Research Article

Estéfano B. Vassoler* and Pedro O.S. Vaz-de-Melo

An Interpretable and Probabilistic Competitive Balance Metric for Sports Leagues

https://doi.org/10.1515/sample-YYYY-XXXX

Received Month DD, YYYY; revised Month DD, YYYY; accepted Month DD, YYYY

Abstract: We propose a new probabilistic and interpretable approach to quantify competitive balance in sports leagues based on the precise moment when a tournament can no longer be perceived as perfectly balanced: the longer it takes, the more balanced it is. We analyzed 1539 seasons from 175 sports leagues in basketball, soccer, handball, and volleyball and observed that only 5% of the seasons could be seen as perfectly balanced throughout their entire duration. For the others, there is a turning point round after which the points distribution permanently diverges from a range of behaviors likely to occur in perfectly balanced tournaments. Our initial results agree with the general literature since soccer has a considerably more random behavior, i.e., its perceived balance is higher overall. Given the explicit temporal dependence, we also proposed a modification to remove the bias that the order of matches could impose, thereby enabling fair comparisons across different leagues and sports. Our modified coefficient highlights the substantial imbalance in heavily debated leagues, such as the Premier League and the NBA. Furthermore, combining both metrics enables the discovery of anomalous tournament schedules where changes to the match order would have improved the perceived competitive balance.

Keywords: Sport Analytics, Competitive Balance, Probabilistic Metrics, Knowledge Discovery from Data.

***Corresponding author: Estéfano B. Vassoler, Pedro O.S. Vaz-de-Melo,** Universidade Federal de Minas Gerais, Departamento de Ciência da Computação, Belo Horizonte, Brazil, e-mail: estefanobvassoler@gmail.com, olmo@dcc.ufmg.br

1 Introduction

A round-robin sports league is a format in which every team competes against each other, and the winner is the one that accrues the highest number of points throughout its duration. While this format accurately showcases the relative skills of each team in their final standings if the skill level does not change throughout the tournament (Sziklai et al., 2022), it can sometimes struggle to sustain audience interest, particularly if there is a significant disparity in skill levels among the teams (Neale, 1964; Fort and Quirk, 1995). In such cases, only a handful of teams remain competitive for the championship as the league advances, leading to matches between teams that lack any significant motivation or aspirations, which may not be as appealing to fans (Neale, 1964; Douvis, 2014).

The concept of competitive balance - referring to a well-balanced distribution of sporting talent across teams - is widely considered fundamental to a league's long-term viability and success (Rottenberg, 1956; Jennett, 1984; Humphreys, 2002). By creating unpredictable match outcomes and maintaining close championship races, competitive balance enhances fan engagement, motivates teams to remain competitive throughout the season, and ensures a healthy rotation of winners over time (Manasis et al., 2022; Gerrard and Kringstad, 2022).

However, the empirical relationship between competitive balance and audience demand is nuanced and context-dependent. Some studies find no significant link to stadium attendance or television viewership, suggesting that factors like team quality or star players dominate audience engagement (Scelles, 2017; Caruso et al., 2019; Wills et al., 2022, 2023; Macedo et al., 2023; Wang, 2025). Others report an inverse relationship, often attributed to loss aversion—fans' preference for their team winning over outcome uncertainty (Coates et al., 2014; Besters et al., 2019; Baydina et al., 2021; Hyun et al., 2023). Conversely, additional research demonstrates positive effects (Jennett, 1984; Cox, 2018; Forrest and Simmons, 2002; Eckard, 2017; Schreyer et al., 2018a; Reilly, 2023; Ferguson and Lakhani, 2023), particularly in high-stakes matches, among neutral spectators, or in leagues with less entrenched fan bases.

Notwithstanding this ongoing debate, a consistent body of evidence confirms that competitive balance significantly shapes overall fan experience (Buraimo and Simmons, 2008; Schreyer et al., 2018b; Collins and Humphreys, 2022; Hyun et al., 2023; van der Burg, 2023; Sheng and Montgomery, 2025). Unsurprisingly, organizations like the NBA have instituted compensation mechanisms such as player drafts to benefit underperforming teams (Kesenne, 2006; Winfree and Fort, 2012), salary caps to limit spending (Dietl et al., 2011; Maxcy and Milwood, 2018), and veto powers to prevent the formation of overly dominant teams (Plumley et al., 2019). All these factors have stimulated substantial methodological innovation, with researchers developing increasingly sophisticated tools to measure competitive balance across different contexts (Humphreys, 2002; Owen, 2014; Doria and Nalebuff, 2021). While this work mainly focuses on analyzing competitive balance within a tournament season, often called seasonal (or within-season) competitive balance, its other dimensions are also noteworthy (Kringstad and Gerrard, 2004; Szymanski, 2003; Gerrard and Kringstad, 2022). At the match level, competitive balance is related to *uncertainty of outcome* and treats each game individually. In this scenario, perfect balance is achieved by teams having equal chances of victory (Benz et al., 2009). *Dominance* or *championship* describes long-term competitive balance, analyzing whether different teams have recently been champions. Ideally, successive seasons should not have identical winners (Ramchandani et al., 2018).

Given that the true skill levels of the teams are unknown, within-season competitive balance is inferred through the observed tournament data, such as the points distribution and match results (Zimbalist, 2002). The problem, however, is that different metrics capture distinct facets of competitive balance (Haan et al., 2007), and there is no consensus about which is the best way to infer and measure it (Manasis et al., 2022; Gerrard and Kringstad, 2022). The simplest approach is to calculate standard measures of dispersion, concentration and inequality on the teams' wins or points in a season (Zimbalist, 2002). However, they do not take into account several factors that directly affect the measurements but are independent of the inherent competitive balance, making comparisons across different sports and leagues challenging (Owen, 2013). Because of this, extensions were proposed to account for season length, number of teams, the structure of schedules and the points allocation system (Michie and Oughton, 2004; Borooah and Mangan, 2011; Ramchandani, 2012; Criado et al., 2013; Owen and King, 2015; Doria and Nalebuff, 2021; Avila-Cano et al., 2023). Typically, these refined versions incorporate normalization factors derived from the upper or lower bounds of competitive balance in ideally balanced leagues with analogous attributes (e.g. season length).

2 Contributions

Although normalized metrics enhance the ability to compare competitive balance across various sports and leagues, their values, particularly intermediate ones, are not very informative and hard to interpret. Put differently, they alone cannot be easily translated into accessible descriptions of the competitive balance within the league, often necessitating a reassessment of the metrics through thorough reviews (Owen, 2013;

Ramchandani et al., 2018; Gerrard and Kringstad, 2022). In addition, these metrics are deterministic and highly sensitive to the observed tournament data (e.g. the points distribution). They neglect the stochastic nature of match outcomes, which can make leagues sharing the same underlying (true) competitive balance exhibit varying point distributions with different levels of dispersion. In Figure 1, the green markers depict the expected variance of the points distribution (termed *competitive imbalance*) after each matchday in perfectly balanced leagues, simulated 1000 times using the actual matches and schedules of real leagues. The shaded areas delineate the 5th and 95th percentiles from simulated match outcomes (home team wins, home team loses or draw) defined randomly using the real league's frequencies. Observe that the size of these regions is not negligible, so even in perfectly balanced tournaments, significant differences in variances may emerge as the league progresses.

After comparing the progression of the competitive imbalance of real tournaments with their corresponding perfectly balanced simulated tournaments, we observed a distinct pattern. For almost all cases, the real competitive imbalance remains within the shaded area until a specific matchday τ is reached. Beyond this point, the real variance permanently diverges, displaying a more pronounced rate of increase. Figure 1 shows this pattern for four real leagues, each one belonging to a different sport, namely, soccer, basketball, volleyball, and handball. In the first τ dates, the real competitive imbalance is indistinguishable from its perfectly balanced counterpart, where all matches are decided solely by chance. Then, as the competitive imbalance of the real league grows more rapidly, it diverges from the expectation. Remarkably, such matchday τ varies greatly across tournaments and sports, suggesting its potential utility for characterizing competitive balance in an interpretable manner.

Thus, our first contribution is to illustrate that both the matchday τ and the fraction of matchdays occurring before τ serve as intuitive and interpretable measures of competitive balance in games from 157 sports leagues, comprising 1,539 seasons across 77 countries. As our second contribution, recognizing the direct influence of match order on τ, we introduce the $\hat{\tau}$ estimator for the *expected* τ of a tournament. This coefficient is independent of the schedule and effectively measures competitive balance, allowing comparisons across different leagues and sports. Additionally, combining our metrics enables identifying tournaments whose *observed* turning point is consistently below its *expected* value throughout the seasons. In these situations, slight changes to scheduling could meaningfully improve the perception of balance.

Finally, our results reveal that both the *observed* and the *expected turning point* τ vary significantly from league to league. Specifically, for basketball, volleyball, and handball, the *expected* τ should occur, on

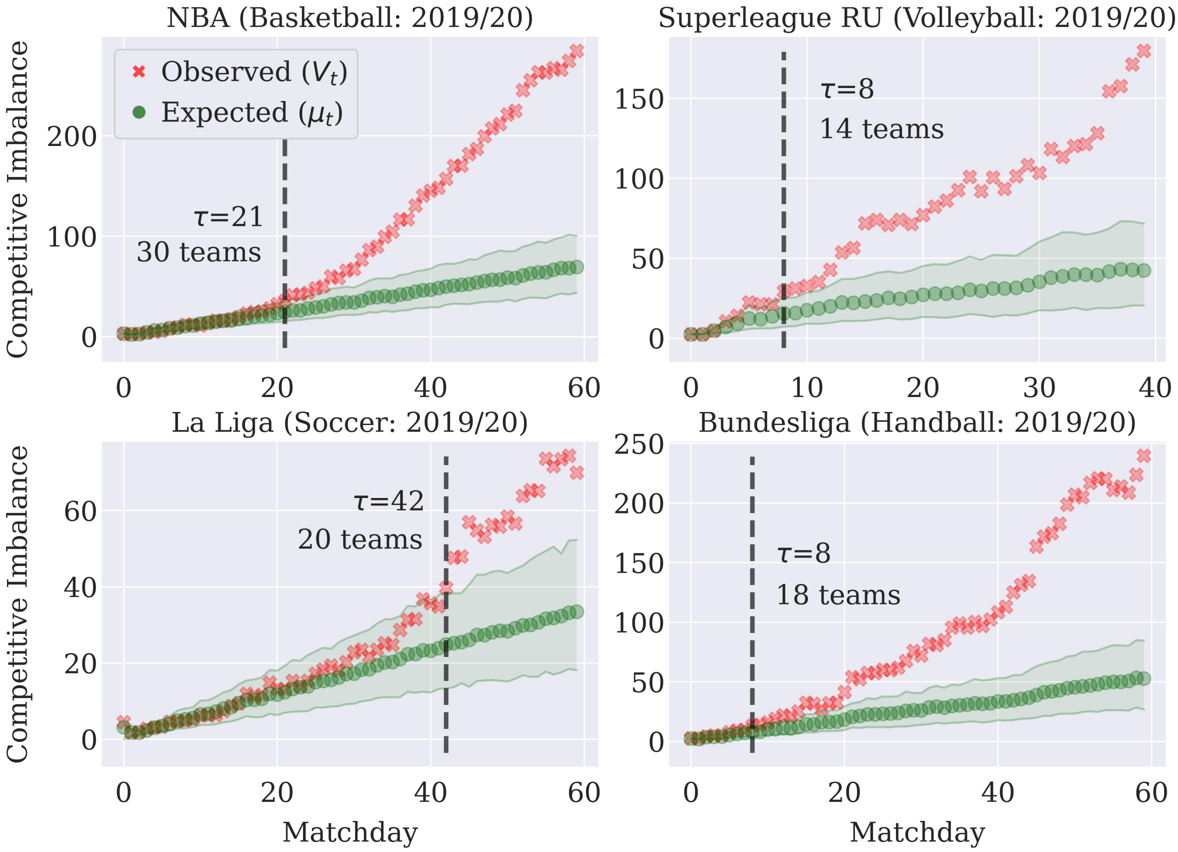

Fig. 1: Temporal progression of the competitive imbalance. The red crosses are the empirical ranking variances $\mathbf{V_t}$ at each time $\mathbf{t}$. The green points represent the expected value $\mathbb{E}(\mathbf{V_t})$ for this variance if all the teams had the same skills, while the shaded green region delimits a 95% confidence envelope. The vertical dashed line is the first time the empirical variance curve leaves the confidence envelope without ever returning: it is the *turning point* τ. Even though the basketball, soccer, and handball leagues lasted longer than 60 dates, their behavior did not change after that.

