**Assessing Parity in the NBA**

Gop Arop

Dr. Ivan Ramler

**Abstract**

This project aimed to evaluate and understand the level of parity (equality, in team success and talent level) in the NBA by analyzing team data from seasons (1947 - 2023). We analyzed data from the basketball reference.com database using the R Studio application. Our goal was to examine the deviation of point differential and win percentage as key metrics for each NBA season. By studying these variables, we gained insights into how parity between NBA teams has changed over time. The analysis revealed a trend in both deviations, indicating levels of parity. Additionally, we explored the impact of the trade deadline on fairness by investigating if there were any changes in parity (using win percentage deviation) after this period. Interestingly, our analysis supported our hypothesis that parity levels tend to decrease after the trade deadline. Furthermore, we used the Herfindahl Hirschman Index (HHI) to evaluate wins during a season and assess balance before and after the trade deadline. The regular calculation of the HHI posed some challenges when it came to determining the limits of the index as the NBA expanded. To address this, my mentor and I opted to use the normalized HHI, which always scales from 0 to 1. Through this research, we gained insights into the dynamics of competitiveness in the NBA. The project involved a combination of applied data analysis skills, including data wrangling, data visualization, program coding for scraping website data, and applying measurements.

**Introduction**

Examining balance within sports leagues is a complex and multifaceted topic that has received significant attention in relation to the National Basketball Association (NBA). Previous studies have provided insights into team success and talent levels, greatly contributing to our understanding of how competition functions within the NBA. However, there is exploration in existing literature regarding factors such as trade deadlines and salary caps’ impact on team performance and parity. This study aims to bridge this gap by examining how these factors influence balance within the league. Starting with an examination of the context of league competitiveness, the study then concentrates on the specific aspects of the NBA. It considers how league rules, salary caps, and trade deadlines impact team performance. Additionally, it analyzes team performance and competitiveness over time by highlighting the overlooked variability in win percentages around the trade deadline. The study also explores the role of the Herfindahl Hirschman Index (HHI) in assessing parity within the NBA. Employing data analysis techniques such as data wrangling, data visualization, and scraping website data, this research aims to uncover dynamics. By doing so, it offers insights and contributes to our broader understanding of competitiveness in professional sports leagues. The findings do not enrich the literature but also have practical implications for strategies within the NBA.

**Methodology**

Data Collection.

This study aimed to investigate the degree of parity in the National Basketball Association (NBA) across the seasons from 1947 to 2023 by using data from "basketball-reference.com", an online database rich with historical statistics about each NBA team and season (Basketball-Reference). The data was collected using RStudio, where three CSV files, 'Team Summaries.csv', 'Opponent Totals.csv', and 'Team Totals.csv' from the database were imported. These files, which held comprehensive team performance summaries, opponent totals, and team totals respectively, were cleaned by eliminating incomplete rows and filtering out data unrelated to the NBA, such as entries for the American Basketball Association (ABA) and league averages. The cleaned 'Opponent Totals.csv' and 'Team Totals.csv' data were then merged into one comprehensive dataset, with the 'opp\_pts' column from the 'opponent\_totals' data frame joined to the 'team\_totals' data frame, enabling the analysis of points scored by opponents within the context of overall team totals. The subsequent data analysis involved the calculation of standard deviations for the total point differential and win percentage for each season, helping identify trends in NBA parity levels over the examined time span.

Standard Deviation of Average Win Percentage.

Further manipulation of the 'team\_summary' data frame was undertaken to facilitate the analysis of the standard deviation of win percentage across different NBA seasons. A new variable 'win\_pct' was created to represent the winning percentage of each team. This was calculated by dividing the number of wins by the total games played, obtained by adding wins and losses. Afterward, the 'team\_summary' data frame was grouped by seasons using the 'group\_by()' function. For each season, two new variables were calculated: 'pct\_over\_500', which represented the percentage of teams with a win percentage over 0.500, and 'sd\_win\_pct', which was the standard deviation of the win percentage for teams in that particular season. The latter variable was instrumental in the analysis of variability in team performance across seasons. The final step in this code snippet involved visualizing the variability of win percentage across seasons. The 'ggplot()' function was employed to create a scatter plot with seasons on the x-axis and the standard deviation of the win percentage on the y-axis. Each point on the graph represented a season, colored blue for visibility. To highlight trends in the data, a smoothed line was added to the plot using the 'stat\_smooth()' function. This line represented the trend in the standard deviation of win percentage across seasons, illustrating how team performance variability evolved over time. The plot was styled with black and white themes using the 'theme\_bw()' function, making it visually appealing and easy to interpret.

