Characterizing and Controlling the Perceived Competitive Balance in Sports Leagues

Paper Track: Business of Sports
Paper ID: 85

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 (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, which refers 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). The prevailing hypothesis is that significant disparities in team strength lead to predictable outcomes (Borland, 2003; Csató, 2023; Fort and Quirk, 1995; Neale, 1964), limiting the number of teams with a realistic chance of winning the championship. This can result in less engaging matches, particularly in the mid-to-late stages of the competition (Di Mattia and Krumer, 2023; Douvis, 2014; Neale, 1964; Pawlowski and Nalbantis, 2015). Additionally, in the last rounds of unbalanced tournaments, matches can become "stakeless" for some participants, threatening fairness as it could be a reason to not exert full effort to win (Chater et al., 2021; Csató, 2025).

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).

This work focuses on competitive balance within a tournament season, often called seasonal (or within-season) competitive balance. 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).

In order to foster competitive balance, major sports leagues and organizations such as the NBA, NFL and UEFA have implemented a range of strategies aimed at leveling the playing field. These include draft systems designed to benefit weaker teams (Késenne, 2006; Winfree and Fort, 2012), salary caps intended to mitigate revenue disparities (Dietl et al., 2011; Maxcy and Milwood, 2018; Totty and Owens, 2011), modifications to tournament formats (Csató, 2020; Di Mattia and Krumer, 2023; Gyimesi, 2024), and transaction veto powers to limit the dominance of top-tier teams (Plumley et al., 2019). However, the effectiveness of these measures is not guaranteed and can even be counterproductive, at times exacerbating the very imbalance they seek to address (Fort and Quirk, 1995). In fact, several studies have documented a decline in competitive balance across major sports leagues despite these interventions, including Major League Baseball (MLB) (Depken, 1999; Lewis, 2008), the NBA (Maxcy and Mondello, 2006), and Europe's top five football leagues: England, Spain, Italy, Germany, and France (Avila-Cano and Triguero-Ruiz, 2023; Gasparetto et al., 2023; Pawlowski et al., 2010; Plumley et al., 2019). This trend is largely driven by growing disparities in team revenues and expenditures (Sanderson and Siegfried, 2003; Szymanski and Késenne, 2004), which translate into unequal access to resources, limited opportunities for performance improvements, and ultimately, a reduction in both league competitiveness and overall appeal (Feddersen and Maennig, 2005).

2. Contributions

Although widely used, 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.

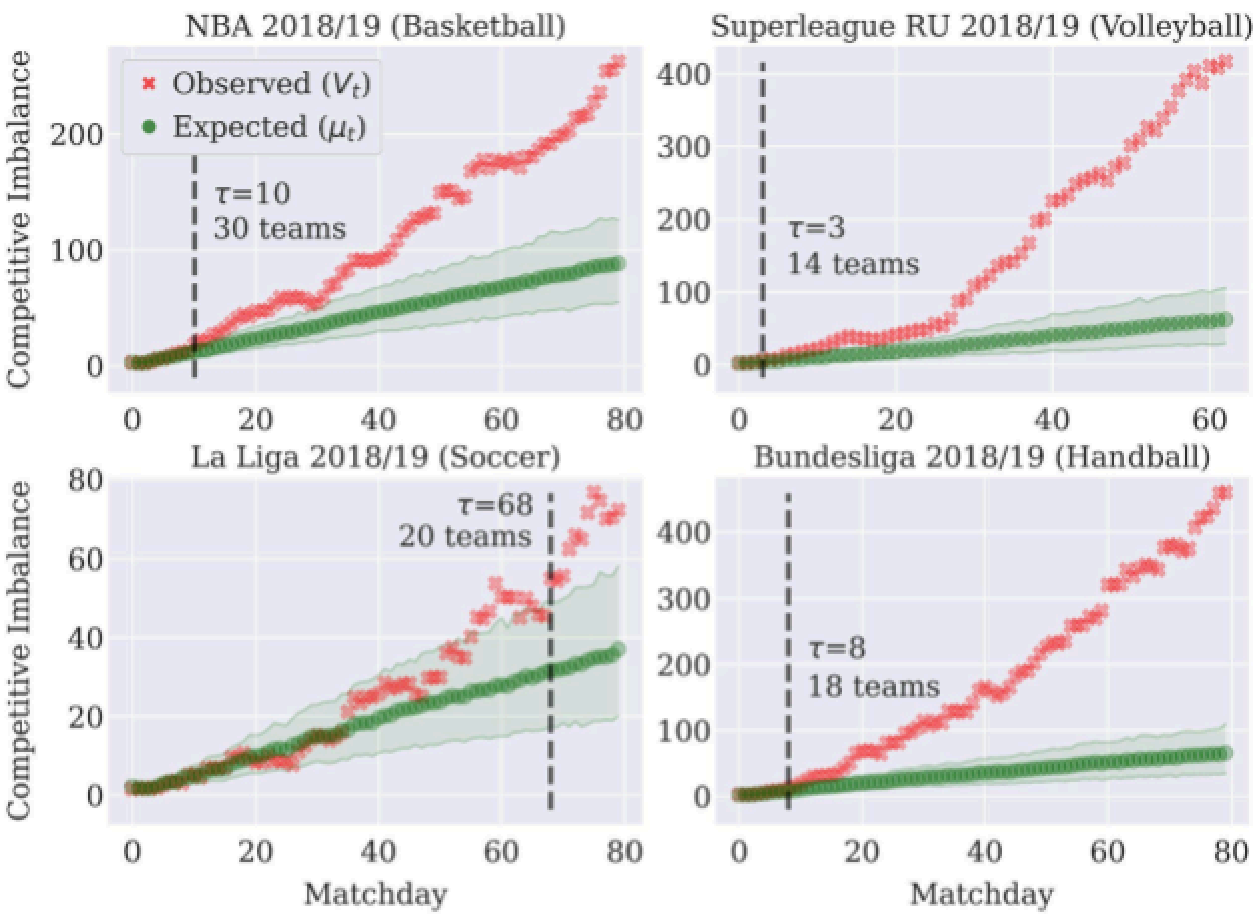

Figure 1. Temporal progression of the competitive imbalance. The red crosses are the empirical standings variances at each time t. The green points represent the expected value for this variance if all the teams had the same skills, while the shaded green region delimits a 95% confidence envelope. The vertical dashed line highlights the first time the empirical variance curve leaves the confidence envelope without ever returning: it is the turning point τ.

Thus, our first contribution is to show that both the matchday τ and the fraction of matchdays occurring before $\tau_{\%}$ are intuitive and interpretable measures of competitive balance. We demonstrate this across 1,392 seasons in 77 countries. Our second contribution addresses the fact that match order directly affects the value of τ. To overcome this issue, we introduce the $\hat{\tau}$ estimator, which represents the expected turning point of a tournament. This coefficient is independent of the schedule and therefore provides a more reliable measure of competitive balance, enabling meaningful comparisons across leagues and sports. Moreover, by combining our metrics, we can identify tournaments whose observed turning point is consistently below its expected value across seasons. In such cases, small adjustments to the schedule could substantially improve the perceived balance of the competition.