average, respectively after 30.7%, 26.0%, and 25.4% of season matches. As expected from the literature (Aoki et al., 2017; Ben-Naim et al., 2005), soccer leagues have a distinguished and more random behavior, so the turning point should occur, on average, after 43.5% of the season matches. Additionally, our metrics also highlight the substantial imbalance in highly scrutinized leagues, such as the Premier League and the NBA.

3 Related Work

In sports leagues, seasonal competitive balance assesses the level of uniformity in the skill distribution of the teams. Since the actual skill levels are typically unknown, competitive balance is instead inferred from observed tournament data, including match records and standings (Zimbalist, 2002). The most common approach is to apply metrics of *dispersion* and *inequality* to the distribution of points and victories in the standings (Quirk and Fort, 1997; Zimbalist, 2002; Humphreys, 2002; Owen, 2014; Michie and Oughton, 2004; Doria and Nalebuff, 2021; Manasis et al., 2022). When applied to the percentage of matches won by each team, the standard deviation (Scully, 1989; Quirk and Fort, 1997) and the Gini coefficient (Quirk and Fort, 1997; Schmidt and Berri, 2001) are examples of simple metrics for measuring competitive balance.

However, these metrics exhibit high sensitivity to factors such as the number of teams and season length (Utt and Fort, 2002; Owen and King, 2015; Gerrard and Kringstad, 2022). To address this challenge, normalized extensions have been proposed to facilitate comparisons across different leagues (Haugen, 2008; Owen, 2010; Aoki et al., 2017; Triguero Ruiz and Avila-Cano, 2019). Among these, variations of the Herfindahl–Hirschman index (HHI) are broadly favored (Michie and Oughton, 2004; Borooah and Mangan, 2011; Manasis et al., 2011; Ramchandani, 2012; Triguero Ruiz and Avila-Cano, 2019; Avila-Cano et al., 2023), particularly for their ability to control for the number of teams, season length, incorporate win-loss records of all teams, and accommodate variations in leagues with tied matches and distinct point systems (Gerrard and Kringstad, 2022). Nonetheless, the Shannon entropy (Horowitz, 1997; Jadbabaie et al., 2020; Manasis, 2022), the Lorentz curve (Michie and Oughton, 2004; Ramchandani, 2012), the top-k concentration ratio (Michie and Oughton, 2004; Ramchandani, 2012), and variations of the inter-quartile range (Ramchandani, 2012) are also utilized, with the last two being noteworthy for their ease of interpretation.

From match records, Groot and Groot (2003) and Ben-Naim et al. (2005) have assessed competitive balance in terms of upset frequency, representing the fraction of times the team with the worse record on the matchday actually won. Naturally, a higher upset frequency is indicative of a more balanced tournament. Criado et al. (2013) proposed a graph-based approach where nodes represent teams, and edges are weighted based on the number of times the teams exchanged positions in the rankings. High competitive balance is associated with both the total number of teams that exchanged positions and the total number of switches. They utilize four metrics to analyze graphs, with the simplest one being the average amount of these exchanges (*normalized mean degree*).

Different from the aforementioned deterministic metrics, our proposed approach is founded on a probabilistic framework, drawing inspiration from a pattern observed in various sports leagues and in Bradley-Terry simulations (Bradley and Terry, 1952; Cattelan et al., 2013). For the initial τ rounds, the standings are not statistically distinguishable from those generated by perfectly balanced tournaments. Subsequently, a permanent divergence occurs. As a result, we characterize competitive balance by the proportion of these initial rounds where the tournament aligns with the characteristics of a perfectly balanced competition. Notably, our metric stands out for its high interpretability, making it easily accessible even to the general public – a feature distinct from most of the existing metrics. Also, it systematically controls for all the previously mentioned factors, enabling meaningful comparisons across different leagues and sports.

4 Dataset

The proposed measures of competitive balance draw inspiration from observations made on real data, specifically tracking the progression of the points distribution across a diverse array of sports tournaments. We conducted a comprehensive study encompassing all seasons from 35 basketball, 30 volleyball, 80 football, and 30 handball leagues, as available on the betting site `betexplorer.com`. The dataset covers seasons from 2011 to 2021, with all 11 seasons available for approximately half of the leagues. For 7% of the leagues, data were available for fewer than 4 seasons. In total, there are 311,714 matches spread out across 1,539 seasons, averaging approximately 202 games per season.

To ensure the stability of our statistical procedures, we only included seasons with a minimum of 8 teams and at least 50 matches. In addition, in cases where league divisions underwent official name changes during the data collection period, we aggregated them under a unified name before our analysis. For each game, we collected information about which team played at home, the result of the match, when it was played, and, when available, the odds for the game. These odds serve as indicators, according to the betting market, of which team is considered the favorite to win the match. For *betexplorer.com*, these values are the average *closing-odds* over various betting sites. As such, they should mitigate some biases, and be a better estimation of the underlying probabilities. This aligns, for instance, with Wilkens (2020), which found that using model ensembles - combining predictions from multiple sources - is the most promising approach for predicting tennis match outcomes.

The inclusion of odds is crucial for our analysis as tournaments with higher competitive balance are expected to be more unpredictable (Forrest and Simmons, 2002; Szymanski, 2003; Owen, 2014). In other words, the betting market is more likely to encounter challenges in accurately predicting the match outcomes in balanced leagues. In total, we managed to collect odds for about 77% of all seasons in our dataset since historical betting data is not always available.

5 Measuring Competitive Balance

We introduce a novel framework for assessing competitive balance in sports leagues. Like traditional approaches, our methodology employs the standings during the regular season as an indicator of underlying team skill disparities - where greater dispersion reflects more pronounced competitive imbalance. However, our approach extends beyond conventional methods in three key aspects:

1. Rather than relying solely on final standings, we implement a probabilistic comparison between observed imbalance patterns at each tournament round and simulations of perfectly balanced tournaments (equally skilled teams).
2. We derive two new metrics based on the duration during which observed and simulated imbalances remain statistically indistinguishable, with longer periods indicating greater competitive balance.
3. As demonstrated in Section 8.2 (*Framework Invariance*), our framework is compatible with various existing imbalance measures (e.g., variance, Gini coefficient), providing methodological flexibility.

In short, a tournament is considered perfectly balanced up to round τ if no statistically significant differences exist between its observed imbalance trajectory and the simulated expectation through round τ.

5.1 Null Model

The standings of balanced tournaments can progress in vastly different ways. In some, a team might start well and falter toward the end. In others, a team might be invincible or even collapse from the start. Unfortunately, standard competitive balance metrics fail to consider these intricacies by looking solely at a snapshot of the final standings. Our competitive balance metrics innovate by incorporating the influence of random *outcome fluctuations* to determine whether the tournament *progressed* as a balanced one. To do so, we compare the observed skill imbalance in the real tournament to those generated by a null model that replicates the characteristics of the tournament, except with all teams being equally skilled. In this scenario, the purpose of the null model is to comprehensively capture the point-distribution variability in tournaments where all teams share an identical skill level. A tournament characterized by a significant gap in skill levels would inevitably lead to a highly skewed points distribution – an unlikely outcome in a random, perfectly balanced tournament.

The null model simulates a perfectly balanced version of a competitions's regular season following the same schedule S as the original tournament $\mathcal{C}$. If all teams $(1, \ldots, n)$ have the same strength, the standings $\mathcal{X}_t = (X_{1,t}, \ldots, X_{n,t})$ up to round t is composed of identically distributed (i.d.) random variables which are not independent since the points a team obtain in a match are directly coupled to its adversary's points. We consider that the points accumulated by a team can be expressed by $X_{i,t} = x_{i,1} + \ldots + x_{i,t}$, that is, the sum of successive games points $x_{i,j}$ with three possible outcomes: 3 points for a win, 1 point for a draw, and 0 points for a loss. In Appendix A, we show the results for volleyball under another point system.

The likelihood of these outcomes is directly tied to the probability of each match result: home win (P_h), draw (P_d), and away win (P_a) – which are empirically estimated based on the observed frequencies in the real tournament $\mathcal{C}$. In the simulations, each match outcome is randomly assigned according to P_h, P_d, and P_a, replicating the distribution observed in $\mathcal{C}$. This approach disregards teams' individual skills, considering only home-court advantage as a determining factor in match results. Notably, in basketball and volleyball tournaments, P_d is always zero since draws are not possible, while in handball, the probability of draws is significantly lower than in soccer.

It is important to note that our simulations are more specific than simulations of a perfectly balanced tournament. Instead, they are tailored to the tournament under consideration, replicating not only characteristics such as the number of matches and teams but also the schedule and home-court advantage. As a result, the confidence envelope (green shaded area in Figure 1) will also reflect all these characteristics, and the comparison between the real tournament and its respective simulations is valid. Simply put, the shaded area is created to answer: given the observed schedule, how much could the variance fluctuate if the results were only defined by P_h, P_d, and P_a?

Our framework only works if the comparison to the expected behaviour is fair, that is, if the same teams that played on a matchday also play during it in our simulation. In a perfect double round-robin where all teams play every round, we could directly compare every round to a simple simulation. However, as will be clearer in Section 5.3, our extensive database contains tournaments with peculiar order of matches. In these cases, forcing the schedule to be the same is crucial to guarantee the validity of our results. Nevertheless, since the simulations obey the same straightforward rules, they serve as a unified reference point, which is only slightly adjusted by the previously mentioned tournament characteristics. Therefore, comparisons across different leagues and sports should also be valid.

5.2 Observed Turning Point τ

Our metrics are based on a recurring pattern we observed in the majority of tournaments within our dataset, and tournaments simulated under the Bradley-Terry model (Bradley and Terry, 1952). Specifically, in competitions featuring uneven skill distributions, the dispersion of the standings closely mirrors the expected dispersion of perfectly balanced ones *only* for a specific number of rounds, denoted as the *observed turning point* τ. It represents the duration that a tournament can be considered perfectly balanced and can be used as a measure of *competitive balance*. Larger τ values indicate tournaments perceived as more balanced by the public. If a tournament has no τ, it is regarded as perfectly balanced throughout its entire duration, and the coefficient value is equivalent to the duration of the tournament.