Standard Deviation of Average Point Differential.

To investigate the variability in the average point differential across different NBA seasons, further manipulation and calculation on the 'team\_totals' data frame were performed. Initially, the data frame was grouped by team and season, and for each team in each season, a new variable 'pts\_diff' was calculated representing the total points scored by a team minus the total points scored against them. This point differential was then averaged per game to obtain the average point differential ('avg\_pt\_diff'), computed by dividing the point differential by the number of games played. The 'avg\_pt\_diff' was subsequently added to the 'team\_pt\_diff' data frame. For the 2023 season, the mean average point differential was computed and printed to the console, providing a single summary measure of point differential for that season. Next, a new data frame 'standard\_dev\_pt' was created, where the standard deviation and mean of the average point differential were calculated for each season. The standard deviation, 'sd\_pt\_diff', was of particular interest as it represented the variability in the average point differential across teams for each season. To visualize this variability, a scatter plot was created with seasons on the x-axis and the standard deviation of the point differential on the y-axis. Each point on the graph, colored red for clarity, represented a season. Furthermore, specific seasons were highlighted and labeled on the plot using the 'annotate()' function. A smoothed line, colored blue, was also added to the plot to highlight trends in the data over time. Finally, the plot was saved as a JPEG image, and a new variable 'period' was created in the 'standard\_dev\_pt' data frame, representing the decade for each season, computed by integer division of the number of seasons since 1947 by 10. This allowed for examination of the standard deviation of average point differential on a decade-by-decade basis.

Difference In Variability of Average Win Percentage After Trade Deadline.

To examine the effects of the NBA trade deadline on team parity, the team standings before and after the trade deadline were scraped from basketball-reference.com for seasons from 1947 to 2023. For seasons from 1971 to 2023, the trade deadline was estimated to be February 1st, with exceptions for the 1999, 2012, and 2021 seasons due to special circumstances. The standings data were gathered in 'all\_midseason\_standings'. For seasons from 1947 to 1970, the trade deadline generally fell on February 1st or 10th, but earlier dates were used for the 1947 to 1949 seasons when the league was still the Basketball Association of America (BAA). This data was gathered in 'all\_midseason\_standings2'. After the data was collected, the CSV files 'midseason\_standings\_1971-2023.csv' and 'midseason\_standings\_1947-1970.csv' were created to store this data for future use. Next, the standard deviation of the win percentage at midseason ('sd\_win\_pct\_ms') and the percentage of teams with a win percentage over 0.500 ('pct\_over\_500\_ms') were calculated for each season using the 'final\_midseason\_standings' data frame. The results were stored in 'midseason\_over\_500'. Following this, the difference in the standard deviation of the win percentage from midseason to the end of the season ('diff\_sd\_win\_pct') was calculated for each season. This data was visualized in a scatter plot, where each point represented a season, and a smoothed line was added to highlight trends over time. Lastly, an approximate line representing the expected change in the standard deviation of win percentage due to the number of games played was added to the plot as a dashed line, providing a reference for interpreting the observed changes.

Calculating Normalized HHI Values.

The code's first section begins by establishing a 'conference' variable within the 'team\_summary' data frame, designating NBA teams into the 'East' or 'West' conference. The allocation considers historical shifts where teams transitioned from one conference to another during certain seasons. The 'ifelse()' function checks for the team's name and the corresponding season to decide the conference allocation. This prepared dataset, 'teams\_split', supports a deeper conference-level analysis. In the subsequent section, the code calculates the HHI, a popular measure of market competitiveness, to evaluate the parity within the East and West conferences for each season. The HHI essentially quantifies the degree of dominance by a few teams within a conference; a high HHI score suggests a more significant dominance by a few teams, thereby indicating a lesser competitive balance. Conversely, a lower HHI score implies a more even distribution of wins amongst teams, indicating a higher competitive balance. The 'calculate\_normalized\_hhi()' function calculates the HHI for each conference, which is subsequently normalized within each season to account for the expansion of the league over time. Normalizing the HHI, which involves dividing the HHI by the maximum possible HHI score, provides a standardized measure allowing for comparison across different time periods or settings. The normalized HHI, hence, ranges from 0 to 1, with 0 representing perfect parity and 1 indicating total dominance by a single team. The function was applied to the 'teams\_split' data frame, generating the 'normalized\_hhi' data frame containing the normalized HHI for each conference across all seasons. Finally, the section culminates with the creation of a line plot illustrating the normalized HHI for each conference across seasons, with the x-axis representing the 'season' and the y-axis denoting the 'Normalized\_HHI'. Each line on the graph symbolizes a conference, distinguished by color-coding. This minimalistic and clear plot facilitates an easy understanding of competitiveness trends within each conference over time, as indicated by the normalized HHI.