We leverage this insight for our third contribution: a novel scheduling approach that provides control over the timing of the divergence point τ. Postponing τ allows the competition to appear balanced for a longer period, increasing its appeal to fans of mid-table teams and enhancing the importance of a larger number of matches. Conversely, advancing τ accelerates the separation between top and bottom teams, helping identify decisive matches earlier and potentially attracting loss-averse fans to more consequential games.

Our results reveal that our algorithm is able to consistently increase the perceived balance across all 1,392 analyzed seasons, opening new possibilities for influencing public perception. 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.

3. 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, 79 football, and 30 handball leagues, as available on the betting site betexplorer.com. The dataset covers seasons from 2011 to 2020, with 10 seasons available for approximately half of the leagues. In total, there are over 279,00 matches spread out across 1,392 seasons, averaging approximately 201 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.

4. Probabilistic Metric

4.1. 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' regular season following the same schedule S as the original tournament C. If all teams $(1, \ldots, n)$ have the same strength, the standings $\mathcal{X}_t = \left(X_{1,t}, \ldots, X_{n,t}\right)$ 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. The likelihood of these outcomes is directly tied to the probability of each match result: home win $\left(P_h\right)$, draw $\left(P_d\right)$, and away win $\left(P_a\right)$. These probabilities are empirically estimated based on the observed frequencies in the real tournament C. In the simulations, each match outcome is randomly assigned according to P_h, P_d, and P_a, replicating the distribution observed in 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.

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 4.3, 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.

4.2. 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.

The procedure to estimate τ for tournament C is as follows. We first run K simulations of the null model following the same schedule S of the original tournament 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 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 $\mathcal{V}_t^{(k)}, \dots, \mathcal{V}_T^{(k)}$ 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 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 $\mathcal{V}_t^{(k)}, \dots, \mathcal{V}_T^{(k)}$ 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:

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

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

$$\hat{\mathcal{F}}_t(q) = \frac{1}{K} \sum_{k=1}^{K} I_{\mathcal{V}_t^{(k)} \leq q_t} \tag{3}$$

with 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 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:

$$\tau = argmax\{t : \mathcal{V}_t < q_t\} \quad (4)$$

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

$$\tau_{\%} = \frac{\tau}{T} \quad (5)$$

Figure 2 shows the boxplots (in purple) 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 $\overline{\tau_{\%}} = 45.9\%$, 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, $\overline{\tau_{\%}} = 32.6\%$, for volleyball, $\overline{\tau_{\%}} = 29.6\%$, 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 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.

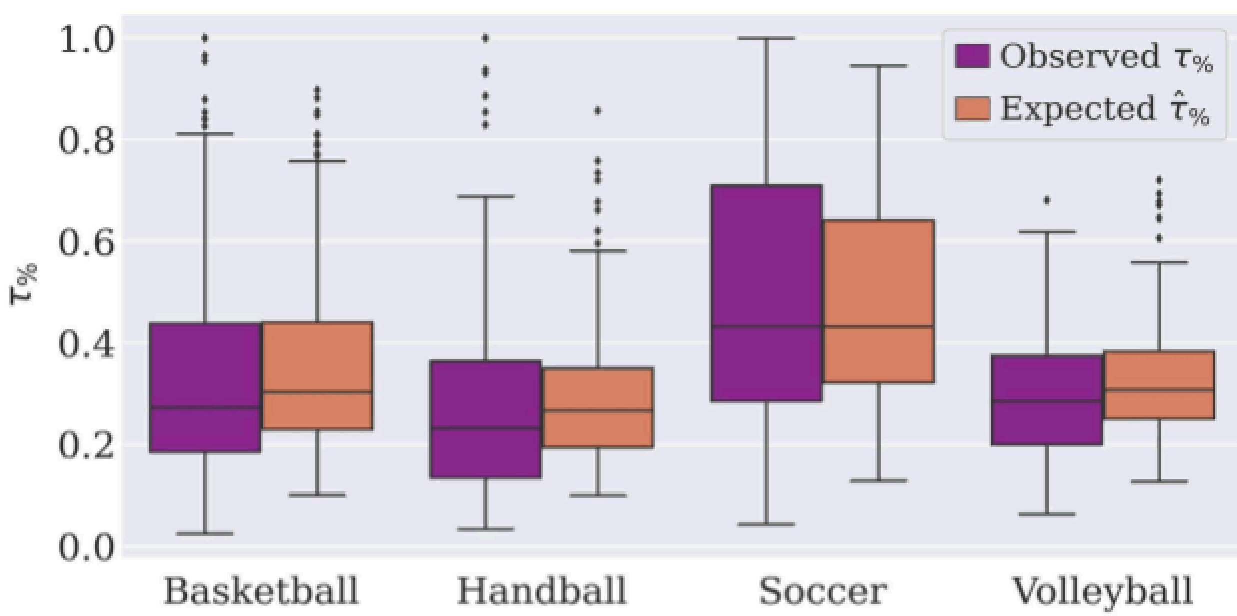

Figure 2. General comparison between the observed and expected turning points for all sports.

4.3. 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 $\left(P_h = 53.8\%\right)$ and away $\left(P_a = 46.2\%\right)$ 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 $E[\tau]$ 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 C ordered by its schedule. It can be seen as a direct result of the underlying skill distribution within the tournament. Now, let S_{M_0} be the set of all possible match orders of M_0, that is, all of its permutations. Then, each permutation $M \in S_{M_0}$ 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 C with observed results M_0 as a random variable $\mathscr{T}$ depending on the schedule $M \in S_{M_0}$ and the underlying skill distribution, represented by the observed match results. To calculate τ for any permutation $M \in S_{M_0}$, we extend the procedure described in the previous section as a function $\tau : S_{M_0} \rightarrow N$ 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

$$E[\tau] \overset{\text{def}}{=} E\left[\mathcal{T} \mid M_0\right] = \sum_{M \in \mathfrak{G}_{M_0}} \tau(M)\, \mathbb{P}(M) \tag{6}$$

Unfortunately, this expression is insufficient to calculate $E[\tau]$ 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,

$$E[\tau] \approx \sum_{M \in \mathfrak{G}^{RR}_{M_0}} \tau(M)\, \mathbb{P}(M) \approx \frac{1}{\left|\mathfrak{G}^{RR}_{M_0}\right|} \sum_{M \in \mathfrak{G}^{RR}_{M_0}} \tau(M) \overset{\text{def}}{=} \hat{\tau} \tag{7}$$

where $\mathfrak{G}^{RR}_{M_0} \in \mathfrak{G}_{M_0}$ represents all possible schedule permutations following a round-robin structure, and $\hat{\tau}$ is how we will denote our estimator henceforth. Since generating all $\left|\mathfrak{G}^{RR}_{M_0}\right|$ 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 $\mathfrak{G}^{RR}_{M_0}$.

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 $M \in \mathfrak{G}^{RR}_{M_0}$ 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.

Figure 2 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, 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.

4.4. Betting Markets

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 3, 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.