This coefficient measures what is *seen* by the public as the tournament progresses, being affected by another factor in addition to the skill distribution of the teams: the order of matches. It loosely represents how balanced the competition would appear to the audience as the tournament occurred. For instance, if a team won all its matches during the tournament's first half, the competition would not seem balanced, and our coefficient would reflect this. As a result, we will use *perceived competitive balance* to denote this idea, and in this work, this term should not be confused with its more common use in the literature: the (subjective) perception of competitive balance by fans in sports leagues or matches (Nalbantis et al., 2017; Pawlowski and Budzinski, 2012).

The procedure to estimate τ for tournament $\mathcal{C}$ is as follows. We first run K simulations of the null model following the same schedule S of the original tournament $\mathcal{C}$. Recall that the probabilities P_h, P_d, and P_a are estimated respectively by the empirical relative frequencies of the home team winning, drawing, or losing a game in $\mathcal{C}$. For simulation (k), at each round t, we record the generated points distribution $\mathcal{X}_t^{(k)}$ and its imbalance $\mathcal{V}_t^{(k)}$ using *any competitive balance metric*. Each simulation generates a stochastic temporal curve $\mathcal{V}_1^{(k)}, \cdots, \mathcal{V}_T^{(k)}$ reflecting the natural variability one can expect in the points distribution $\mathcal{X}_t^{(k)}$ when there is no difference in team strength. With these quantities, we estimate the *expected imbalance of the point distribution* at round t for a perfectly balanced tournament by calculating the average imbalance μ_t:

$$\mu_t = \frac{1}{K}\sum_{k=1}^{K}\mathcal{V}_t^{(k)}. \tag{1}$$

Subsequently, we compare the imbalance of the real tournament $\mathcal{C}$ with those generated by the null model. For each round t, we employ a significance level α to determine the imbalance values likely to be observed in a perfectly balanced tournament. More specifically, we use the $(1-\alpha)$-quantile q_t of $\mathcal{V}_t^{(1)}, \cdots, \mathcal{V}_t^{(K)}$ to estimate the threshold below which the expected fluctuation should be if the assumption that all teams are equally skilled holds. Formally, it is the smallest value greater than $100(1-\alpha)\%$ of all simulated imbalances at round t:

$$q_t = \inf\{q : (1-\alpha) \leq \hat{\mathbb{F}}_t(q)\}, \tag{2}$$

where $\hat{\mathbb{F}}_t$ is the empirical cumulative function:

$$\hat{\mathbb{F}}_t(q) = \frac{1}{K}\sum_{k=1}^{K} \mathcal{I}_{\mathcal{V}_t^{(k)} \leq\ q_t}$$

with $\mathcal{I}$ being an indicator variable: 1 if true and 0 otherwise.

This *quantile envelope* represents what one can expect for the extreme deviation of the observed imbalance $\mathcal{V}_t$ from its expected behavior μ_t *if the real tournament $\mathcal{C}$ was perfectly balanced*. It allows us to measure *enduring detachment* between these quantities since $\mathcal{V}_t$ lying for a long time above q_t is highly unlikely if all teams have the same skill. We define the *observed turning point* τ as the first moment the observed imbalance $\mathcal{V}_t$ becomes at least as large as q_t and stays as such until the end of the tournament:

$$\tau = \arg\max_t\ \mathcal{V}_t < q_t. \tag{3}$$

If a tournament ends at time T, we trivially normalize this measure to provide a more equitable comparison across different leagues:

$$\tau_{\%} = \frac{\tau}{T}. \tag{4}$$

It is important to emphasize that $\mathcal{V}_t$ may exceed q_t at a given round t and, after some rounds, return to being smaller than it. In perfectly balanced tournaments, this behavior can be explained by the well-studied regression toward the mean phenomenon (Barnett et al., 2005), which has been used, for example, to debunk the notion of team momentum often attributed to timeouts in basketball games (Assis et al., 2021). This principle states that if an observation of a random variable is extreme, subsequent observations are likely to be closer to the mean.

We begin by validating our metric through simulated tournaments based on the Bradley-Terry model Bradley and Terry (1952), with results presented in Figure 2. The figure tracks the evolution of competitive imbalance ($\mathcal{V}_t$) across four synthetic tournaments, each consisting of 10 teams playing five complete double round-robin schedules. Team skill distributions are encoded in each subplot title using the notation $[p_1] \times n_1 + [p_2] \times n_2 + \cdots$, where n_i denotes the number of teams with skill level p_i. For example, the most imbalanced tournament (bottom-right subplot) comprises: two weak teams ($p = 1/9$), six average teams ($p = 1$), and two strong teams ($p = 9$). Under the Bradley-Terry model, the probability of team i defeating team j depends exclusively on their skill levels p_i and p_j, calculated as $p_i/(p_i + p_j)$. In our most imbalanced tournament configuration, this results in a weak team ($p = 1/9$) having a 10% probability of defeating a strong team ($p = 9$), since $(1/9)/(1/9 + 9) = 0.10$. For each simulated tournament, we generated match outcomes by sampling from these probabilities. Note that the Bradley-Terry framework assumes binary outcomes (win/lose) and excludes the possibility of draws.

Our use of synthetic Bradley-Terry tournaments serves two key purposes. First, these simulations provide a controlled validation framework for our proposed metrics. By precisely defining the underlying skill distributions beforehand, we can verify whether our competitive balance measures accurately reflect the predetermined imbalance levels in each simulated tournament. Second, the Bradley-Terry model offers an ideal baseline for initial validation as its simplicity - relying solely on team skill to generate match outcomes without confounding factors - creates a clear, interpretable testing environment for our metrics.

In Figure 2, the observed imbalances $\mathcal{V}_t$, denoted by red crosses, are computed using the *variance* of the standings. In contrast, the null model's expected imbalances (μ_t), also computed using the *variance*, are represented by green circles, while the quantile envelope q_t, under a significance level of $\alpha = 5\%$, is shown as the shaded green region. We note that we only add the expected imbalaces (μ_t) for visual reference, since only the (shaded) quantile region is required to estimate the coefficient.

Observe that when all teams are equally skilled (top-left), the imbalance $\mathcal{V}_t$ remains within the quantile envelope q_t throughout the tournament. Conversely, in tournaments with increasing skill discrepancies among teams, the deviation from the envelope accelerates. A vertical dashed line marks our *observed turning point* measure τ, which is the point where the observed imbalance $\mathcal{V}_t$ permanently separates from the quantile envelope q_t. Although these simulations are influenced by some degree of randomness, repeated runs consistently reveal key trends: (i) the vast majority of balanced tournaments remain within the envelope, (ii) for moderate skill discrepancy levels, the actual imbalance eventually diverges from the envelope, and (iii) as the skill gap widens, this divergence initiates earlier, resulting in a smaller τ.

This pattern is also observed across the four real tournaments, each representing a different sport, shown in Figure 1. Initially, in all tournaments, there is no distinction between the observed imbalance, $\mathcal{V}_t$, and the expected imbalance, μ_t. However, given time, the uneven skill distribution causes the observed

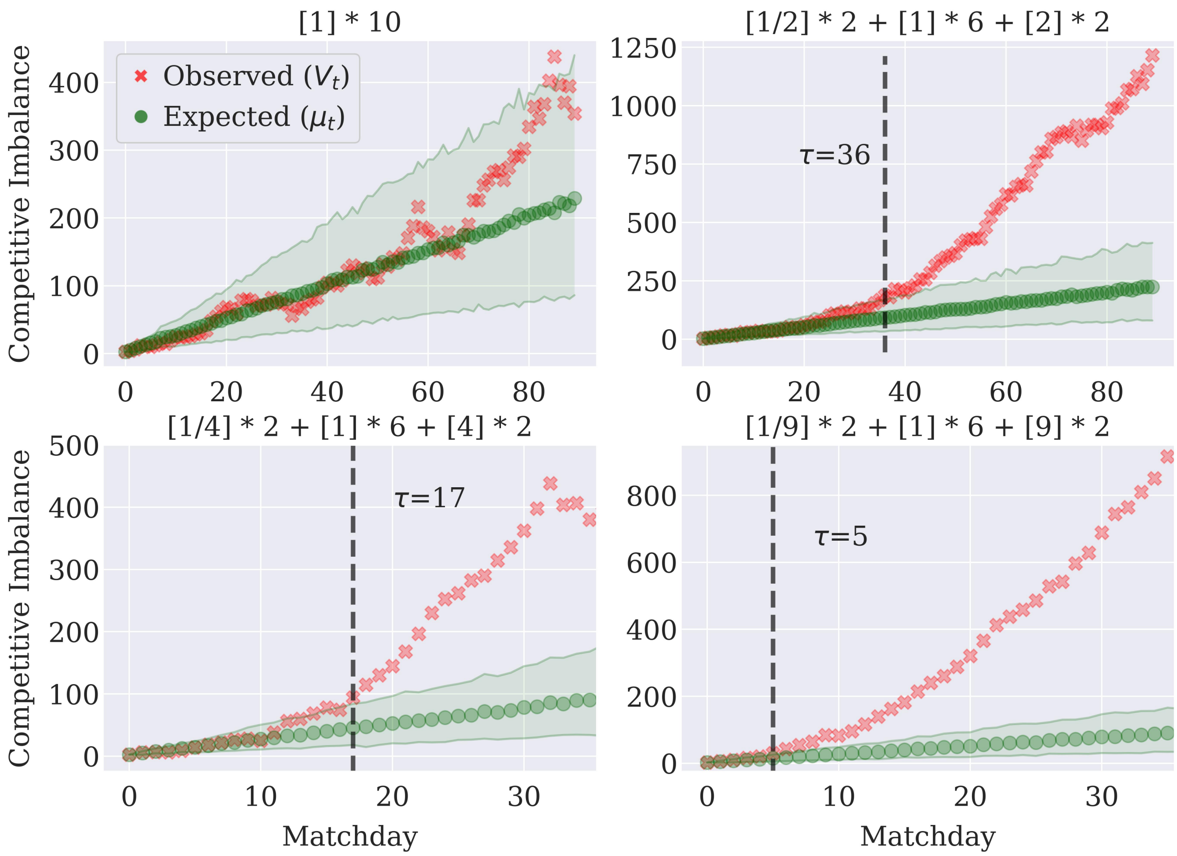

Fig. 2: Competitive imbalance $\mathcal{V}_t$ divergence in Bradley-Terry simulations.

imbalance $\mathcal{V}_t$ to diverge from its expected behavior μ_t under the assumption that the real tournament is perfectly balanced. Note that τ is a measure of *permanent divergence* from the perfectly balanced behavior. Therefore, the observed imbalance $\mathcal{V}_t$ (red crosses) *must* stay above the quantile envelope q_t (shaded region) for all dates after τ. In particular, the red crosses first escaped the green envelope on dates 5 for *Russian Superleague* and 39 for *La Liga*, but the *turning points* τ only happened slightly later (dates 8 and 42) since they immediately dropped inside again.

