 2(T-1)$ which implies $\frac{n}{n-2+\frac{2}{T}} = \frac{n+2}{n}$.

7.2 Game-level variance

While it is worthwhile to have an improved estimate for team-level variance, the main purpose of the model is to

Table 2: Est. standard deviation of team winning percentages – 1991–2018.

	Power-rank model s	Lower bound s	Ratio of estimates
MLB	5.71%	5.69%	1.003
NHL	7.80%	7.75%	1.006
NBA	14.65%	14.56%	1.006
NFL	15.08%	14.60%	1.033

develop an estimate of variance at the game level. The hard work has already been done. Since $\sigma^2_{game} = \frac{2(T-1)}{T}\sigma^2$, the estimate of game-level variance, denoted by s^2_{game}, follows directly from (41):

$$s^2_{game} = \frac{2(T-1)}{T}\ s^2 = 2\left[\frac{n(T-1)}{(n-2)T+2}\right]\left[\hat{\sigma}^2 - \frac{0.25}{n}\right]. \quad (44)$$

Since the game measure is almost exactly double the team-level measure, the relative rankings and the ratios across leagues are almost exactly the same. We don't have to measure the observed variance at the game level—our model allows us to estimate game-level variance using only team-level data.

To normalize the scores, we compare the results to the worst-possible case. The worst possible case at the game level is when each game is perfectly predictable: $q_{ij} = 1$ and $q_{ji} = 0$. In this case, both probabilities are 0.5 away from the average probability of 0.5. From Table 3A, we see that at the individual game level, MLB and the NHL are at 16% and 22% of the standard deviation of the worst case, while the NBA and NFL are above 40%.

Table 3A: Est. standard deviation at game level – 1991–2018.

	s_{game}	Worst case	Ratio of s_{game} to worst case
MLB	7.94%	50%	15.88%
NHL	10.83%	50%	21.66%
NBA	20.36%	50%	40.73%
NFL	20.98%	50%	41.96%

Up to this point, our estimates have focused on the four major sports leagues in North America. It is interesting to compare the results with both European soccer and with college basketball. Game-level variances for the soccer leagues lie around halfway between the NHL and the NBA/NFL levels; see Table 3B.[9]

Table 3B: Est. standard deviation at game level – major European soccer leagues.

	Time period	s_{game}	Worst case	Ratio of s_{game} to worst case
Bundisliga	1999–2018	13.88%	50%	27.76%
EPL	1992–2018	15.25%	50%	30.49%
La Liga	1999–2018	13.78%	50%	27.55%
Series A	1998–2018	16.12%	50%	32.25%

9 Here, in calculating the observed variance, we allocate each team a half of a win when there is a tie. The resulting ranking across soccer leagues is similar to the HICB ranking calculated in Ramchandani et al. (2018).

Given that college basketball does not employ many of the competitive balance tools used in professional leagues (e.g., a player draft), it is perhaps not surprising that we see significantly less balance at the game level in all the major conferences compared to the NBA and the other professional leagues. This is in spite of the fact that teams tend to play most of their schedule against teams comparable in quality. While schedules are not perfectly balanced, our estimates are close to the upper part of the range in equation (21) and the lower bounds on s_{game} are within 0.6% from the estimates below. Thus, even at the lower bound, the estimated s_{game} for college basketball is above that for the NBA; see Table 3C.

Table 3C: Est. standard deviation at game level – college basketball – 1991–2018.

	s_{game}	Worst case	Ratio of s_{game} to worst case
ACC	22.99%	50%	45.98%
Big 12	23.71%	50%	47.43%
Big Ten	24.70%	50%	49.40%
Pac 12	24.36%	50%	48.71%
SEC	21.49%	50%	42.97%

7.3 Adding a home-team advantage

It is straightforward to add a home-team advantage to the game-level model. If the advantage is of size h then we have a home and away levels of q_{ij}: $q_{ij} \pm h$. Assuming all the probabilities remain between 0 and 1, and that home and away games are balanced for each team, the average value of q_i is unchanged and hence the true value of team-level variance is unchanged.[10] In contrast, the variance at the game level goes up by h^2,