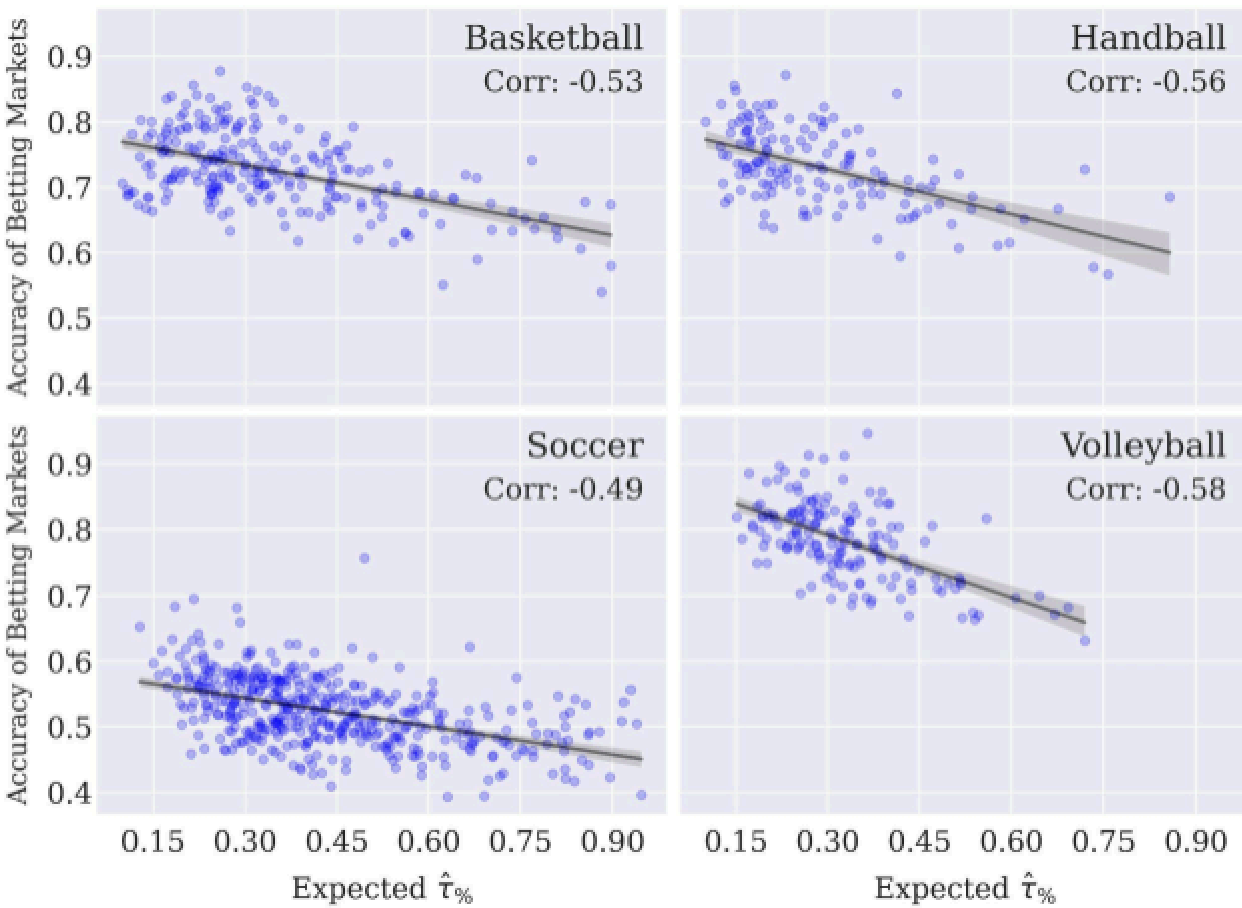

Figure 3. Comparison between the betting market prediction accuracy and competitive balance for all seasons containing betting data. Competitive balance is measured by our normalized expected turning point. The black line represents the linear fit for the data and the number below each sport name is the corresponding Pearson Correlation coefficient.

4.5. Observed vs Expected

Figure 4 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 $CI_{\hat{\tau}}$, 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 the expected turning point $\hat{\tau}$. The number below each sport name is the Pearson Correlation coefficient between $\tau_{\%}$ and $\hat{\tau}_{\%}$.

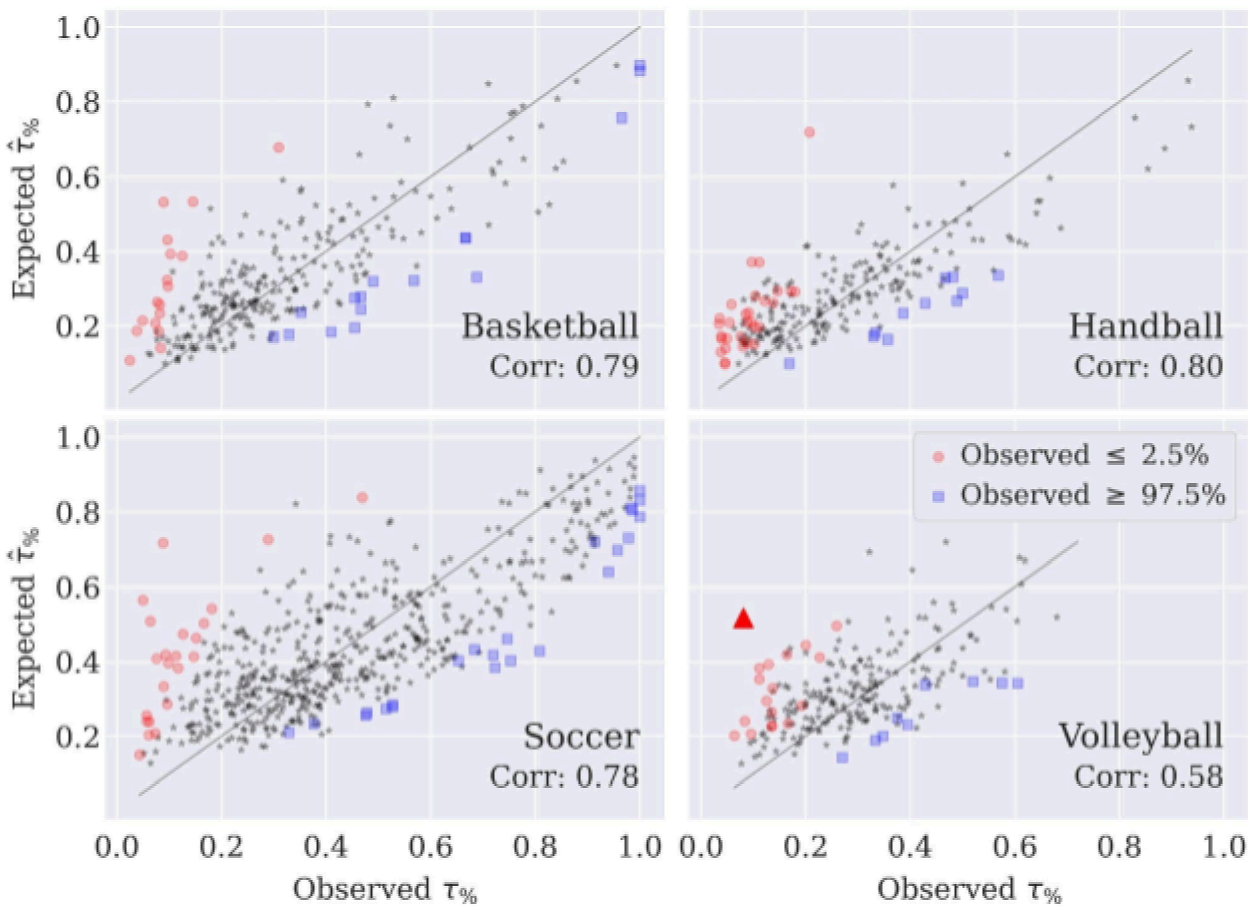

Figure 4. Comparison between the observed and expected turning points for all studied seasons. The red circles represent the tournament whose observed was below the expected confidence interval, whereas for the blue squares, it was above it. The number below each sport name is the Pearson Correlation. We highlight the 2013/2014 volleyball season for Brazilian Superliga as a red triangle.