Figure 3 shows the boxplots (in blue) for the *normalized observed turning points* $\tau_{\%}$ for all seasons in our dataset grouped by sport. Soccer leagues are the most balanced ones: the ones that present, on average, the highest *turning point* values. The mean *turning point* for soccer is $\overline{\tau_{\%}} = 45.7\%$, indicating that the tournaments tend to exhibit points distributions akin to perfectly balanced tournaments for nearly half of the tournament duration. For the other sports, the *turning points* come, on average, significantly earlier. For basketball, $\overline{\tau_{\%}} = 33.0\%$, for volleyball, $\overline{\tau_{\%}} = 29.2\%$, and handball, $\overline{\tau_{\%}} = 27.4\%$. This discrepancy can be attributed to the higher likelihood of underdogs winning in soccer, primarily due to the minimal number of scores (goals) required for a victory. This phenomenon results in more frequent upsets, leading to teams

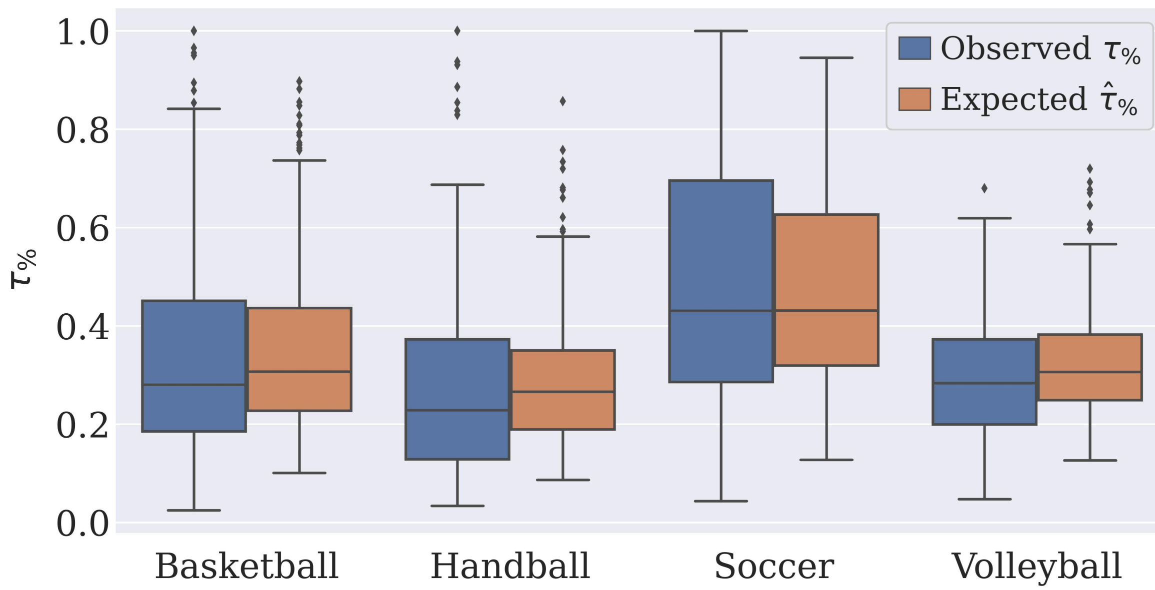

Fig. 3: General comparison between the observed $\tau_{\%}$ and expected $\hat{\tau}_{\%}$ turning points for all sports.

being closer in the standings and consequently, a higher perceived competitive balance. Moreover, despite the substantial differences in average values, all sports exhibit tournaments with notably high *turning points*. Even in volleyball, certain leagues experience detachment as late as 60% of their duration, with two seasons never reaching a *turning point* at all.

5.3 Expected Turning Point $\hat{\tau}$

One limitation of the previous *observed turning point* τ is its dependency on the game order (or schedule). This is particularly notable for the 2013/2014 season of the Brazilian volleyball tournament *Superliga*. The tournament's schedule was unique because *Sada Cruzeiro*, which eventually won both the regular season and playoffs, had qualified for the *World Championship*, a title they would also claim. Typically, such a qualification would not pose a problem, but the *World Championship* occurred shortly after the Brazilian season commenced, leading to rescheduled matches for *Sada Cruzeiro* played earlier than usual. In summary, while most teams in the tournament had played only two matches, the top-ranked team in the world at the time played five. Based on the estimated probabilities of home ($P_h = 53.8\%$) and away ($P_a = 46.2\%$) wins for this season, the likelihood of *Sada Cruzeiro* winning all five matches was only 3.9% if all teams were equally matched, yet they won them. Consequently, the observed imbalance $\mathcal{V}_t$ diverged q_t after *Sada Cruzeiro*'s third match and never reverted to the quantile envelope again, resulting in $\tau = 4$ for this tournament.

While extreme situations such as this are infrequent, they undoubtedly illustrate the influence scheduling has on the *observed turning point* τ which only estimates the *perceived* competitive balance for the *real instance* of the tournament. If, on the other hand, we intend to measure the *actual competitive balance* solely by analyzing the tournament data, it is crucial to mitigate the effect of match order as much as possible. To address this, we introduce a method for computing the expected value $\mathbb{E}[\tau]$ when the sequence of matches is unknown, ensuring that the metric remains invariant to the tournament schedule.

Let M_0 be the list of observed match results in the real tournament $\mathcal{C}$ ordered by its schedule. It can be seen as a direct result of the underlying skill distribution within the tournament. Now, let $\mathfrak{S}_{M_0}$ be the set of all possible match orders of M_0, that is, all of its permutations. Then, each permutation $M \in \mathfrak{S}_{M_0}$ has a different temporal progression for its point distribution $\mathcal{X}_t$ throughout the tournament, and, consequently, a different progression for the observed imbalance $\mathcal{V}_t$. In other words, the competitive imbalance $\mathcal{V}_t$ at each time t is directly reliant on which permutation M is being considered. Naturally, as t approximates T, the imbalances become similar independently of the match order, with the final $\mathcal{V}_T$ always being the same for every possible ordering.

Aware of this dependence, we view the *turning point* for a tournament $\mathcal{C}$ with observed results M_0 as a random variable $\mathcal{T}$ depending on the schedule $M \in \mathfrak{S}_{M_0}$ and the underlying skill distribution, represented by the observed match results. To calculate τ for any permutation $M \in \mathfrak{S}_{M_0}$, we extend the procedure described in the previous section as a function $\tau : \mathfrak{S}_{M_0} \to \mathbb{N}$ that takes as input a permutation of the schedule (along with its outcomes) M, and returns its *turning point* τ. Naturally, this function outputs the *observed turning point* of the original tournament by simply taking $M = M_0$ as input. More importantly, it allows us to effectively dissipate the impact of scheduling on our estimation by defining the actual competitive balance of a tournament as the **expected** *turning point* value given its observed match results M_0 while disregarding their order:

$$\mathbb{E}[\tau] \coloneqq \mathbb{E}\Big[\mathcal{T} \;\Big|\; M_0\Big] = \sum_{M \in \mathfrak{S}_{M_0}} \tau(M)\, \mathbb{P}(M). \tag{5}$$

Unfortunately, this expression is insufficient to calculate $\mathbb{E}[\tau]$ due to the probability $\mathbb{P}(M)$ being complicated to ascertain. Only permutations similar to a round-robin would ever be considered for real tournaments, and those adhering to fairness considerations would have a higher likelihood of being chosen. Since it is impractical to account for these subtleties, we simply assume that all non-round-robin schedules can never happen and that all round-robin schedules are equally likely. Mathematically,

$$\mathbb{E}[\tau] \approx \sum_{M \in \mathfrak{S}_{M_0}^{RR}} \tau(M)\ \mathbb{P}(M) \approx \frac{1}{\left|\mathfrak{S}_{M_0}^{RR}\right|} \sum_{M \in \mathfrak{S}_{M_0}^{RR}} \tau(M) =: \hat{\tau}, \tag{6}$$

where $\mathfrak{S}_{M_0}^{RR} \subset \mathfrak{S}_{M_0}$ represents all possible schedule permutations following a round-robin structure, and $\hat{\tau}$ is how we will denote our estimator henceforth.

Through these simplifications, we avoid all orderings that would never appear in a real setting, but we also ignore slight modifications that could occur under extreme circumstances, such as the COVID-19 pandemic. All in all, this process can be seen as a *format standardization* where we convert all tournaments to well-ordered round robins. Back to our previous example, this conversion implies that in none of our permutations $M \in \mathfrak{S}_{M_0}^{RR}$ *Sada Cruzeiro* could have played more games in the beginning than its competitors; all teams would have had about one game per round. As such, the *expected turning point* $\hat{\tau}$ could never be as small as τ was.

We detail the algorithmic procedure for calculating (6) in Algorithm 1. Since generating all $\left|\mathfrak{S}_{M_0}^{RR}\right|$ permutations for tournaments with as few as 15 teams is unfeasible (Rasmussen and Trick, 2008), we instead generate a sample of K round-robin-based permutations to represent the set $\mathfrak{S}_{M_0}^{RR}$. For each sample, we calculate the *turning point* τ using the function $\tau : \mathfrak{S}_{M_0} \to \mathbb{N}$ defined earlier, giving us a list of values from which to estimate the *expected turning point* $\hat{\tau}$ and an *expected confidence interval* $\mathrm{CI}_{\hat{\tau}}$. According to Equation 6, the *expected turning point* is simply the average *turning point* over the K permutations. It measures the coefficient's expected value for *standard* round-robin schedules if all match outcomes remain unchanged. Due to the well-defined and constant schedule format, the non-uniformity of teams' strengths is the most relevant factor in its estimation. As for the *confidence interval*, we define it as the $(1 - \alpha)$ probability mass around the *expected turning point* $\hat{\tau}$. For a significance level of $\alpha = 5\%$, this is the interval between the 0.025 and the 0.975 empirical quantiles. It represents a range inside which a tournament's *turning point* should be given the underlying teams' skill distribution. Any instance whose τ is outside of it must have had a schedule that hindered the natural point fluctuation.

Our methodology for creating a round-robin permutation $M \in \mathfrak{S}_{M_0}^{RR}$ consists of concatenating as many random, mirrored double round-robin schedules as necessary to complete the set of observed match outcomes M_0. Initially, we create a single round-robin schedule with the same number of teams as the real tournament $\mathcal{C}$ using the *Circle Method* (de Werra, 1981), a popular method for generating initial solutions for round-robin sport scheduling problems (Rasmussen and Trick, 2008; Goossens et al., 2012). It takes an

Algorithm 1: Expected Turning Point Algorithm

```
Input   : Schedule and matches outcomes: M_0
Output  : τ̂ and CI_τ̂

The M_0 input is a list of matchdays composed of matches/outcomes: (home team, away team,
  outcome).

Function Matching(M_0):
    τ_list ← []
    for k ∈ {1, ···, K} do
        M ← RandomizeRR(M_0)
        turning_point ← τ(M)
        τ_list.append(turning_point)
    end
    τ̂ ← Mean(τ_list)
    CI_τ̂ ← Quantile(τ_list, q = [0.025, 0.975])
    return τ̂, CI_τ̂
```

ordered list of n teams as input and generates a full schedule where each team plays in every round and against each other once. Matches are represented by ordered tuples where the first team plays at home.