10 In the NHL and MLB, even when the best team faces the worst, the chances of winning (on neutral ground) is at or below 75%; see Table 5. Given an estimated home-field advantage of 5% in the NHL and 3% in MLB, the win probabilities are never above 80% or below 20%. In the NFL and NBA, the situation is a bit different. There are 435 possible pairings, each of which can be played with two options for the home team. There is never an issue in 435 pairings when the stronger team is the away team. For the other 435 pairings, there is a potential issue in our linear rank-order model. In the NFL, the estimated win probability would slightly exceed 1 in 10 of the pairings and for 15 pairings in the NBA. In these cases, we set the win probability to 1 (and the corresponding loss probability to 0). Given the small number of pairings effected (1.2% in NFL and 1.8% in the NBA) and the small change required, this has a very small effect on our estimates.

Table 4: Est. standard deviation at game level with home advantage – 1991–2018.

	h	s_{game} no home	s_{game} with home
NBA	10.0%	20.36%	22.75%
NHL	5.3%	10.83%	12.12%
NFL	7.9%	20.98%	22.67%.
MLB	3.8%	7.94%	8.83%

$$\sigma^2_{game\,(h)} = \frac{1}{2T(T-1)} \sum_{i=1}^{T} \sum_{j \neq i} \left[\left(q_{ij} + h - 0.5\right)^2 + \left(q_{ij} - h - 0.5\right)^2 \right] \tag{45}$$

$$= \frac{1}{T(T-1)} \sum_{i=1}^{T} \sum_{j \neq i} \left(q_{ij} - 0.5\right)^2 + h^2. \tag{46}$$

The intuition for this is straightforward—adding a home team advantage makes the individual game results more spread out and thus increases their variance even if the season-level records are unchanged. Our new estimate $s^2_{game(h)}$ is

$$s^2_{game\,(h)} = 2\frac{T-1}{T}\, s^2 + h^2. \tag{47}$$

And, just as before in (18) and (19), our observed value of the sample-size variance should be adjusted upward by h^2/n. This leads to an unbiased estimate of game-level variance taking into account a home-game advantage,

$$s^2_{game\,(h)} = 2\left[\frac{n(T-1)}{(n-2)T+2}\right]\left[\widehat{\sigma}^2 - \frac{0.25 - h^2}{n}\right] + h^2. \tag{48}$$

The direct effect in (48) accounts for most of the effect: adding a home-team advantage increases the estimate of s_{game} by 8% in the NFL, 11% in the MLB, 12% in the NBA and NHL.[11] Adding this factor switches the position of the NBA and NFL, although they are still essentially tied—the NBA's average s_{game} rises to 22.75% and the NFL's rises to 22.67%; see Table 4.

The calculations that follow all include the home-team advantage, using our estimates from the 1991 through 2018 seasons.

7.4 Additional game-level measures of balance

The Power-Rank model allows us to solve for Δ to see the implied estimate for the winning advantage of being one rank ahead. Rearranging equation (32) leads to:

$$\Delta^2 = \frac{12(T-1)}{T^2(T+1)}\,\sigma^2 \tag{49}$$

$$= \frac{12(T-1)}{T^2(T+1)}\left[\frac{n}{n-2+\frac{2}{T}}\right]\left[\widehat{\sigma}^2 - \frac{0.25-h^2}{n}\right] \tag{50}$$

We use the formulation in Eq. (50) to provide an estimated value of Δ. One might be tempted to estimate Δ directly from (31): take the win percentage of the top team minus the win percentage of the bottom team and divide by T. This observed value of Δ overstates the true value. Even if the true value of Δ is zero, so that all teams have equal win probabilities, we would still expect the best record to be above 0.500 and the worst to be below 0.500 in any finite season. (More generally, because teams other than $i = T$ may end up with the best record and teams other than $i = 1$ with the worst, there is a wider gap in the observed records than the true Δ would indicate.) The fact that there is a linear relationship between σ^2 and Δ^2 and that σ^2 can be estimated from $\widehat{\sigma}^2$ leads to a simple and straightforward way to get an unbiased estimate of Δ^2, which we take the square root for our estimate of Δ.