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 4 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. In total, only 4.02% of all seasons were below the $CI_{\hat{\tau}}$, while 3.09% were above. We conjecture that an observed turning point below the $CI_{\hat{\tau}}$ 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 $CI_{\hat{\tau}}$ might be from some matches being played earlier than they should, breaking the well-behaved round-robin structure.

5. Controlling Competitive Balance

5.1. Formulation

In the previous section (4.2), we formalized when a tournament's imbalance is within the expectation of a perfectly balanced one by defining the observed turning point τ . Here, we focus on what it means to extend this metric. Simply put, we wish to determine which feasible tournament schedule maximizes the duration the imbalance $\mathcal{V}_t$ stays inside the envelope $\mathbb{Q}_{1-\alpha}\left[\mathcal{V}_t\right]$. Consider a double round-robin tournament C composed of n teams and T rounds. Let $S \in \{0,\ 1\}^{T\times n\times n}$ be a third-order tensor indicating when the ordered match (i, j) happened: $S_{tij} = 1$ if the match happened at the round t and 0 otherwise. By ordered match (i, j), we mean that team i plays at home while j plays away. In this formulation, finding which schedule extends the τ for as long as possible is the same as finding the optimal integer entries for S.

Building upon our previous notation, let $x_i(i, j)$ and $x_i(j, i)$ represent the points team i earned in its matches against i, respectively, playing as home and as away. These values are predetermined and remain unchanged because they are the direct projection of the underlying skill distribution in the real world, i.e., the results of these matches are recorded in the tournament data. Thus, the points team i earns until round t can be explicitly calculated for any possible schedule simply by adding up the points accrued in the matches until this round. In particular, for an arbitrary schedule S, the points $X_{i,t}$ obtained by team i up to round t is

$$X_{i,t}(S) = \sum_{\tilde{t}=1}^{t}\sum_{j=1}^{n}\left(x_i(i, j)S_{\tilde{t}ij} + x_i(j, i)S_{\tilde{t}ji}\right). \tag{8}$$

Just as we can express $X_{i,t}$ as a function of the schedule, we can also calculate the standings imbalance $\mathcal{V}_t(S)$. Let the variance be the metric used to calculate the imbalance, and let $\overline{X}_t$ be the average point distribution across all n teams. Then, the imbalance $\mathcal{V}_t(S)$ in round t, given a schedule S, is:

$$\mathcal{V}_t(S) = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_{i,t}(S) - \overline{X}_t\right)^2. \tag{9}$$

Now, since this imbalance directly depends on the schedule, we can carefully define how to find the match order that extends the perceived competitive balance as measured by τ . Essentially, we consider the optimal schedule S^* to be the one that maximizes the duration for which its standings' dispersion $\mathcal{V}_t\left(S^*\right)$ remains within the quantile envelope $\mathbb{Q}_{1-\alpha}\left[\mathcal{V}_t\left(S^*\right)\right]$. As we did with our previous definitions of $X_{i,t}(S)$ and $\mathcal{V}_t(S)$, we extend the τ as a function of the schedule S

$$\tau(S) = argmax\left\{t:\ \mathcal{V}_t(S) \leq \mathbb{Q}_{1-\alpha}\left[\mathcal{V}_t(S)\right]\right\}, \tag{10}$$

from which we can trivially define S^* as

$$S^* = argmax\,\{S:\ \tau(S)\} \tag{11}$$

And the maximized perceived competitive balance τ^* as

$$\tau^* = \tau\left(S^*\right). \tag{12}$$

Note that no constraints are yet associated with Equation 11. They must be added according to the type of tournament at hand to ensure S^* is a feasible schedule. The Double Round-Robin (DRR) format is employed in most professional soccer leagues world- wide, including the top five European leagues. In our dataset, they represent about 90% of the studied seasons. For this format, the maximization should have the following restrictions:

$$S_{tii} = 0 \qquad \forall\, i\ \in \{1,\ ...,\ n\},\ t\ \in \{1,\ ...,\ T\} \tag{13}$$

$$\sum_{t=1}^{T} S_{tij} = 1 \qquad \forall\,(i, j)\ \in \{1,\ ...,\ n\}^2,\ i\ \neq j \tag{14}$$

$$\sum_{j=1}^{n}\left(S_{tij} + S_{tji}\right) = 1 \qquad \forall\, i\ \in \{1,\ ...,\ n\},\ t\ \in \{1,\ ...,\ T\} \tag{15}$$

$$\sum_{t=1}^{T/2}\left(S_{tij} + S_{tji}\right) = 1 \qquad \forall\,(i, j)\ \in \{1,\ ...,\ n\}^2,\ i\ \neq j \tag{16}$$

Equation 13 states that no team will ever face itself during the competition, while Equation 14 indicates that each team will play against all others exactly once at home and once away. Equation 15 limits the number of times each team can play in a round t to 1. Finally, Equation 16 divides the tournament into two turns, forcing teams to face each other once before facing any other team again. A more general formulation, applicable to tournaments that follow formats other than the double round-robin (e.g., the NBA), can be found in the Appendix.

5.2. Iterative Maximum Weighted-Matching Schedules (iMWM)

To simplify, we generate S^* for a single round-robin tournament, meaning each team competes against every other team exactly once. This schedule S^* can be extended to a double round-robin tournament by mirroring the generated schedule and switching the home-court advantage for the second half of the tournament. Similar methods can be applied to create schedules for multiple round-robin tournaments.

Moreover, we construct a schedule S' that minimizes τ using a τ-minimizer algorithm. Then, considering that S_t is the tensor component equivalent to the $n \times n$ matrix denoting all the matches in S scheduled in round t, we set S^* to be the schedule S' played in reverse order. There is no guarantee that reversing the minimizer schedule necessarily converts it into a maximizer one, it is simply a good heuristic to approximate it.

We note that reversing S' also involves changing the home-away status. Fairness aside, setting the best team to play at home in the initial matches is optimal to minimize the perceived competitive balance (PCB) since it will increase their likelihood of winning. Consequently, it will accelerate the separation between good and bad teams, increasing the point distribution variance and, as a consequence, reducing the perceived competitive balance. For a maximization approach, the opposite occurs. It is favorable to set the worst team as home, leveraging the home-court advantage to mitigate the difference in skill in a match. It is possible, then, to better mimic a perfectly balanced tournament by ensuring that each team’s winning probability is as even as possible.