Since the *Circle Method* is deterministic - it always generates the same schedule if the order of teams in the list remains unchanged - the initial step to produce a permutation $M \in \mathfrak{S}^{RR}_{M_0}$ involves shuffling the team list before feeding it as input to the *Circle Method*. The subsequent step is to designate which team plays at home in each match. Here, we simply iterate over the matches generated by the *Circle Method* and alter the tuple order with a 50% probability. Moving on, the third step entails iterating over each round of the schedule and shuffling the match order within each round. This step is crucial as some tournaments feature rounds where matches are spread across multiple matchdays, which might potentially occur after the start of the subsequent round. The fourth and final step to complete a full first phase of the round-robin permutation involves shuffling the $n-1$ rounds of the schedule. Algorithm 2 describes these steps for a single round-robin tournament, which can then be mirrored to create the second phase, completing the double round-robin tournament.

For general tournaments, we generate additional round-robin schedules, mirror and concatenate them with the previous ones until all actual matches are accommodated. These general tournaments introduce a new randomization step. In a simple double schedule, team i faces team j once at home and once away. In a multi-schedule, both cases can happen more than once; thus, we also randomize them when required.

We emphasize that we only alter the order of the matches, maintaining their results as they were. As a consequence, $\hat{\tau}$ is only an approximation of what would have been expected had the tournaments been

Algorithm 2: Round Robin Randomization

```
Input   : Real schedule and matches outcomes: M_0
Output  : Randomized round-robin schedule: M

RemoveUnseenGames removes from M all games that were not played in M_0.
AddRealOutcomes adds to M the outcomes of all matches played in M_0.

Function RandomizeRR(M_0):
    teams ← GetTeams(M_0)
    Shuffle(teams)
    M ← CircleMethod(teams)
    for matchday ∈ M do
        for match ∈ matchday do
            Shuffle(match)
        end
        Shuffle(matchday)
    end
    Shuffle(M)
    RemoveUnseenGames(M, M_0)
    AddRealOutcomes(M, M_0)
    return M
```

played again in a different order. In particular, this procedure disregards that the order of matches can affect their outcome. Although this may be a strong assumption for high-stakes matches, its overall impact on a season-wide coefficient should be considerably smaller than the scheduling had on the *observed turning point*. The reason being that the effect of disregarding match order would be spread over the permutations rounds, potentially canceling out. However, the impact of a peculiar schedule could easily be concentrated in a specific portion of the tournament, akin to our Brazilian *Superliga* example. As such, the *expected turning point* $\hat{\tau}$ provides a better estimation for the **actual** competitive balance of tournaments. It also provides an *expected confidence interval* inside of which the *observed turning point* should reside, allowing outlier cases such as the one used to motivate its definition (Brazilian *Superliga*) to be found.

Figure 3 illustrates the behavior of the *expected turning point* $\hat{\tau}_{\%}$ contrasted to the *observed turning point* $\tau_{\%}$. Note that the *expected* $\hat{\tau}_{\%}$ is considerably more stable than the *observed* one, as illustrated by its concentrated quartiles and whiskers. We verified that there was only a single league (2020/21 Portugal's *Andebol 1*) with a $\hat{\tau}_{\%}$ below 10%, whereas the same happened in 94 seasons for $\tau_{\%}$: a strong piece of evidence for its reliability against peculiar schedules such as the one for Brazil's *Superliga*. Also, the *observed* values are lower in general, possibly indicating that tournament organizers are, deliberately or not, using schedules that slightly favor a rapid distinction between good and bad teams.

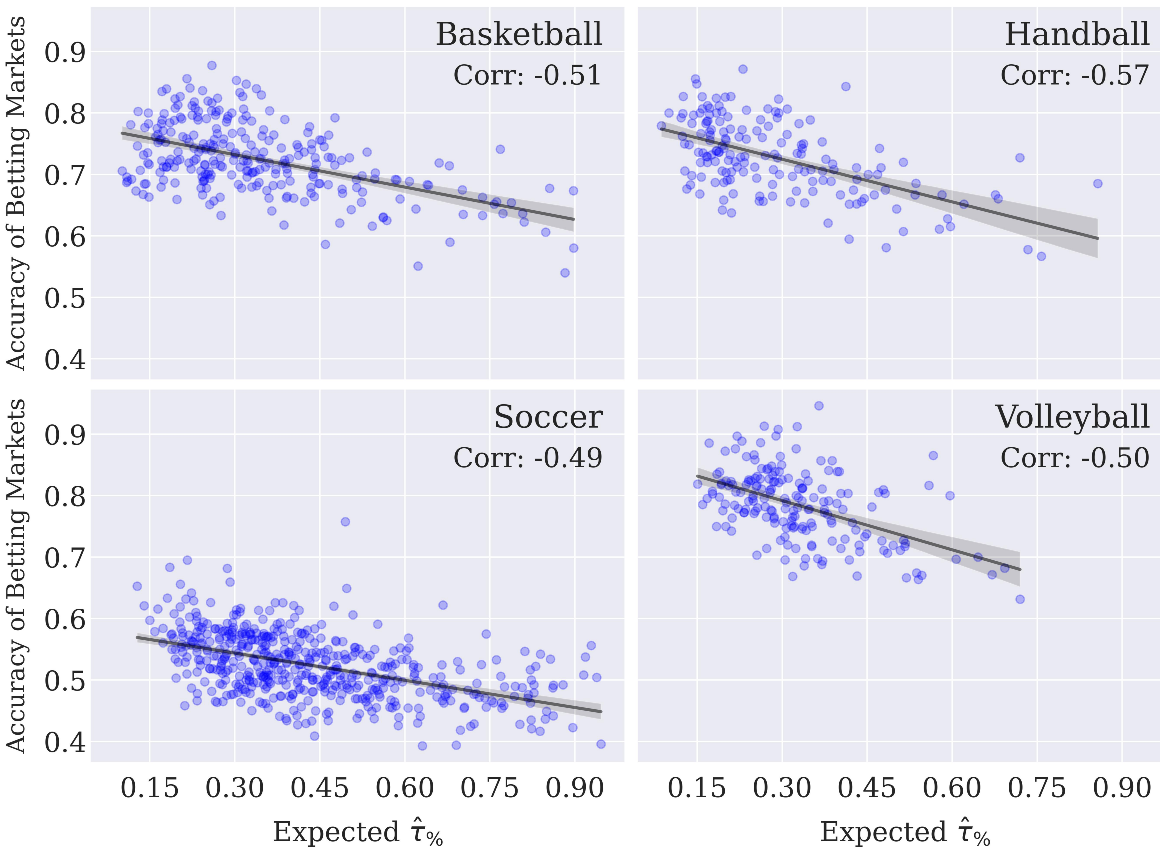

Fig. 4: Comparison between the betting market's prediction accuracy and competitive balance for all seasons containing betting data. Competitive balance is measured by our normalized expected turning point $\hat{\tau}_{\%}$. The black line represents the linear fit for the data and the number below each sport name is the corresponding Pearson Correlation coefficient.

We validate our proposed *expected turning point* $\hat{\tau}$ by evaluating its correlation with the predictability of the tournaments in our dataset. As per the *uncertainty of outcome* theory, tournaments that are more balanced tend to exhibit greater difficulty in predicting match outcomes (Rottenberg, 1956; Forrest and Simmons, 2002; Szymanski, 2003; Owen, 2014). We measure predictability using the accuracy of the betting market for the tournament, that is, what fraction of the season matches the team with the smallest odds won. As shown in Figure 4, there is a notable negative correlation across all four sports between the *expected turning point* $\hat{\tau}$ and the accuracy of the betting market: tournaments with lower $\hat{\tau}$ (indicating more imbalanced tournaments) tend to have more accurate market predictions. A simple linear model confirms this trend, demonstrating a significant drop in accuracy of at least 10% for all sports as $\hat{\tau}$ increases by about 0.3.

6 Anomalous Seasons

Figure 5 shows a comparison between the normalized observed $\tau_{\%}$ and the expected $\hat{\tau}_{\%}$ for each tournament in our dataset. The red circles represent the tournament whose observed $\tau_{\%}$ was below the *expected confidence interval* $\mathrm{CI}_{\hat{\tau}}$, whereas for the blue squares, $\tau_{\%}$ was above it. We also highlight the 2013/2014 volleyball season for the Brazilian *Superliga* as a red triangle since it was our motivation for defining $\hat{\tau}$ in *Expected Turning Point* $\hat{\tau}$. The number below each sport name is the Pearson Correlation coefficient between $\tau_{\%}$ and $\hat{\tau}_{\%}$.

Observe that there is a strong correlation between the metrics, suggesting that the *observed turning point* still is a reasonable measure for competitive balance. In part, this correlation is explained by all seasons already following a well structured round-robin schedule, so the expected and observed coefficients should be similar. For volleyball, the lower correlation might be a consequence of the lack of tournaments with significantly high turning points. More importantly, Figure 5 highlights all tournaments outside the *expected confidence interval* (red circles and blue squares), that is, those whose schedules resulted in observed *turning points* way higher (or lower) than what is expected from the corresponding match outcomes (i.e. $\hat{\tau}_{\%}$). In total, only 3.96% of all seasons were below the $\mathrm{CI}_{\hat{\tau}}$, while 3.18% were above. We conjecture that an *observed turning point* below the $\mathrm{CI}_{\hat{\tau}}$ can be a result of easy schedules for good teams at the beginning, or a large discrepancy in teams' strengths; whereas an *observed turning point* being above the $\mathrm{CI}_{\hat{\tau}}$ might be from some matches being played earlier than they should, breaking the well-behaved round-robin structure.

7 Most Imbalanced Leagues

When it comes to competitive balance, a natural question is: what are the most imbalanced leagues? Table 1 displays the three most imbalanced leagues for each sport, along with their normalized *turning points* averaged over all available seasons. We also list three other popular soccer leagues along with their imbalanced ranks. Furthermore, we show the *Theil-Sen slope* to characterize the general temporal tendency of $\hat{\tau}_{\%}$ for all the leagues across their available seasons. The Theil-Sen slope is the median slope of all lines through pairs of points, which mitigates the effect of outliers in our small (at most 11 seasons) sample.