There is a debate in the literature about the value of measures of competitive balance versus measures of uncertainty of outcome; see Fort and Maxcy (2003). We agree that competitive balance measures at the team level are not directly connected to uncertainty of outcome. For that reason, we prefer balance measures at the game level. If games are well balanced then their outcomes will be hard to predict. Game-level variance is just one such measure. We consider two additional measures which indicate both balance and uncertainty at the game level, followed in the next section by an uncertainty-of-outcome metric at the season level.

One revealing measure of competitive imbalance and uncertainty of outcome is the chance the highest-ranked team beats the lowest-ranked team: $q_{T1} = 0.5 + (T{-}1)\Delta$. Applying (49), we have:

$$(T-1)\Delta = \frac{T-1}{T}\,\frac{\sigma}{\sqrt{\frac{T+1}{12(T-1)}}} \tag{51}$$

The right-hand side is $\frac{T-1}{T}$times the ratio of the true standard deviation in team records to the maximum possible standard deviation (that arises when higher-ranked teams are sure to beat lower ranked teams). If that ratio is ½, the highest-ranked team is nearly sure to beat the lowest-ranked team.

The Highest-Rank Team column in Table 5 calculates the predicted average win probability for the best team in each league. In our estimated model, the best NBA and NFL teams are predicted to win a little over 76% and 77% of their

11 The home advantage for each league is based on the average difference in winning probabilities for home games versus away games. Our estimates are similar to those in Trandel and Maxcy (2011), updated for additional years, and those in Lopez, Matthews, and Baumer (2018).

Table 5: Estimated "One-Spot" ranking advantage – 1991–2018 seasons.

	Estimated Δ	Highest-rank team	Highest vs. Lowest	Obs. Highest vs. Lowest
NBA	1.69%	76.3%	96.5%	96.2%
NHL	0.93%	64.0%	75.2%	80.3%
NFL	1.65%	77.2%	96.9%	97.1%
MLB	0.65%	60.2%	68.5%	74.9%

Table 6: Estimated probability of high-ranked team winning – 1991–2018 seasons.

	Estimated Δ	q_{HRW}
NBA	1.69%	67.0%
NHL	0.93%	59.0%
NFL	1.65%	67.6%
MLB	0.65%	56.6%

games. For MLB and the NHL, the highest-ranked teams are predicted to win 60% and 64% of the time. That's one sense in which the NHL and MLB are comparable with relatively little dispersion in win probabilities, while the NBA and NFL are also comparable with greater dispersion.

The Highest versus Lowest column highlights the issue of competitive balance at the game level. Here we look at the estimated q_{ij} when the top-ranked team competes against the lowest-ranked team. Even the lowest-ranked team has a 31% chance of beating the top-ranked team in MLB and a 25% chance in the NHL. In contrast, the model predicts there is a 3.5% chance that the lowest-ranked team beats the top-ranked one in the NBA, and that such an upset is even less likely in the NFL.

The last column in Table 5 provides the empirical probability of the best team beating the worst. This is not a perfect measure as we designated the best and worst teams by their end-of-season records. In addition, the leagues don't have perfectly balanced schedules. Even so, we see that the historical data line up reasonably well with the model's predicted results.

We can also use our model to calculate a related metric of competitive balance, q_{HRW}, the chance that in any matchup (at a neutral site) the higher-ranked team wins; see Lopez, Matthews, and Baumer (2018). In Calculation #6 of the Appendix, we derive the value of q_{HRW}. The result is a simple formula:

$$q_{HRW} = \frac{1}{T(T-1)} \sum_{i=1}^{T} \sum_{j \neq i}^{T} \mathrm{Max}\left[q_{ij}, 1 - q_{ij}\right] \tag{52}$$

$$= \frac{1}{2} + \frac{\Delta\,(\mathrm{T}+1)}{3}. \tag{53}$$

In a league with 30 teams, the average matchup has one team ranked 31/3 or roughly 10 spots ahead of its rival.