Algorithm 1 illustrates how to implement the τ-minimizer interface by maximizing the dispersion of the points distribution as the tournament progresses. To achieve this, we propose a greedy approach that matches teams with the greatest skill discrepancy in each round, meaning the best teams are scheduled to face the worst possible teams. This strategy increases the probability of good teams earning points by making their matches as easy as possible while decreasing the probability of bad teams earning points by making their matches as tricky as possible. With the best teams regularly gaining points and the worst teams not, the dispersion of the points distribution should increase rapidly.

Algorithm 1. Iterative **M**aximum **W**eighted-**M**atching (iMWM)

Input : List of team rankings or strengths R

Output: A schedule $\mathcal{S}'$ that minimizes τ

Function τ-minimizer(R):

  $G_1 \leftarrow$ CompleteGraph($|R|$)

  for $1 \leq t \leq T$ **do**

    $A_t \leftarrow$ MaxWeightMatching(G_t)

    $\mathcal{S}'_t \leftarrow A_t \odot U$

    $G_{t+1} \leftarrow G_t - A_t$

  end

  return $\mathcal{S}'$

The only input required to match teams based on their skills is the list $R = \{r_1, ..., r_n\}$ of teams ordered by their skill levels. From R, we create an undirected weighted graph $G_t(N, E_t, w)$, where the set of nodes N are the n teams $\{r_1, ..., r_n\}$ and an edge exists between two teams r_i and r_j if they are allowed to play in round t. The function $w: E \rightarrow R$ assigns a real number to each edge (r_i, r_j) describing the skill discrepancy between teams r_i and r_j. The graph G_t represents all matches that can be scheduled in round t. In the first round, G_1 is a complete graph, meaning that all teams can play each other. In the following rounds, all edges representing matches already scheduled are removed from G_1.

To maximize the skill discrepancy in the matches scheduled for round t, we need to find a set of edges in Gt with no duplicate nodes and with the maximum possible weight sum, which is equivalent to solving a max-weight matching problem for this graph (Galil, 1986). However, this solution corresponds to a set of undirected edges, meaning we still need to define which team plays at home and which plays away. To minimize τ, we set the home court advantage to the higher-ranked team for all edges, as previously explained.

More formally, let A_t be the adjacency matrix corresponding to the list of undirected edges returned by the max-weight matching algorithm applied to graph G_t. Moreover, let U be a strictly upper triangular matrix of size $n \times n$, with ones in the cells above the main diagonal.

Then, we can assign the matches S'_t scheduled for round t by setting $S'_t = A_t \odot U$, where $\odot$ denotes element-wise matrix multiplication, since U filters out all elements below the main diagonal. This process will yield a strictly upper triangular matrix S'_t, containing the matches in A_t, with the highest-ranked team always playing at home.

As for the weighting function, we employed a straightforward approach that considers only the ranks of the teams since ranking is easier than determining the exact team skill. Specifically, we define the discrepancy $w\left(r_i, r_j\right)$ as the squared difference between their ranks:

$$w\left(r_i, r_j\right) = (i - j)^2 \qquad (17)$$

We chose the squared distance over the more natural $|i - j|$ due to its faster growth. Using the absolute difference often results in ambiguity, since it would be too easy to obtain the same final sum when adding the weights of all edges.

5.3. Results

To construct our τ-maximizer schedules, the only input required is a list $R = \left\{r_1, ..., r_n\right\}$ of teams ordered by skill level, which can be derived from a prediction model or provided by experts. To assess the effectiveness of our algorithms independently of rank prediction accuracy, we use an *"oracle"* approach, where R is generated directly from the actual final tournament rankings. Additionally, we evaluate our algorithms using the *"yesterday"* prediction model, which trivially reflects the final rankings of the previous season in R. In practice, the quality of R is expected to fall between the predictions of the yesterday model (as a lower bound) and the oracle approach (as an upper bound).

Figure 5 illustrates the impact of the iMWM algorithms on all double round-robin tournaments using the oracle team skill estimation. Each point on the graphs represents a season of a sports league, with the horizontal axis indicating the normalized perceived competitive balance $\tau_{\%}$ for that season and the vertical axis showing the maximized $\tau_{\%}^{max}$ produced by the τ-maximizer algorithm. The red circles represent seasons where the algorithm would have increased the perceived competitive balance τ, the blue stars indicate those where the algorithms would have decreased it, and the black squares denote seasons where it would have remained unchanged. Notably, regardless of the algorithm used, the dominance of red circles strongly suggests that an accurate team-ranking estimation would consistently and significantly increase the perceived competitive balance across all the sports seasons analyzed. For all eight graphs, at least 96% of the seasons considered experienced a strict increase in their perceived competitive balance.

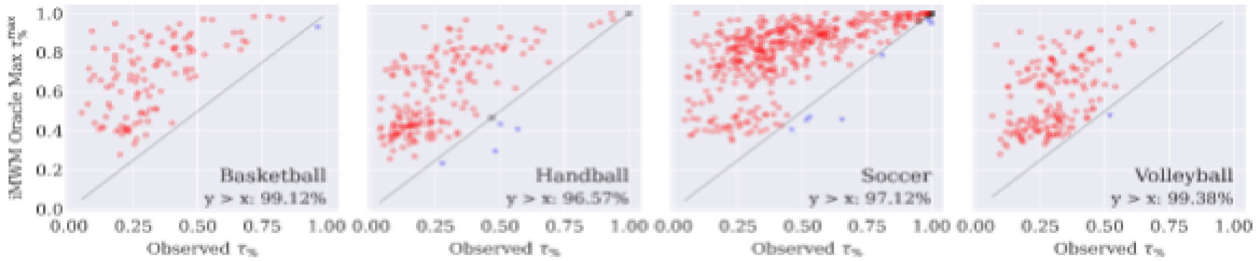

Figure 5. Comparison between the normalized observed turning point and the maximized version generated by our τ-maximizer scheduler using an oracle iMWM approach. Tournaments where the maximized is significantly greater than the observed are represented by red circles, while those where it is smaller are blue stars. If they are within 1% of each other, we illustrate them as black squares around the black x = y line.

Nevertheless, in a few cases, the real schedule ended up being more competitive than the one proposed by our oracle algorithm. The reason could be that the standings were not an accurate estimation of the skill distribution for the entire tournament. For example, this could happen if some teams started the season strong but faltered towards the end. Another option would be if the skill discrepancy is so gigantic that even the difference between the first and second-ranked teams is already too significant. In this situation, it is impossible to manipulate the tournament in a way that mimics a competitively balanced tournament since the games will never be closely matched. Fortunately, due to the maximizing nature of our algorithm, even when there is a decrease in τ, its magnitude tends to be small. The figures illustrate this behavior since the blue stars are usually considerably closer to the black equality line than most red circles.

As for the yesterday model , we show the average results for all sports in Table 1 where column *Oracle?* is unchecked. Only first-division tournaments are considered for this approach since dealing with promotion and relegation quickly becomes complicated. Also, we placed all promoted teams in alphabetical order at the bottom of the standings, i.e., we assume they are worse than all teams that stayed. Although this approach is simple, the perceived competitive balance τ for about 69% of all seasons still increased with our iMWM algorithm.

Table 1. Summary results for double round-robin tournaments. The checkmark columns represent respectively *Oracle/Yesterday* methodology, *Mirrored/Reversed* second tournament phase. The first result column indicates the proportion of tournaments whose τ increased with our algorithm. The other two illustrate, for the tournaments where τ increased (decreased), how much did it increase (decrease) by.