For our entire dataset, the slopes reveal a trend toward decreasing balance in soccer tournaments, with an average slope of -0.38%, and a small increasing balance in all other sports: 0.12% for basketball, 0.31%

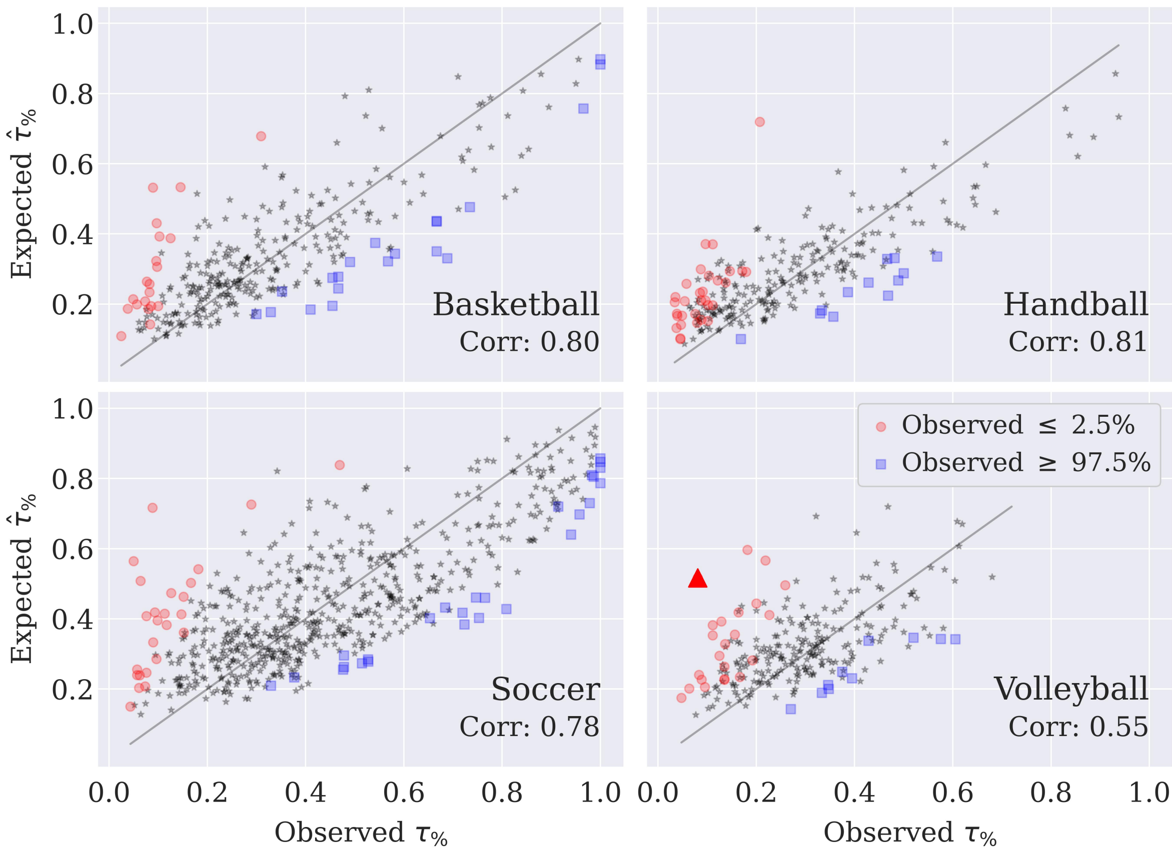

Fig. 5: Comparison between the observed $\tau_{\%}$ and expected $\hat{\tau}_{\%}$ turning points for all studied seasons. The red circles represent the tournament whose observed $\tau_{\%}$ was below the *expected confidence interval* $\mathrm{CI}_{\hat{\tau}}$, whereas for the blue squares, $\tau_{\%}$ was above it. The number below each sport name is the Pearson Correlation. We highlight the 2013/2014 volleyball season for Brazilian *Superliga* as a red triangle.

for handball, and 0.44% for volleyball. This outcome is unsurprising since soccer tournaments are already generally perceived as more balanced when compared to the others, making a general increase in balance less likely. This outcome is unsurprising since soccer tournaments are already perceived as more balanced than others, making further increases in balance less likely.

Notably, there are some of the most popular leagues in each sport among the most imbalanced leagues. For soccer, *Big Five* leagues - England, Italy, Germany, Spain and France - are often studied given the revenue disparity between the best and worst teams (Wonga, 2023; Plumley et al., 2018). As with most literature work, the French league is the most balanced of all *Big Five* tournaments, while the English, Spanish, and Italian leagues are usually the least competitive (Avila-Cano and Triguero-Ruiz, 2023; Kringstad, 2021; Michie and Oughton, 2004). Additionally, throughout the 11 seasons we consider, our results agree with the temporal trend in Kringstad (2021), since only *La Liga* did not show a decreasing trend in competitive balance. The order in Table 1 is also in line with the (average) concentration results over 22 seasons for these five main European leagues shown in Ramchandani et al. (2018). Finally, notice how τ (first column) is usually close to $\hat{\tau}$ (second). The one clear discrepancy is the *La Liga*, which is perceived as more competitive than its teams' skill distribution would make people believe.

For basketball, the evolution of balance in the NBA is notable given both its popularity and its compensation mechanisms (del Corral et al., 2016). In Bowman et al. (2013), the authors analyzed such progression from the fans' perspective and found that balance seems to have increased through time. Although the temporal progression trend for NBA is small, our work indicates an increase in competitive balance as well. Unfortunately, it is hard to find a comparison between the NBA and other basketball leagues. An exception is the work of García Unanue et al. (2014), which revealed that the NBA is as imbalanced as European leagues when conferences were analyzed separately. Similarly, the analyses presented in Jungić et al. (2015) revealed that the NBA is less competitive than Spain's ACB, which also is in line with our results.

For handball, Hantau et al. (2014) presented concerns regarding the imbalance in Romania's *Women Liga Nationala* and in Germany's *Bundesliga*, two leagues that appear in our list. Similarly, Haugen and Guvåg (2018) showed that both *Bundesliga* and *Liga Nationala* have low competitive balance. The three *Big Five* soccer leagues they analyzed (Spain, Germany, and France) also follow the same order as ours.

For volleyball, literature concerning competitive balance is not as developed as it is for the other sports. Nonetheless, qualitatively, while none of the most famous tournaments are present in Table 1, the Russian Superleague is the 4-th least competitive, and Italy's *Superlega* is the 8-th. We note, however, that *Division A (Belarus)* showed a tremendous positive trend in its competitive balance during the studied period. This tendency is significant, as a consistent linear increase of 3% over ten seasons represents a substantial shift (+30% in total) in the perceived competitive balance.

8 Other Metrics of Imbalance

8.1 Coefficients Comparison

The main contribution of this paper is our $\hat{\tau}$ as a new, probabilistic metric for measuring competitive balance, which accounts for the fluctuations inherent to sports leagues. As such, it is expected that our metric have different results from the usual, purely deterministic competitive balance metrics in the literature. In the following, we compare our $\hat{\tau}$ with five of the most commonly used metrics for quantifying the competitive

Tab. 1: Average normalized turning point values for the three least competitive leagues of each sport.We also included the missing Big Five soccer leagues with their respective rankings in parenthesis. Competitive Balance is measured by the expected turning point $\hat{\tau}$ (second column). The *Theil Slope* column represents the trend of increase/decrease during our 11-season analysis.

Soccer	$\tau_{\%}$	$\hat{\tau}_{\%}$	Theil Slope ($\hat{\tau}_{\%}$)
Serie A (Italy)	24.0%	23.3%	-1.3%
Premiere League (England)	23.5%	23.7%	-0.4%
La Liga (Spain)	34.8%	24.8%	1.1%
Bundesliga (13th, Germany)	31.2%	32.9%	-1.0%
Ligue-1 (16th, France)	37.1%	34.8%	-1.1%
Serie A (45th, Brazil)	50.7%	48.8%	-2.5%
Basketball			
NBA (USA)	11.8%	13.1%	0.5%
B-League (Japan)	11.7%	17.0%	-0.3%
BBL (Germany)	23.6%	21.2%	-0.5%
Volleyball			
Superleague (Ukraine)	17.2%	20.9%	0.3%
Division A (Belarus)	23.7%	22.6%	3.1%
Extraliga (Hungary)	24.6%	23.8%	-0.2%
Handball			
Andebol 1 (Portugal)	14.2%	15.6%	0.2%
Bundesliga (Germany)	13.8%	16.1%	0.1%
Liga Nationala (Romania)	12.2%	19.3%	-0.1%

balance in a tournament. For ease of notation, let s_i denote the final points in the standings of the i-th ranked team, and let S represent the sum of the points of all n teams at the end of the tournament. In our previous notation, $s_i \equiv X_{i,T}$ and $S \equiv \sum_i X_{i,T}$. To simplify, we apply each metric to the final standings following the point system used for our metric: 3 points for a win, 1 for a tie, and 0 for a loss. Given the number of metrics considered, it should be sufficient to highlight the overall tournament behavior.

nHHI: The normalized Herfindahl–Hirschman Index (HHI), in the context of sports leagues, measures the proportion of points (or winning percentage) teams contain compared to the total: $\text{HHI} = \sum_i \left(\frac{s_i}{S}\right)^2$. For a tournament with N teams, it can be normalized as $\text{nHHI} = \frac{\text{HHI}-\frac{1}{n}}{1-\frac{1}{n}}$, where $\frac{1}{n}$ is the HHI for a perfect balanced tournament (Ramchandani, 2012).

Improved nHHI: This coefficient improves the previous **nHHI** normalization by considering an appropriate upper bound in the context of sports leagues: $\text{nHHI} = \frac{\text{HHI}-\frac{1}{n}}{\text{UP}-\frac{1}{n}}$ (Owen et al., 2007). For more details, see Appendix B.

Tab. 2: Correlation between our expected turning point $\hat{\tau}$ and other popular competitive balance metrics.

	nHHI	Improved nHHI	HICB	Gini	nGini	IQR	25% NCR
Basketball	-0.20	-0.49	-0.37	-0.46	-0.54	-0.63	-0.25
Handball	-0.14	-0.60	-0.51	-0.54	-0.61	-0.58	-0.50
Soccer	-0.36	-0.64	-0.59	-0.62	-0.66	-0.60	-0.45
Volleyball	-0.18	-0.56	-0.49	-0.54	-0.59	-0.59	-0.05

HICB: Another option related to the HHI is the Herfindahl Index of Competitive Balance, given by $\text{HICB} = \frac{\text{HHI}}{\frac{1}{n}}$. It fixes the HHI's sensitivy to the number of teams, bringing it back to a variance measure (Michie and Oughton, 2004; Ramchandani et al., 2018; Doria and Nalebuff, 2021).

Gini: The Gini coefficient is a common way to measure equality in normalized distributions. It can be geometrically defined using the Lorentz curve, but if $\overline{S}$ is the average points in the standings, it can also be algebraically calculated by $\text{Gini} = \frac{\sum_i \sum_j |s_i - s_j|}{2\overline{S}n^2}$.

nGini: Similar to **Improved nHHI**, this metric improves the **Gini** index by normalizing it with an appropriate upper bound (Utt and Fort, 2002). For more details, see Appendix B.

IQR: The interquartile range is a traditional concentration metric that measures the difference between both quartiles, that is, between the 0.75-quantile and 0.25-quantile (Ramchandani, 2012).

25% NCR: The 25% concentration ratio (25% CR) measures the points accrued by the top 25% teams when compared to the total number of points: $25\%\text{CR} = \sum_i^{\lfloor 0.25n \rfloor} \frac{s_i}{S}$. It can be normalized by the number of teams in the top 25%, that is, $25\%\text{NCR} = \frac{25\%\text{CR}}{\lfloor 0.25n \rfloor}$ (Ramchandani, 2012).

Table 2 presents the Pearson correlation coefficient between $\hat{\tau}$ and these five metrics. The results demonstrate a generally significant correlation between them, except for the nHHI and NCR metrics in some sports. Overall, the correlations suggest a significant association between our $\hat{\tau}$ coefficient and these concentration metrics. The fact that the correlations are not very strong indicates that our proposed $\hat{\tau}$ metric offers a complementary approach to characterizing competitive balance in sports tournaments. This is a direct result of our model's capability to consider the random fluctuations in tournaments, which may go unnoticed by other measurements that solely rely on the final standings.