In the average game in MLB, the better team has a 56.6% chance of winning, while in the NFL the better team has a 67.6% of winning; see Table 6. This is another way of indicating the greater unpredictability in MLB and the NHL compared to the NBA and NFL. Our results are very similar to the estimates in Lopez, Matthews, and Baumer (2018), although their estimates use data from 2006 to 2016 and rely on a much more complicated estimation.

8 Full-season balance

Season length amplifies the differences in winning probabilities. Even though games in MLB are more evenly matched than in the NFL, over a 162-game season those small differences will get compounded and so the end-of-season ranking may be more predictable. This leads us to explore the question of full-season balance. We do so by looking at the predictability of the final rankings as seen from the start of the season. If final rankings are more predictable at the season start, then there is less likelihood of reversals in league standings. Thus, this approach can also be thought of as a baseline for intra-season competitive intensity metrics such as how often team rankings change; see Neale (1964), Scelles, Desbordes, and Durand (2011), Andreff and Scelles (2015).

The two drivers of this measure are the Δs for each league and the season length. So far, our competitive balance metrics have been designed to factor out season length. To measure the predictability of end-of-season rankings, season length is a key component. Our measure of predictability is based on the chance that a team with rank i comes out at the top at the season end. We look at the chance the top-ranked team ends up first, how many teams have at least a 1% chance of coming in first, and the variance in these probabilities.

The probabilities are calculated via 10 million simulations. Each team's final record is the average of a season-length's number of independent draws, which converges to a normal distribution. In the simulation, we model each team's full-season record as a draw from a normal distribution with mean equal to its average winning probability q_i as calculated in (31) using the average Δ from the 1991–2018 seasons and variance that reflects the win probabilities against each opponent.[12] The ordering of the end-of-season records determines the ranking.

12 $\sigma_i^2 = \frac{1}{n(T-1)} \sum_{j \neq i} q_{ij}(1 - q_{ij})$.

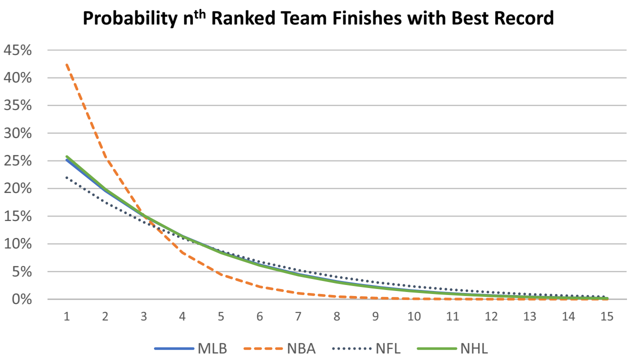

Figure 3: Probability nth-ranked team finishes with best record.

The chance that the best team in the NHL will end the season with the best record is 26%, and the corresponding number is 25% for MLB. While the best team in the NHL has a higher average win probability than the best team in MLB (64.0% vs. 60.2%), the nearly double season length in MLB means there is almost the same chance the best team in MLB ends the season at the top.

When comparing the NFL and the NBA, the differences become larger as the NBA season is more than five times as long. While the average win rate of the best teams in both leagues are almost equal (76.3% vs. 77.2%), the chance that the best team ends the season at the top is 42% in the NBA versus 22% in the NFL. In a representative season in the NBA, 7 teams have more than a 1% chance of having the best record, while there are 10 such teams in the NHL, 11 in MLB and 12 in the NFL; see Figure 3. Of course, the number of teams viewed at the outset as having a practical chance of ending with the best record may be larger due to uncertainty in identifying the top teams at the beginning of each season.

This greater dispersion can also be seen in terms of the variances in the probabilities of coming in first place. For the NBA, the variance in probabilities is 0.84%, while the variance is 0.41% in MLB and the NHL, and 0.31% for the NFL. Greater variance in the probabilities means the top teams have much more than a $1/T$ (roughly 3.2%) chance of coming in first, and the bottom teams have lower chances.