Oracle?	*Mirrored?*	**Success Rate**	**Avg Increase**	**Avg Decrease**
✔	✔	95.7%	33.3%	6.3%
✔		86.1%	22.2%	10.1%
	✔	69.4%	22.9%	10.4%
		64.3%	18.7%	10.8%

5.4. Practical Scenarios

In this section, we present the results of our algorithms across various settings. These settings are essentially combinations of the following three configurations:

1. **Oracle?** Whether the algorithm uses the oracle model or the yesterday model .
2. **Mirrored?** Whether the single round-robin schedule is replicated in subsequent turns or reversed, leaving the most balanced games for the final rounds.

Additionally, all results are summarized using three key metrics:

1. **Success Rate:** The percentage of tournaments where the algorithm increases the perceived competitive balance τ .
2. **Average Increase:** The average percentage increase in τ for tournaments where the algorithm increases τ .
3. **Average Decrease:** The average percentage decrease in τ for tournaments where the algorithm does not increase τ .

While balanced tournaments are desirable, it is crucial to recognize that the optimal τ-maximizer schedule may not always align with certain public interests. More specifically, mimicking a competitive league through such a schedule, even ignoring fairness implications, could violate broadcasting interests since most attractive matches would happen way before the end of the tournament. With that in mind, we propose an intermediate scheduler approach, namely τ-maxmin, which uses the τ-maximizer for the first half and the τ-minimizer for the second half of the tournament. Intuitively, by placing a τ-minimizer scheduler in the second half of the tournament, we sacrifice the perception of competitive balance to ensure the most thrilling matches occur towards the tournament's conclusion. However, we acknowledge that, in some cases, these matches could become stakeless and have less impact on the standings, which could in turn lower attendance (Buraimo et al., 2022; Csató, 2025).

Naturally, the τ-maxmin approach is more effective for tournaments already following a double round-robin (DRR) structure. It can, however, be extended for any multi-round-robin tournament by alternating between τ-maximizer and τ-minimizer turns. Table 1 shows the results of the τ-maxmin approach for DRR tournaments where the column *Mirrored?* is unchecked. Observe that about 86% of all seasons still showed an increase in τ for the oracle approach, whereas about 64% for the yesterday model. As expected, the τ-maxmin approach is applied to general tournaments, the results are slightly worse, respectively, around 79% and 62% for the oracle and yesterday model approaches. We provide the table for all tournaments in the Appendix.

Nevertheless, it is important to emphasize that only the first half is being maximized. Thus, this approach would manage to extend the perceived competitive balance of the vast majority of tournaments whose real τ happened before the first half. Yet, it could decrease for tournaments where the real τ is already really high — especially if the rankings estimation is not optimal, as in the yesterday model . Put differently, the τ-maximizer effectively simulates the expected fluctuations of a competitively balanced tournament during its initial phase. Beyond this point, the larger the actual perceived competitive balance, the less likely it became for our second half to prolong its perceived competitive balance, given that the τ-minimizer sacrificed balance to prioritize the most captivating matches towards the end. We illustrate this behavior in Figure 6. Observe that the blue dots (tournaments that would have their τ decreased) mainly occur after the tournament's first half.

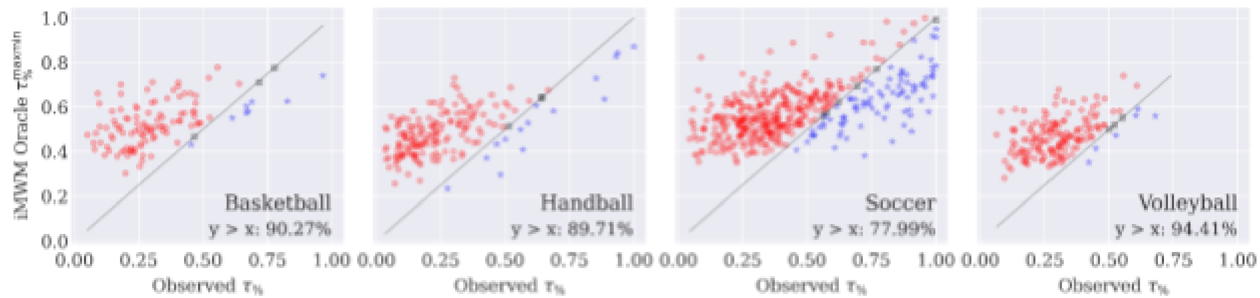

Figure 6. Comparison between the normalized observed turning point and the half-maximized version generated by our τ maxmin scheduler using an oracle iMWM approach.

6. 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 1,392 sports leagues seasons 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.

Moreover, we leverage the effect scheduling has on our observed coefficient to propose a new avenue to control the perception of competitive balance in sports leagues. Unlike traditional methods, our solution does not require changes to tournament regulations or format, nor does it rely on matchups being determined as the tournament progresses. Instead, we focus on designing a schedule that keeps the points distribution tight for as long as possible, enhancing the perception that the tournament features teams of similar skill levels. We initially provide a general formulation of the problem aimed at manipulating the observed turning point by only altering the order (i.e., the schedule) of the games while keeping all other aspects of the tournament unchanged. Finally, we propose an heuristic algorithm, named Iterative Maximum Weighted-Matching Scheduler, to create great approximations for the solution. We tested it on our dataset, and demonstrated that it can significantly extend (or delay) the number of initial rounds in which the teams exhibit skill parity, potentially enhancing audience engagement over longer periods.

References

Avila-Cano, A., Owen, P. D., & Triguero-Ruiz, F. (2023). Measuring competitive balance in sports leagues that award bonus points, with an application to rugby union. *European Journal of Operational Research*, *309*(2), 939–952.

Avila-Cano, A., & Triguero-Ruiz, F. (2023). On the control of competitive balance in the major European football leagues. *Managerial and Decision Economics*, *44*(2), 1254–1263.

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

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

Borland, J. (2003). Demand for Sport. *Oxford Review of Economic Policy*, *19*(4), 478–502.

Borooah, V., & Mangan, J. E. (2011). Measuring Competitive Balance in Sports using Generalised Entropy with an Application to English Premier League Football. *Applied Economics*, 1.

Bradley, R. A., & Terry, M. E. (1952). Rank Analysis of Incomplete Block Designs: I. The Method of Paired Comparisons. *Biometrika*, *39*(3/4), 324–345.

Buraimo, B., Forrest, D., McHale, I. G., & Tena, J. D. (2022). Armchair fans: Modelling audience size for televised football matches. *European Journal of Operational Research*, *298*(2), 644–655.

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

Caruso, R., Addesa, F., & Di Domizio, M. (2019). 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.

Chater, M., Arrondel, L., Gayant, J.-P., & Laslier, J.-F. (2021). Fixing match-fixing: Optimal schedules to promote competitiveness. *European Journal of Operational Research*, *294*(2), 673–683.

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

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

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

Criado, R., García, E., Pedroche, F., & Romance, M. (2013). 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).

Csató, L. (2020). Optimal Tournament Design: Lessons From the Men's Handball Champions League. *Journal of Sports Economics*, *21*(8), 848–868.

Csató, L. (2023). How to avoid uncompetitive games? The importance of tie-breaking rules. *European Journal of Operational Research*, *307*(3), 1260–1269.