8.2 Framework Invariance

All results in this paper utilized the *variance* to quantify the imbalance of the points distribution. In this section, we demonstrate that our proposed τ metrics are indeed invariant to the metric used to calculate

Tab. 3: Correlation between the observed turning point τ calculated using variance compared to other popular competitive balance metrics.

	nHHI	Improved nHHI	HICB	Gini	nGini	IQR	25% NCR
Basketball	0.99	0.99	0.99	0.96	0.96	0.59	0.67
Handball	0.99	0.98	0.99	0.98	0.97	0.54	0.74
Soccer	0.98	0.98	0.98	0.95	0.95	0.48	0.81
Volleyball	0.99	0.96	0.99	0.95	0.92	0.45	0.52

the tournament imbalance at each tournament round. To support this statement, in Table 3 we present the correlation between the *observed turning point* calculated using the *variance* and the same coefficient calculated using the competitive imbalance metrics described in *Coefficients Comparison*. There is an almost perfect correlation with nHHI, HICB and Gini, as well as a moderately high correlation with the 25% NCR and IQR. These high correlations strongly suggest that our result would also hold if other metrics were used to quantify the imbalance in the points distribution at each tournament round. Additionally, Table 3 also indicates that using the *variance* is similar to any other *global* metric of dispersion, that is, those that consider the entire standings. Both IQR and NCR only account for a portion of it, respectively, teams between the quartiles and teams at the top of the standings, so the correlation is not as strong.

9 Conclusion

We introduce a new framework for studying competitive balance that focuses on the progression of a competition rather than relying solely on its final standings. Central to this approach is the *observed turning point* (τ), which quantifies the perceived competitive balance of a tournament by measuring the number of rounds in which it appears indistinguishable from a perfectly balanced competition. However, due to its temporal nature, this metric can be significantly influenced by the tournament schedule. To address this, we propose the *expected turning point* ($\hat{\tau}$), which represents the expected value of τ if the tournament followed a more uniform schedule. This is achieved by computing τ across multiple schedule permutations while preserving the original match outcomes, enabling fair comparisons across different leagues and sports by mitigating the effects of scheduling irregularities.

Our analysis highlights the trade-offs between these two metrics. While the *observed* coefficient retains the sequence of matches, making it sensitive to scheduling effects, the *expected* coefficient neutralizes this influence but disregards the potential importance of match order. Applying these metrics to a dataset of 157 sports leagues across four sports, we found that soccer generally exhibits the most balanced tournaments. However, competitive balance varies widely across sports, with both highly balanced and imbalanced leagues present in each. Moreover, comparing the *observed* and *expected* coefficients across all studied seasons revealed several anomalous cases where the *observed* balance deviates significantly from the expected one.

Future Directions. Our framework could be extended to provide a temporal perspective in related contexts, such as measuring *competitive intensity*. Additionally, our approach opens new possibilities for influencing public perception through scheduling. Rather than implementing complex and costly compensation mechanisms to adjust real competitive balance, carefully designed schedules could alter how balanced a tournament appears to audiences.

References

R. Y. Aoki, R. M. Assuncao, and P. O. Vaz de Melo. Luck is hard to beat: The difficulty of sports prediction. In *Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 1367–1376, 2017.

N. Assis, R. Assunção, and P. O. Vaz-de Melo. Stop the clock: Are timeout effects real? In *Machine Learning and Knowledge Discovery in Databases. Applied Data Science and Demo Track: European Conference, ECML PKDD 2020, Ghent, Belgium, September 14–18, 2020, Proceedings, Part V*, pages 507–523. Springer, 2021.

A. Avila-Cano and F. Triguero-Ruiz. On the control of competitive balance in the major european football leagues. *Managerial and Decision Economics*, 44:1254–1263, 3 2023. ISSN 10991468.

A. Avila-Cano, P. D. Owen, and F. Triguero-Ruiz. Measuring competitive balance in sports leagues that award bonus points, with an application to rugby union. *European Journal of Operational Research*, 309:939–952, 9 2023. ISSN 03772217.

A. G. Barnett, J. C. Van Der Pols, and A. J. Dobson. Regression to the mean: what it is and how to deal with it. *International journal of epidemiology*, 34(1):215–220, 2005.

K. Baydina, P. Parshakov, and M. Zavertiaeva. Uncertainty of Outcome and Attendance: Evidence from Russian Football. *International Journal of Sport Finance*, 16(1):33–43, Feb 2021.

E. Ben-Naim, F. Vazquez, and S. Redner. What is the most competitive sport? *arXiv preprint physics/0512143*, 2005.

M.-a. Benz, L. Brandes, and E. Franck. Do soccer associations really spend on a good thing? empirical evidence on heterogeneity in the consumer response to match uncertainty of outcome. *Contemporary Economic Policy*, 27(2): 216–235, 2009.

L. M. Besters, J. C. van Ours, and M. A. van Tuijl. How outcome uncertainty, loss aversion and team quality affect stadium attendance in Dutch professional football. *Journal of Economic Psychology*, 72:117–127, Jun 2019.

V. Borooah and J. E. Mangan. Measuring competitive balance in sports using generalised entropy with an application to english premier league football. *Applied Economics*, page 1, 2011.

R. A. Bowman, J. Lambrinos, and T. Ashman. Competitive balance in the eyes of the sports fan: Prospective measures using point spreads in the nfl and nba. *Journal of Sports Economics*, 14(5):498–520, 2013.

R. A. Bradley and M. E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

B. Buraimo and R. Simmons. Do sports fans really value uncertainty of outcome? Evidence from the English premier league. *International Journal of Sport Finance*, 3(3), 2008.

R. Caruso, F. Addesa, and M. Di Domizio. The Determinants of the TV Demand for Soccer: Empirical Evidence on Italian Serie A for the Period 2008-2015. *Journal of Sports Economics*, 20(1):25–49, 2019.

M. Cattelan, C. Varin, and D. Firth. Dynamic Bradley–Terry modelling of sports tournaments. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 62(1):135–150, 2013.

D. Coates, B. R. Humphreys, and L. Zhou. Reference-dependent preferences, loss aversion, and live game attendance. *Economic Inquiry*, 52(3):959–973, 2014.

C. Collins and B. R. Humphreys. Contest Outcome Uncertainty and Fan Decisions: A Meta-Analysis. *Journal of Sports Economics*, 23(6):789–807, 2022.

A. Cox. Spectator demand, uncertainty of results, and public interest: Evidence from the english premier league. *Journal of Sports Economics*, 19(1):3–30, 2018.

R. Criado, E. García, F. Pedroche, and M. Romance. A new method for comparing rankings through complex networks: Model and analysis of competitiveness of major european soccer leagues. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 23(4), 2013.

D. de Werra. Scheduling in Sports. *Studies on graphs and discrete programming*, 11:381–395, 1981.

J. del Corral, J. García-Unanue, and F. Herencia-Quintanar. Are nba policies that promote long-term competitive balance effective? what is the price? *The Open Sports Sciences Journal*, 9(1), 2016.

H. M. Dietl, M. Lang, and A. Rathke. The Combined Effect of Salary Restrictions and Revenue Sharing in Sports Leagues. *Economic Inquiry*, 49(2):447–463, 4 2011. ISSN 0095-2583.

M. Doria and B. Nalebuff. Measuring competitive balance in sports. *Journal of Quantitative Analysis in Sports*, 17:29–46, 3 2021. ISSN 15590410.

J. Douvis. What makes fans attend professional sporting events? a review. *Advances in sport management research journal*, 1:40–70, 2014.

E. W. Eckard. The Uncertainty-of-Outcome Hypothesis and the Industrial Organization of Sports Leagues. *Journal of Sports Economics*, 18(3):298–317, Apr 2017.

P. J. Ferguson and K. R. Lakhani. Consuming Contests: The Effect of Outcome Uncertainty on Spectator Attendance in the Australian Football League*. *Economic Record*, 99(326):410–435, 2023.

D. Forrest 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(2):229–241, 6 2002. ISSN 00390526.

R. Fort and J. Quirk. Cross-subsidization, incentives, and outcomes in professional team sports leagues. *Journal of Economic Literature*, 33:1265–1299, 1995.

R. Fort and J. Quirk. Introducing a competitive economic environment into professional sports. 1997.

J. García Unanue, A. Godoy, L. Villarrubia, J. Sánchez Sánchez, L. Gallardo, et al. Competitive balance in european basketball leagues and the nba. *Cultura_Ciencia_Deporte [CCD]*, 9(27), 2014.

B. Gerrard and M. Kringstad. The multi-dimensionality of competitive balance: evidence from european football. *Sport, Business and Management: An International Journal*, 12:382–402, 8 2022. ISSN 20426798.

D. R. Goossens, J. Beliën, and F. C. R. Spieksma. Comparing league formats with respect to match importance in belgian football. *Annals of Operations Research*, 194:223–240, 4 2012. ISSN 0254-5330.

J. Groot and L. Groot. The competitive balance of French football 1945-2002. *Journal of Sports Economics*, 3(2), 2003.

M. Haan, R. H. Koning, and A. Van Witteloostuijn. Competitive balance in national european soccer competitions. In *Statistical thinking in sports*, pages 75–88. Chapman and Hall/CRC, 2007.

C. Hantau, A. Alexandru, A. Yannakos, and C. Hantau. Analysis of the competitional balance in the romanian women handball. *Procedia-Social and Behavioral Sciences*, 117:672–677, 2014.

K. K. Haugen. Point score systems and competitive imbalance in professional soccer. *Journal of Sports Economics*, 9(2): 191–210, 2008. ISSN 15270025.

K. K. Haugen and B. Guvåg. Uncertainty of outcome and rule changes in european handball. *European Journal of Sport Studies*, 2018.

I. Horowitz. The Increasing Competitive Balance in Major League Baseball. *Review of industrial Organization*, 12, 1997.

B. R. Humphreys. Alternative measures of competitive balance in sports leagues. *Journal of sports economics*, 3(2):133–148, 2002.

M. Hyun, G. J. Jones, W. F. Jee, J. S. Jordan, J. Du, and Y. Lee. Revisiting the Uncertainty of Outcome Hypothesis and the Loss Aversion Hypothesis in the National Basketball Association: Adding a Predicted Game Quality Perspective. *Journal of Sports Economics*, 24(8):1076–1096, dec 2023.

A. Jadbabaie, A. Makur, and D. Shah. Estimation of skill distribution from a tournament. *Advances in Neural Information Processing Systems*, 2020.

N. Jennett. Attendances, Uncertainty of Outcome and Policy in Scottish League Football. *Scottish Journal of Political Economy*, 31(2):176–198, 6 1984. ISSN 0036-9292.

S. Jungić, J. Jovanović, M. Mihajlović, and S. Simović. Comparative analysis of competitive balance of basketball leagues. *Choregia*, 11(1), 2015.

S. Kesenne. The Win Maximization Model Reconsidered. *Journal of Sports Economics*, 7(4):416–427, 11 2006. ISSN 1527-0025.

B. Kringstad and M. Gerrard. The concepts of competitive balance and uncertainty of outcome. *The Economics and Management of Mega Athletic Events*, 115, 2004.

M. Kringstad. Comparing competitive balance between genders in team sports. *European Sport Management Quarterly*, 21 (5):764–781, 2021.

A. Macedo, M. Ferreira Dias, and P. R. Mourão. European Men's Club Football in the Eyes of Consumers: The Determinants of Television Broadcast Demand. *Journal of Sports Economics*, 24(5):579–623, 2023.

V. Manasis. Measurement of competitive balance in professional team sports using the adjusted entropy. *Economics Bulletin*, 42(2):1124–1134, 2022.

V. Manasis, V. Avgerinou, I. Ntzoufras, and J. J. Reade. Measurement of competitive balance in professional team sports using the normalized concentration ratio, 2011.