9 Conclusions

In order to make comparisons of competitive balance across sports leagues, we need to take into account how different season lengths influence observed measures of balance. To our knowledge, we develop the first measures of competitive balance that are invariant to season length. By looking at balance at the game level rather than just the team level, our competitive balance measures also provide insight into uncertainty of outcomes.

The commonly used N-S metric is known to be biased. The HHI and related measures are also biased. To develop an unbiased metric requires a model. In particular, it is not correct to just subtract the ideal standard deviation from the observed value rather than divide by it. We present a general model of competition which shows the relationship between the true variance and the expected value of the observed variance. The difference depends on the number of games as well as the variability of each team's winning probability from its individual mean. While we do not have a closed-form solution for the estimated variance based on observables, we provide sufficiently tight bounds that allow us to make comparisons across leagues.

These bounds show that the fundamental flaw in the N-S metric invalidates its conclusions related to comparisons of competitive balance across leagues. We find that MLB and the NHL have the greatest level of balance at the team-level. At the other end, we find that the NFL and NBA have roughly equal levels of team-level competitive balance using a metric that correctly controls for season length.

While variance estimates and related measures of static competitive balance metrics provide an indication of how close teams are in terms of their win probabilities, this doesn't necessarily indicate how evenly they are matched at the game-level. Game-level balance is better connected to excitement and uncertainty of outcome. This leads us to create a measure of balance that focusses on game-level competition.

To do so, we simplify the general model to the case where win probabilities depend linearly on the difference in ranks between the competing teams. This rank-order model leads to easy-to-apply unbiased estimates for game-level variance and HHI metrics based solely on aggregate statistics. We find that the relative ranking of game-level balance across leagues is unchanged from team-level rankings. Including a home-game advantage leads to a small increase in imbalance at the game level, although it has almost no effect at the team level.

Along with making comparisons across leagues, we use the rank-order model to help interpret what the imbalance means in terms of uncertainty of outcomes. In an average MLB game, the higher-ranked team wins 57% of the time and 59% in the NHL, while in the NBA and the NFL, the higher-ranked team wins about 67% of the time. At the extreme, when the top-ranked team meets the lowest-ranked, the top-ranked team wins 97% of the time in the NFL and 96% in the NBA, while the lowest-ranked team has a 25% chance in the NHL and nearly a 32% chance in MLB.

Fans care about full-season results in addition to winning individual games. This leads us to look for a way to measure the predictability of end-of-season rankings at the season start. The ranking at the end of the season is the result of many individual imbalanced games, and thus differences in team winning probabilities become amplified as the season length grows. End-of-season rankings are least predictable in the NFL, closely followed by MLB and the NHL. The NFL rankings are unpredictable due to its short season, while the MLB rankings are unpredictable due to its small dispersion of winning probabilities (in spite of having the longest season). The NHL is unpredictable due to having nearly as little dispersion as MLB combined with a shorter season. The NBA is the league where the best teams are most likely to end up at the top of the standings; the difference in win probabilities between teams is comparable to the NFL, but over a season that is more than five times longer, higher win probabilities conquer randomness.

In a well-balanced sports league, teams are evenly matched, game are exciting, and championships are hard to predict. We believe the models and metrics presented in this paper provide a simple unbiased view of how sports leagues stack up based on competitive balance at the team, game, and season levels. We hope our approach will also help shift focus from team-level measures to competitive balance and predictability at the game and season level.

Link to public data: https://www.dropbox.com/s/k0v3ne7xrpiprt6/JQAS%20Data.xlsx?dl=0.

Author contribution: All the authors have accepted responsibility for the entire content of this submitted manuscript and approved submission.
Research funding: None declared.
Conflict of interest statement: The authors declare no conflicts of interest regarding this article.