Csató, L. (2025). Mitigating the risk of tanking in multi-stage tournaments. *Annals of Operations Research*, *344*(1), 135–151.

Depken, C. A. (1999). Free-agency and the competitiveness of Major League Baseball. *Review of Industrial Organization*, *14*(3), 205–217.

Di Mattia, A., & Krumer, A. (2023). Fewer teams, more games, larger attendance? Evidence from the structural change in basketball's EuroLeague. *European Journal of Operational Research*, *309*(1), 359–370.

Dietl, H. M., Lang, M., & Rathke, A. (2011). The Combined Effect of Salary Restrictions and Revenue Sharing In Sports Leagues. *Economic Inquiry*, *49*(2), 447–463.

Doria, M., & Nalebuff, B. (2021). Measuring competitive balance in sports. *Journal of Quantitative Analysis in Sports*, *17*(1), 29–46.

Douvis, J. (2014). What makes fans attend professional sporting events? A review. *Advances in Sport Management Research Journal*, *1*, 40–70.

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

Feddersen, A., & Maennig, W. (2005). Trends in Competitive Balance: Is There Evidence for Growing Imbalance in Professional Sport Leagues? *SSRN Electronic Journal*.

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

Forrest, D., & Simmons, R. (2002). Outcome uncertainty and attendance demand in sport: the case of English soccer. *Journal of the Royal Statistical Society: Series D (The Statistician)*, *51*(2), 229–241.

Fort, R., & Quirk, J. (1995). Cross-Subsidization, Incentives, and Outcomes in Professional Team Sports Leagues. *Journal of Economic Literature*, *33*(3), 1265–1299.

Galil, Z. (1986). Efficient algorithms for finding maximum matching in graphs. *ACM Computing Surveys (CSUR)*, *18*(1), 23–38.

Gasparetto, T., Mishchenko, D., & Zaitsev, E. (2023). Factors influencing competitive balance across European football top tier leagues. *Managerial and Decision Economics*, *44*(4), 2068–2078.

Gerrard, B., & Kringstad, M. (2022). The multi-dimensionality of competitive balance: evidence from European football. *Sport, Business and Management: An International Journal*, *12*(4), 382–402.

Gyimesi, A. (2024). Competitive balance in the post-2024 Champions League and the European Super League: A simulation study. *Journal of Sports Economics*, *25*(6), 707–734.

Haan, M., Koning, R. H., & Van Witteloostuijn, A. (2007). Competitive balance in national European soccer competitions. In *Statistical thinking in sports* (pp. 75–88). Chapman.

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

Hyun, M., Jones, G. J., Jee, W. (Frank), Jordan, J. S., Du, J., & Lee, Y. (2023). 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.

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

Késenne, S. (2006). The Win Maximization Model Reconsidered. *Journal of Sports Economics*, *7*(4), 416–427.

Lewis, M. (2008). Individual Team Incentives and Managing Competitive Balance in Sports Leagues: An Empirical Analysis of Major League Baseball. *Journal of Marketing Research*, *45*(5), 535–549.

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

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

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

Maxcy, J., & Mondello, M. (2006). The Impact of Free Agency on Competitive Balance in North American Professional Team Sports Leagues. *Journal of Sport Management*, *20*(3), 345–365.

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

Neale, W. C. (1964). 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), 1.

Owen, D. (2014). Measurement of competitive balance and uncertainty of outcome. In *Handbook on the Economics of Professional Football*. Edward Elgar Publishing.

Owen, P. D. (2013). Measurement of competitive balance and uncertainty of outcome. *Handbook on the Economics of Professional Football*, 41–59.

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

Pawlowski, T., Breuer, C., & Hovemann, A. (2010). Top Clubs' Performance and the Competitive Situation in European Domestic Football Competitions. *Journal of Sports Economics*, *11*(2), 186–202.

Pawlowski, T., & Nalbantis, G. (2015). Competition format, championship uncertainty and stadium attendance in European football – a small league perspective. *Applied Economics*, *47*(38), 4128–4139.

Plumley, D., Ramchandani, G. M., & Wilson, R. (2019). 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*(2), 118–133.

Ramchandani, G. (2012). 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.

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

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

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

Rottenberg, S. (1956). The Baseball Players' Labor Market. *Journal of Political Economy*, *1*.

Sanderson, A. R., & Siegfried, J. J. (2003). Thinking about competitive balance. *Journal of Sports Economics*, *4*(4), 255–279.

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

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

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

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

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

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

Szymanski, S., & Késenne, S. (2004). Competitive balance and gate revenue sharing in team sports. *Journal of Industrial Economics*, *52*(1), 165–177.

Totty, E. S., & Owens, M. F. (2011). Salary caps and competitive balance in professional sports leagues. *Journal for Economic Educators*, *11*(2), 45–56.

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

Wang, X. (2025). 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*, 1–16.

Wills, G., Addesa, F., & Tacon, R. (2023). 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.

Wills, G., Tacon, R., & Addesa, F. (2022). 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.

Winfree, J., & Fort, R. (2012). Nash Conjectures and Talent Supply in Sports League Modeling. *Journal of Sports Economics*, *13*(3), 306–313.

Zimbalist, A. S. (2002). Competitive Balance in Sports Leagues: An Introduction. *Journal of Sports Economics*, *3*(2), 111–121.

Appendix

Synthetic Validation

Metric

We validate our metric through simulated tournaments based on the Bradley-Terry model (Bradley and Terry, 1952), with results presented in Figure 7. The figure tracks the evolution of competitive imbalance $\left(\mathcal{V}_t\right)$ 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 $\left[p_1\right] \times n_1 + \left[p_2\right] \times n_2 + \ ...$, 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 algorithm. 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 7, 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 $\left(\mu_t\right)$, 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.

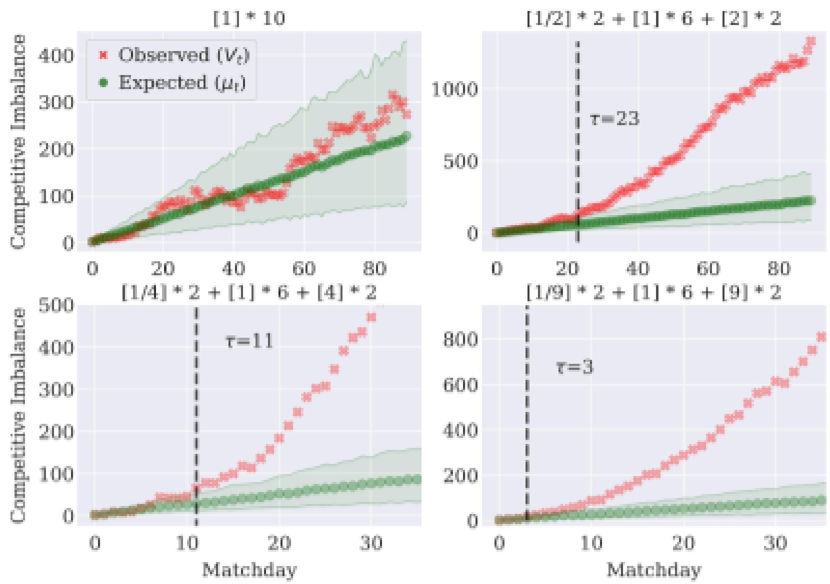

Figure 7. Competitive imbalance divergence in Bradley-Terry simulations.