V. Manasis, I. Ntzoufras, and J. J. Reade. Competitive balance measures and the uncertainty of outcome hypothesis in European football. *IMA Journal of Management Mathematics*, 33(1):19–52, 2022. ISSN 1471-678X.

J. Maxcy and P. Milwood. Regulation by taxes or strict limits. *Sport, Business and Management: An International Journal*, 8 (1):52–66, 1 2018. ISSN 2042-678X.

J. Michie and C. Oughton. *Competitive balance in football: Trends and effects*. The sportsnexus London, 2004.

G. Nalbantis, T. Pawlowski, and D. Coates. The fans' perception of competitive balance and its impact on willingness-to-pay for a single game. *Journal of Sports Economics*, 18(5):479–505, 2017.

W. C. Neale. The peculiar economics of professional sports: A contribution to the theory of the firm in sporting competition and in market competition. *The Quarterly Journal of Economics*, 78:1, 2 1964. ISSN 00335533.

D. Owen. *Measurement of competitive balance and uncertainty of outcome*. Edward Elgar Publishing, 11 2014.

P. D. Owen. Limitations of the relative standard deviation of win percentages for measuring competitive balance in sports leagues. *Economics Letters*, 109(1):38–41, 10 2010. ISSN 01651765.

P. D. Owen. Measurement of competitive balance and uncertainty of outcome. *Handbook on the economics of professional football*, pages 41–59, 2013.

P. D. Owen and N. King. Competitive balance measures in sports leagues: The effects of variation in season length. *Economic Inquiry*, 53(1):731–744, 2015. ISSN 14657295.

P. D. Owen, M. Ryan, and C. R. Weatherston. Measuring competitive balance in professional team sports using the herfindahl-hirschman index. *Review of Industrial Organization*, 31:289–302, 2007.

T. Pawlowski and O. Budzinski. The (monetary) value of competitive balance for sport consumers: A stated preferences approach to european professional football. Ilmenau Economics Discussion Papers 77, Ilmenau, 2012.

D. Plumley, G. Ramchandani, and R. Wilson. Mind the gap: an analysis of competitive balance in the english football league system. *International Journal of Sport Management and Marketing*, 18(5):357–375, 2018.

D. Plumley, G. M. Ramchandani, and R. Wilson. The unintended consequence of financial fair play: An examination of competitive balance across five european football leagues. *Sport, Business and Management: An International Journal*, 9:118–133, 5 2019. ISSN 20426798.

J. Quirk and R. D. Fort. *Pay dirt: The business of professional team sports*. Princeton University Press, 6 1997. ISBN 9780691187945.

G. Ramchandani. Competitiveness of the english premier league (1992-2010) and ten european football leagues (2010). *International Journal of Performance Analysis in Sport*, 12(2):346–360, 2012.

G. Ramchandani, D. Plumley, S. Boyes, and R. Wilson. A longitudinal and comparative analysis of competitive balance in five european football leagues. *Team Performance Management: An International Journal*, 24(5/6):265–282, 2018.

R. V. Rasmussen and M. A. Trick. Round robin scheduling–a survey. *European Journal of Operational Research*, 188(3), 2008.

B. Reilly. Testing a variant of match-level outcome uncertainty using historical data from the European Champion Clubs' Cup. *Sports Economics Review*, 4(June):100022, 2023.

S. Rottenberg. The baseball players' labor market. *Journal of political economy*, 64(3):242–258, 1956.

N. Scelles. Star quality and competitive balance? Television audience demand for English Premier League football reconsidered. *Applied Economics Letters*, 24(19):1399–1402, Nov 2017.

M. B. Schmidt and D. J. Berri. Competitive balance and attendance: The case of major league baseball. *Journal of Sports Economics*, 2:145–167, 5 2001. ISSN 1527-0025.

Schreyer, B. Torgler, and S. L. Schmidt. Game Outcome Uncertainty and Television Audience Demand: New Evidence from German Football. *German Economic Review*, 19(2):140–161, May 2018a.

D. Schreyer, S. L. Schmidt, and B. Torgler. Game Outcome Uncertainty in the English Premier League: Do German Fans Care? *Journal of Sports Economics*, 19(5):625–644, 2018b.

G. Scully. *The Business of Major League Baseball*. University of Chicago Press, 1989. ISBN 9780226743899.

D. Sheng and H. A. Montgomery. Football league brands: consumer perceptions and the role of competitive balance. *Sport, Business and Management: An International Journal*, 15(2):204–224, Mar 2025.

B. R. Sziklai, P. Biró, and L. Csató. The efficacy of tournament designs. *Computers and Operations Research*, 144, 2022.

S. Szymanski. The economic design of sporting contests. *Journal of economic literature*, 41(4):1137–1187, 2003.

F. Triguero Ruiz and A. Avila-Cano. The distance to competitive balance: a cardinal measure. *Applied Economics*, 51(7): 698–710, 2 2019. ISSN 0003-6846.

J. Utt and R. Fort. Pitfalls to Measuring Competitive Balance With Gini Coefficients. *Journal of Sports Economics*, 3(4), 2002. ISSN 15270025.

T. van der Burg. Competitive balance and demand for European men's football: a review of the literature. *Managing Sport and Leisure*, pages 1–16, 2023.

X. Wang. Shining stars, unpredictable outcomes: the impact of outcome uncertainty and star power on the online attention of the Chinese Football Association Super League. *Applied Economics*, pages 1–16, Feb 2025.

S. Wilkens. Sports prediction and betting models in the machine learning age: The case of tennis. *Econometrics: Econometric & Statistical Methods - General eJournal*, 2020.

G. Wills, R. Tacon, and F. Addesa. Uncertainty of outcome, team quality or star players? What drives TV audience demand for UEFA Champions League football? *European Sport Management Quarterly*, 22(6):876–894, Nov 2022.

G. Wills, F. Addesa, and R. Tacon. Stadium attendance demand in the men's UEFA Champions League: Do fans value sporting contest or match quality? *PLoS ONE*, 18(2 February):1–22, 2023.

J. Winfree and R. Fort. Nash Conjectures and Talent Supply in Sports League Modeling. *Journal of Sports Economics*, 13 (3):306–313, 6 2012. ISSN 1527-0025.

P. Wonga. Relationship between football clubs'competitive performance and revenue streams. 2023.

A. S. Zimbalist. Competitive balance in sports leagues: An introduction. *Journal of Sports Economics*, 3:111–121, 5 2002. ISSN 1527-0025.

A Other Point Systems

Considering that our coefficient directly depends on the variance of the standings, how many points teams make after matches is a relevant factor. For soccer and handball, our 3, 1, and 0 points for, respectively, a win, draw, and loss is standard for most competitions. For basketball and volleyball, the impossibility of drawing makes our point system equivalent to counting the number of wins each team achieved. However, given that volleyball matches are based on the *set* result, rather than the points the teams accrued in the match, there is another common point system. Specifically, 3-0 and 3-1 victories still provide 3 points for the winner and 0 for the loser; whereas a 3-2 victory indicates that both teams are similarly matched, so it is worth 2 points for the winner and 1 for the loser.

Figure 6 compares both point systems for all volleyball tournaments in our dataset. While the difference is not large, a slight discrepancy can be observed. The perceived competitive balance in this new system tends to be smaller than in the previous one, suggesting that it makes tournaments more predictable than they could be. We believe that an explanation for this difference is that it mostly affects the teams in the middle of the standings, grouping them closer and reducing the average points slightly. Strong (weak) teams will most likely win (lose) matches by 3-0 or 3-1, so they will not be influenced very much by this change. Thus, the real setting has a bias for these teams: the more (less) points teams have, the more likely they are to get a 3-0 or 3-1 victory (defeat). This is not accounted for in the simulations, so the teams at the top (bottom) of the standings do not separate themselves from the rest as much as they do in real life, accelerating the τ detachment.

B Most Imbalanced Tournament

Given their broad application in other fields, the Gini index and the HHI coefficients already have normalized versions. However, they rely on upper bounds which are impossible to obtain in conventional round-robin tournaments, since a single team cannot win all matches. Therefore, these coefficients should be normalized in sports leagues using the upper bound from the most unbalanced tournament possible (Fort and Quirk,

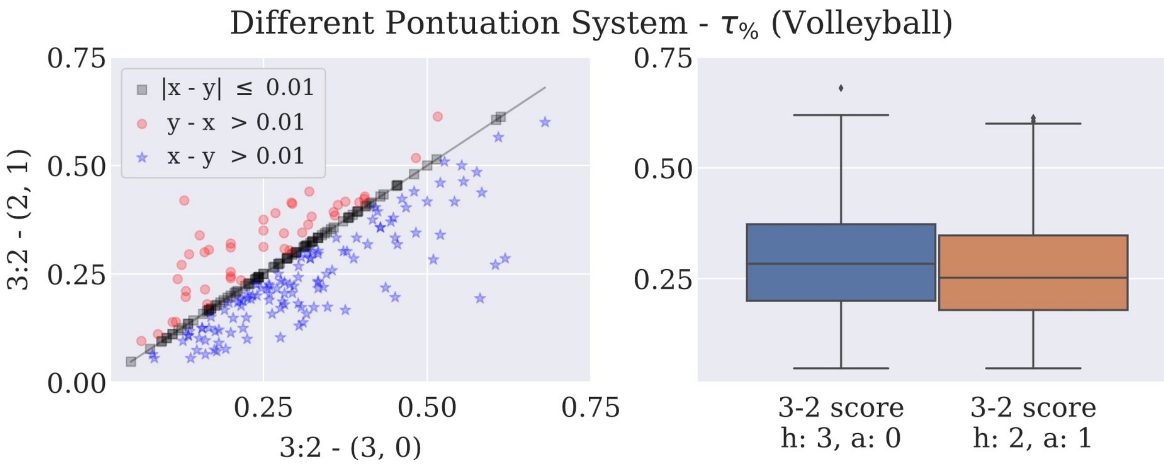

Fig. 6: Comparison between two different points systems: a 3-2 victory gives 2 (3) points to the winner and 1 (0) to the loser. On the left, we show the one-to-one comparison for all volleyball seasons, and on the right, we illustrate them as a boxplot.

1997; Utt and Fort, 2002). In such tournament tournament, the first-place team wins all matches they play, the second-place team wins all the matches except those against the first-place team, and so on. From this definition, it is possible to estimate the upper bound threshold – for example, the HHI upper bound for a balanced round-robin tournaments has a closed form given by $HHI_{ub} = \frac{2(2n-1)}{3n(n-1)}$ (Owen et al., 2007).

Unfortunately, two factors prevent us from using analytical expressions directly: (i) our database is not composed only of competitions with balanced schedules; (ii) calculating the imbalance progression in Section 8.2 requires applying these coefficients to portions of real tournaments that might not be balanced either. With that in mind, we choose to build the most unbalanced tournament and calculate the upper limit directly from it. To achieve this, we slightly generalize the definition of the most unbalanced tournament to work with schedules where different teams play a different number of games. In this new version, the team that wins all its games is the one (or one of) that played the most amount of games; the second-place team is the one who played the second-largest number of games; and so forth. In particular, if all teams play the same number of matches, we have a situation identical to the old definition. Additionally, in the unrealistic limit where a team plays in every tournament match, this team would win every game, and the upper limit would be the same as the one used in the other fields.