Appendix: Calculations

Calc #1

$$E[\hat{\sigma}^2] = \frac{1}{T}\sum_{i=1}^{T} E\,[w_i - 0.5]^2 \tag{A1}$$

$$= \frac{1}{T}\sum_{i=1}^{T} E\left(\sum_{k=1}^{n}\frac{(w_{ik} - p_{ik} + p_i - 0.5)}{n}\right)^2 \tag{A2}$$

$$= \frac{1}{T}\sum_{i=1}^{T}\left[(p_i - 0.5)^2 + E\sum_{k=1}^{n}\left(\frac{w_{ik} - p_{ik}}{n}\right)^2\right] \tag{A3}$$

$$= \sigma^2 + \frac{1}{T}\sum_{i=1}^{T}\sum_{k=1}^{n}\frac{p_{ik}(1 - p_{ik})}{n^2} \tag{A4}$$

$$= \sigma^2 + \frac{1}{T}\sum_{i=1}^{T}\sum_{k=1}^{n}\frac{0.25 - (p_{ik} - 0.5)^2}{n^2} \tag{A5}$$

$$= \sigma^2 + \frac{0.25}{n} - \frac{1}{T}\sum_{i=1}^{T}\sum_{k=1}^{n}\frac{(p_{ik} - p_i + p_i - 0.5)^2}{n^2} \tag{A6}$$

$$= \sigma^2 + \frac{0.25}{n} - \frac{1}{T}\sum_{i=1}^{T}\sum_{k=1}^{n}\frac{(p_{ik} - p_i)^2 + (p_i - 0.5)^2}{n^2} \tag{A7}$$

$$= \sigma^2\frac{n-1}{n} + \frac{0.25}{n} - \frac{1}{T}\sum_{i=1}^{T}\sum_{k=1}^{n}\frac{(p_{ik} - p_i)^2}{n^2} \tag{A8}$$

Calc #2

$$E = -\frac{2}{Tn}\sum_{i=1}^{T}\sum_{k-1}^{n} p_{ik}\mathrm{Log}_2\,(p_{ik}) \tag{B1}$$

Let $h(p) = -p\ \mathrm{Log}_2(p)$. Then $h'(p) = -\mathrm{Log}_2(p) - 1/Ln(2)$ and $h''(p) = -\frac{1}{pLn\,(2)}$. Using a Taylor expansion around $p_{ik} = 1/2$, we have

$$E \approx \frac{2}{Tn}\sum_{i=1}^{T}\sum_{k=1}^{n} h\left(\frac{1}{2}\right) + h'\left(\frac{1}{2}\right)\left(p_{ik} - \frac{1}{2}\right) + \frac{1}{2}h''\left(\frac{1}{2}\right)\left(p_{ik} - \frac{1}{2}\right)^2. \tag{B2}$$

As the average value of p_{ik} over all (i, k) is 1/2, the linear term is zero. This leaves

$$E \approx 2h\left(\frac{1}{2}\right) + \frac{h''\left(\frac{1}{2}\right)}{Tn}\sum_{i=1}^{T}\sum_{k=1}^{n}(p_{ik} - 1/2)^2 \tag{B3}$$

$$= 1 - \frac{2}{Ln(2)}\ \sigma^2_{game}, \tag{B4}$$

where σ^2_{game} is the variance at the game level,

$$\sigma^2_{game} = \frac{1}{Tn}\sum_{i=1}^{T}\sum_{k=1}^{n}(p_{ik} - 1/2)^2. \tag{B5}$$

Calc #3

$$\sigma^2 = \frac{1}{T}\sum_{i=1}^{T}(q_i - 0.5)^2 \tag{C1}$$

$$= \frac{T}{(T-1)^2}\sum_{i=1}^{T}\left(i - \frac{T+1}{2}\right)^2\Delta^2 \tag{C2}$$

$$= \frac{T\Delta^2}{(T-1)^2}\sum_{i=1}^{T} i^2 - i(T+1) + \frac{(T+1)^2}{4} \tag{C3}$$

$$= \frac{T\Delta^2}{(T-1)^2}\left[\frac{T(T+1)(2T+1)}{6} - \frac{T(T+1)^2}{2} + \frac{T(T+1)^2}{4}\right] \tag{C4}$$

$$= \frac{T^2(T+1)\Delta^2}{(T-1)^2}\left[\frac{2T+1}{6} - \frac{T+1}{4}\right] \tag{C5}$$