**Results**

Results of the Standard Deviation of Average Win Percentage.

The analysis of the standard deviation of win percentage across 77 NBA seasons revealed a cyclical pattern, indicative of parity shifts. The mean of this trend hovered around 0.15, suggesting a standard variability in most seasons. Notably, before 1977, when the NBA consisted of approximately 12 teams on average, outliers were more common due to the smaller sample size.

A graph showing the number of seasons

Description automatically generated

Fig. 1. Visual graph of the standard deviation of average win percentage for each NBA season.

Results of Standard Deviation of Average Point Differential.

The standard deviation of the average point differential over 77 NBA seasons presents a cyclical pattern akin to the previously discussed win percentage plot (Fig. 1), suggesting a correlation between these two aspects and confirming our suggestions of league-wide parity depending on the occurrence of certain events. The trend centers around a five-point mark, demonstrating a consistent point difference among the teams. Influential points on the plot, detailed in Table 1, correspond to notable NBA historical events, highlighting the impact of such changes on team performance.

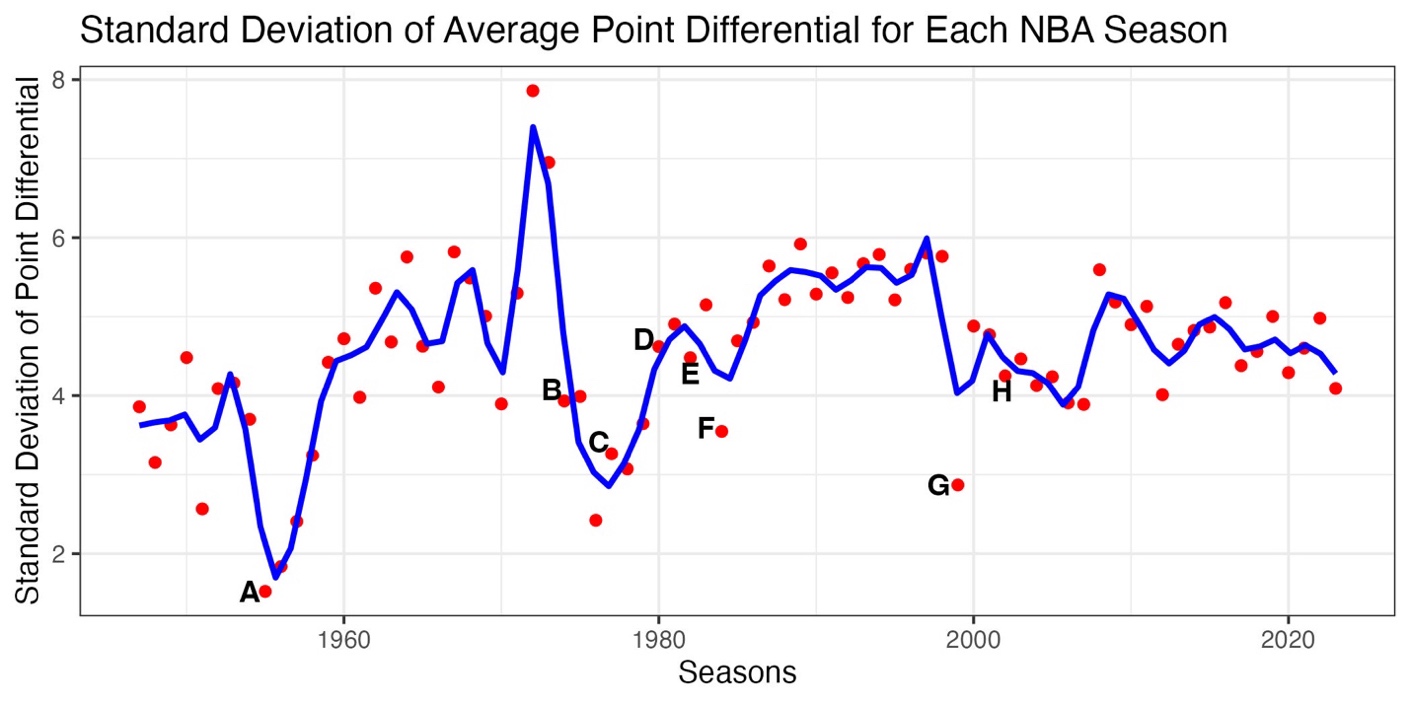
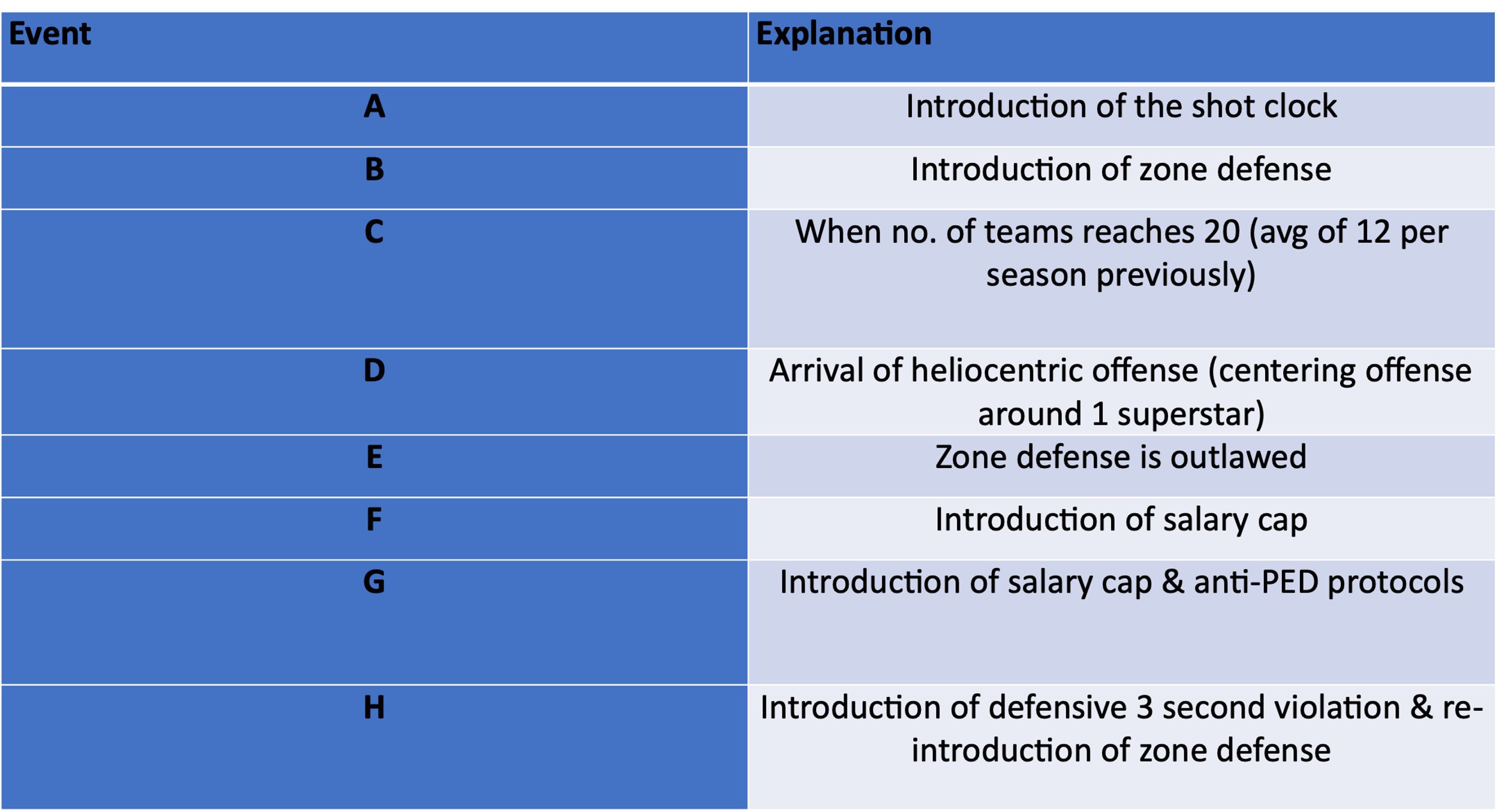


Fig. 2. Visual graph of standard deviations of average point differential for each NBA season.