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. 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.

Controlling

Given a list of team strengths S, we run M double round-robin simulations with random schedules and random results, including home-advantage. Then, for each simulation we reorder the matches following our heuristic schedules: minimizer iMWM schedule; and maximizer iMWM schedule. The results of these experiments are shown in Figure 8, where the title of each image represents the strengths of the best 10 teams (first half of S). The rest can be directly obtained by inverting each value s in the title, i.e., taking $1/s$. When playing at home, the team strength is boosted by 50% to simulate home-advantage. The figure only shows results for this scenario; however, the general behavior is similar for other numbers of teams and home-advantage magnitudes. From left to right the tournaments become progressively more balanced and, as expected, $\tau_{\%}$ increases. This can be seen in the simulations by the positive vertical shift in the boxplots for random tournaments: the closer the team skills, the longer it takes, on average, for tournaments to be distinguishable from balanced ones

More importantly, Figure 8 shows that our heuristics consistently outperform random schedules, particularly in highly imbalanced tournaments (first two), where both maximizer and minimizer significantly shift the perceived competitive balance distribution. Although match reordering only temporarily delays the divergence, our maximizer still raises the median $\tau_{\%}$ by about 20-25%. In more balanced settings, where chance plays a larger role, the effects are subtler but still clear: maximizer pushes the median up, and minimizer drops the median below the random lower quartile. Even in the most balanced case, our methods meaningfully influence the outcome — raising the maximizer median to 90% (15% increase) and reducing the minimizer's to 20% (50% decrease).

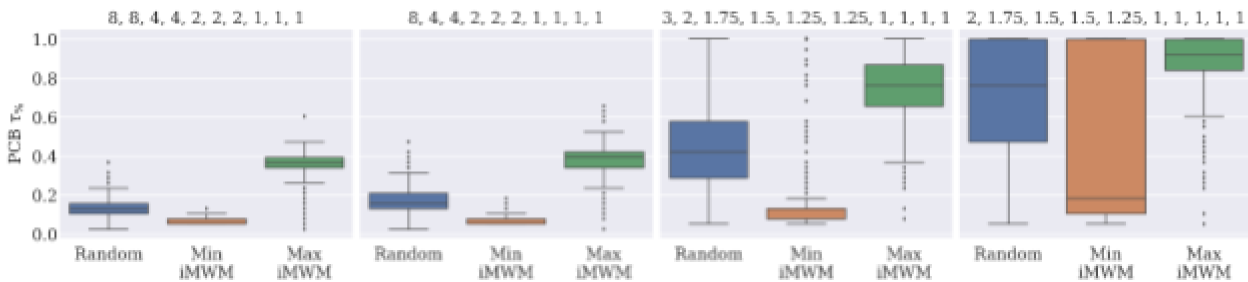

Figure 8. Effect of our iMWM scheduler when compared to random schedules. All tournaments are simulations of 20-team double round-robin competitions using the Bradley–Terry model. The title shows the best 10 teams' strength s, and the other strengths are their inverse: 1/s. The boxes represent the lower and upper quantiles, with the median as the center line. Additionally, the whiskers are drawn to the farthest datapoint within 1.5IQR from the nearest edge.

General Controlling Formulation

While the DRR formalization works perfectly for most double round-robin tournaments, its strictness on the format limits its reach, making it impossible to apply it to similar tournaments such as the regular season of the NBA. One issue is its inflexibility on the number of matches teams can play in a round. Its most noticeable consequence is that the formalization does not work with an odd number of participants since a team will always have a bye. Another problem is the unnecessary restraint on the number of matches between two teams to only once at home and once away.

We propose a few modifications — in Equation 18 — to circumvent these issues and generalize our formalization. We change the third constraint in Equation 15 to an inequality, allowing teams to not play in a round. Furthermore, we add an $n \times n$ matrix ς in the second constraint (Equation 14) to control how many times a team faces each other. The element $ς_{ij}$ represents how many times i faced j as a home team. Critically, this last change adds a new dimension (or index) to the scheduling tensor S. An element S_{tijk} of this new fourth-order tensor indicates in which round t the k-th ordered match (i, j) will occur. As before, $S_{tijk} = 1$ if the game will be placed at round t and 0 otherwise. Since matches can happen a different number of times now, $S_{tijk} = 1$ will also be 0 for any k greater than the number of times the match (i, j) happened in the actual tournament $\left(ς_{ij}\right)$.

Mathematically, we can use matrix $ς_{ij}$ to ensure a team cannot face itself as described in Equation 18. Equations 19 and 20 ensure that each ordered match (i, j) can only be assigned to the schedule S if they exist in the real tournament as well as guarantee that all of them are assigned. Furthermore, as mentioned before, the constraint in Equation 21 has been relaxed, allowing teams to play at most once in each round. Finally, let K be the maximum number of times a team faced another at home in the tournament, that is, $K = max\{ς\}$. Then, the last constraint in Equation 22 forces the tournament to be composed by turns. If the most played ordered match happened K times in the tournament, then K turns are sufficient to cover all matches as long as each match happens at most once in each.

$$\varsigma_{ii} = 0 \qquad \forall\, i \in \{1, ..., n\},\ t \in \{1, ..., T\} \tag{18}$$

$$S_{tijk} = 0 \qquad \forall\,(i,j) \in \{1, ..., n\}^2,\ t \in \{1, ..., T\},\ k \in \left\{\varsigma_{ij} + 1, ..., K\right\} \tag{19}$$

$$\sum_{k}^{K} \sum_{t=1}^{T} S_{tijk} = \varsigma_{ij} \qquad \forall\,(i,j) \in \{1, ..., n\}^2 \tag{20}$$

$$\sum_{k}^{K} \sum_{j=1}^{n} \left(S_{tijk} + S_{tjik}\right) \le 1 \qquad \forall\, i \in \{1, ..., n\},\ t \in \{1, ..., T\} \tag{21}$$

$$\sum_{k}^{K} \sum_{t=1+\frac{(s-1)T}{K}}^{s/TK} \left(S_{tijk} + S_{tjik}\right) \le 1 \qquad \forall\,(i,j) \in \{1, ..., n\}^2,\ s \in \{1, ..., K-1\} \tag{22}$$

Unfortunately, adding flexibility to when matches are played also affects the previous objective function. Allowing teams to not play in a given round means that the optimal schedule S^* would always be achieved by making each match a different, separate round. To account for this, we also change the objective function to maximize the relative number of rounds rather than the absolute one. As an added benefit, this procedure allows the comparison between different leagues and sports since it effectively serves as a normalization over the length of the competition.

$$S^* = argmax\left\{S:\ \tau_{\%}(S)\right\}. \tag{24}$$

All Tournaments Results (iMWM)

Table 2. Summary results for all tournaments.

Oracle?	*Mirrored?*	**Success Rate**	**Avg Increase**	**Avg Decrease**
✔	✔	90.8%	30.4%	9.7%
✔		78.8%	21.5%	11.7%
	✔	65.8%	22.2%	11.1%
		61.6%	18.5%	11.8%