$$= \Delta^2 T^2\frac{T+1}{12(T-1)}. \tag{C6}$$

Calc #4

$$\sigma^2_{game} = \frac{1}{T(T-1)} \sum_{i=1}^{T} \sum_{j \neq i} \left(q_{ij} - 0.5\right)^2 \tag{D1}$$

$$= \frac{1}{T(T-1)} \sum_{i=1}^{T} \sum_{j=1}^{T} (i-j)^2 \Delta^2 \tag{D2}$$

$$= \frac{\Delta^2}{T(T-1)} \sum_{i=1}^{T} Ti^2 - 2i\frac{T(T+1)}{2} + \frac{(2T+1)T(T+1)}{6} \tag{D3}$$

$$= \frac{\Delta^2 T}{T-1}\left[\frac{(2T+1)(T+1)}{6} - \frac{(T+1)^2}{2} + \frac{(2T+1)(T+1)}{6}\right] \tag{D4}$$

$$= \Delta^2 \frac{T(T+1)}{6}. \tag{D5}$$

Calc #5

$$\sigma^2_{samplesize} = \frac{1}{T(T-1)^2} \sum_{i=1}^{T} \sum_{j \neq i} E\left(w_{ij} - q_{ij}\right)^2 \tag{E1}$$

$$= \frac{1}{T(T-1)^2} \sum_{i=1}^{T} \sum_{j \neq i} \frac{T-1}{n}\left[\frac{1}{4} - (i-j)^2 \Delta^2\right] \tag{E2}$$

$$= \frac{1}{4n} - \frac{\Delta^2}{nT(T-1)} \sum_{i=1}^{T} \sum_{j \neq i} (i-j)^2 \tag{E3}$$

$$= \frac{1}{4n} - \frac{\Delta^2}{nT(T-1)} \sum_{i=1}^{T} \sum_{j=1}^{T} (i-j)^2 \tag{E4}$$

$$= \frac{1}{4n} - \frac{\Delta^2 \; T(T+1)}{6n}, \tag{E5}$$

where we use the same tools to calculate the sum of $(i–j)^2$ as in (D2)–(D5) above.

Calc #6

$$q_{\mathrm{HRW}} = \frac{1}{T(T-1)} \sum_{i=1}^{T} \sum_{j \neq i}^{T} \mathrm{Max}\left[q_{ij}, 1 - q_{ij}\right] \tag{F1}$$

$$= \frac{1}{T(T-1)} \sum_{i=1}^{T} \sum_{j \neq i}^{T} \frac{1}{2} + \left[\mathrm{Max}\,(i,j) - \mathrm{Min}\,(i,j)\right] \Delta \tag{F2}$$

$$= \frac{1}{2} + \Delta\left[\frac{2}{3}(T+1) - \frac{1}{3}(T+1)\right] \tag{F3}$$

$$= \frac{1}{2} + \frac{\Delta(\mathrm{T}+1)}{3}. \tag{F4}$$

References

Andreff, W., and N. Scelles. 2015. "Walter C. Neale 50 Years After: Beyond Competitive Balance, the League Standing Effect Tested with French Football Data." *Journal of Sports Economics* 16 (8): 819–34.

Berri, D., M. Schmidt, and S. Brook. 2007. *The Wages of Wins: Taking Measure of the Many Myths in Modern Sport*. Stanford: Stanford University Press.

Birnbaum, P. 2016. *Noll-Scully Doesn't Measure Anything Real*. https://blog.philbirnbaum.com/2016/04/noll-scully-doesnt-measure-anything-real.html (accessed January 17, 2020).

Boulier, B. L., and H. O. Stekler. 2003. "Predicting the Outcomes of National Football League Games." *International Journal of Forecasting* 19: 257–70.

Bradley, R. A., and M. E. Terry. 1952. "Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons." *Biometrika* 39 (3/4): 324–5.

Chen, M. K., J. E. Ingersoll, and E. H. Kaplan. 2008. "Modeling a Presidential Prediction Market." *Management Science* 54 (8): 1381–94.