Table 1. Explanation of all pinpointed events that occurred in Fig. 2.



Results For the Difference in Variability of Win Percentage After Trade Deadline.

The analysis of the trade deadline's effect on the variability of average win percentage required an adjustment to the y-axis center point, due to the uneven number of games played mid-season and at the season's end, setting the approximated line at -0.0185052150. An examination of the plot revealed 66 of the 77 seasons above this line, suggesting increased win percentage variability post-trade deadline. This implies that the trade deadline potentially influences significant shifts in team performance, thus impacting the NBA's competitive dynamics.

A graph showing the growth of the price of a trade

Description automatically generated with medium confidence

Fig. 3. Visual graph of the difference in variability of win percentage at the middle of the season (before trade deadline) & at the end of the season (after trade deadline).

Results of Normalized HHI Values for Each Conference

The analysis of the normalized Herfindahl-Hirschman Index (HHI) values revealed an evenly distributed parity within the NBA. Since 1971, exactly half of the entries for both the Western and Eastern conferences registered below the average threshold of 0.50 for normalized HHI values (scaled from 0 to 1).

A graph showing the number of events

Description automatically generated with medium confidence

Fig. 4. Visual graph of normalized HHI values from 1971- 2023 for Western and Eastern conference.

**Conclusion**

This research investigated the degree of parity in the NBA over seven decades, revealing cyclical patterns in the standard deviation of win percentage and total point differential. These patterns corroborate our thesis that the NBA's competitive balance is influenced by various events and rules, including the implementation of salary caps and trades. In fact, the salary cap appears to be a crucial factor in maintaining league-wide balance.

Our analysis further identified a significant shift in team performance following the trade deadline, suggesting this deadline potentially triggers notable changes in the NBA's competitive dynamics. Moreover, the normalized Herfindahl-Hirschman Index (HHI) values indicate a consistent distribution of parity within the NBA, reinforcing our findings.

While this study provides valuable insights, it is not without limitations. Further advanced statistical tests and comparison with other sports leagues' data are necessary for a more comprehensive understanding of parity trends. The quest to pinpoint an "ideal" time of parity in NBA history remains open for future research.

In summary, this study highlights the influence of various factors in shaping the NBA's competitiveness. It serves as a solid foundation for further explorations into the competitive balance in professional sports leagues, paving the way for more in-depth studies on the ideal competitive environment.

**References:**

* Basketball Reference. “NBA & ABA Team Index | Basketball-Reference.com.” *Basketball-Reference.com*, 2019, www.basketball-reference.com/teams/.
* Depken, Craig A. “Free-Agency and the Competitiveness of Major-League Baseball.” *SSRN Electronic Journal*, 1996, https://doi.org/10.2139/ssrn.1711.
* “NBA Stats (1947-Present).” *Www.kaggle.com*, www.kaggle.com/datasets/sumitrodatta/nba-aba-baa-stats.
* Owen, P. Dorian, et al. “Measuring Competitive Balance in Professional Team Sports Using the Herfindahl-Hirschman Index.” *Review of Industrial Organization*, vol. 31, no. 4, Dec. 2007, pp. 289–302, <https://doi.org/10.1007/s11151-008-9157-0>.
* Sorensen, Mike. “NBA Rules Have Adapted over the Years to Make the Game More Fun for Players, Fans.” *Deseret News*, 21 Feb. 2019, www.deseret.com/2019/2/21/20666425/nba-rules-have-adapted-over-the-years-to-make-the-game-more-fun-for-players-fans.
* “NBA Trade Deadline History - RealGM.” Basketball.realgm.com, basketball.realgm.com/nba/transactions/trade-deadline. Accessed 15 Aug. 2023.