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

Eckard, E. W. 1998. "The NCAA Cartel and Competitive Balance in College Football." *Review of Industrial Organization* 13: 347–69.

Ely, J., A. Frankel, and E. Kamenica. 2015. "Suspense and Surprise." *Journal of Political Economy* 123 (1): 215–60.

Fair, R. 2009. "Interpreting the Predictive Uncertainty of Elections." *The Journal of Politics* 71 (2): 612–26.

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

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

Glickman, M. E., and H. S. Stern. 1998. "A State-Space Model for National Football League Scores." *Journal of the American Statistical Association* 93: 25–35.

Goossens, K. 2006. "Competitive Balance in European Football: Comparison by Adapting Measures: National Measure of Seasonal Imbalance and Top 3." *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: 373–87.

Humphreys, B. 2002. "Alternative Measures of Competitive Balance in Sports Leagues." *Journal of Sports Economics* 3 (2): 133–48.

James, B. 1981. *1981 Baseball Abstract: The 5th Annual Edition*. Lawrence, KS: Bill James.

Knowles, G., K. Sherony, and M. Haupert. 1992. "The Demand for Major League Baseball: A Test of the Uncertainty of Outcome Hypothesis." *The American Economist* 36 (2): 72–80.

Lopez, M. J., G. J. Matthews, and B. S. Baumer. 2018. "How Often Does the Best Team Win? A Unified Approach to Understanding Randomness in North American Sport." *The Annals of Applied Statistics* 12 (4): 2483–516.

Mauboussin, M. 2012. *The Success Equation: Untangling Skill and Luck in Business, Sports, and Investing*. Boston: Harvard Business Review Press.

Michie, J., and C. Oughton. 2004. *Competitive Balance in Football: Trends and Effects*. London: Football Governance Research Centre FGRC research papers.

Neale, W. C. 1964. "The Peculiar Economics of Professional Sports." *Quarterly Journal of Economics* 78: 1–14.

Noll, R. 1988. "*Professional Basketball*" *Studies in Industrial Economics. Paper no. 144*. Stanford, CA: Stanford University.

Noll, R. 1991. "Professional Basketball: Economic and Business Perspectives." In *The Business of Professional Sports*, 18–47, edited by P. D. Staudohar, and J. A.Mangan. Champaign, IL: University of Illinois Press.

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., M. Ryan, and C. R. Weatherston. 2007. "Measuring Competitive Balance in Professional Team Sports Using the Herfindahl-Hirschman Index." *Review of Industrial Organization* 31: 289–302.

Owen, P. D., and N. King. 2015. "Competitive Balance Measures in Sports Leagues: The Effects of Variation in Season Length." *Economic Inquiry* 53 (1): 731–44.

Plumley, D. J., and S. Flint. 2015. "The UEFA Champions League: Maintaining the Status Quo?." *Team Performance Management* 21 (5/6): 247–58.

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

Rottenberg, S. 1956. "The Baseball Players' Labor Market." *Journal of Political Economy* 64 (3): 242–58.

Ramchandani, G., D. Plumley, S. Boyes, and R. Wilson. 2018. "A Longitudinal and Comparative Analysis of Competitive Balance in Five European Football Leagues." *Team Performance Management: An International Journal* 24: 265–82.

Scelles, N., M. Desbordes, and C. Durand. 2011. "Marketing in Sport Leagues: Optimising the Product Design. Intra-Championship Competitive Intensity in French Football Ligue 1 and Basketball Pro A." *International Journal of Sport Management and Marketing* 9 (1/2): 13–28.

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

Tango, T. 2006. *True Talent Level for Sports Leagues*. https://www.insidethebook.com/ee/index.php/site/article/true_talent_levels_for_sports_leagues/ (accessed January 17, 2020).

Trandel, G., and J. Maxcy. 2011. "Adjusting Winning-Percentage Standard Deviations and a Measure of Competitive Balance for Home Advantage." *Journal of Quantitative Analysis in Sports* 7 (1): 1–17.