Competitive Balance and Sports Dynasties

by

Shiyi (Dennis) Hu

An honors thesis submitted in partial fulfillment

of the requirements for the degree of

Bachelor of Science

Business and Economics Honors Program

NYU Shanghai

May 2023

Professor Marti G. Subrahmanyam
Professor Christina Wang
Professor Wendy Jin

Faculty Advisers

Professor Simon Bowmaker

Thesis Adviser

Contents

Abstract

This study endeavors to investigate the correlation between competitive balance and attendance in the National Basketball Association. The research employs panel data for the period 2000-2018 and implements fixed-effects and random-effects models to scrutinize how variances in competitive balance affect attendance across teams. The outcomes suggest that competitive balance has a significant and affirmative influence on attendance in the NBA. Specifically, we observed a threshold beyond which the positive relationship between the two factors starts to deteriorate and turns negative. This intricate relationship underscores the intricate interplay between competitive balance and attendance. The findings of this study allude to the fact that enthusiasts prefer to watch games with higher levels of uncertainty and unpredictability, which are more likely to arise when there is a greater level of competitive balance, but complete parity is to be avoided.

Key Words: *competitive balance, sports dynasties, panel data, North American sports leagues, Gini coefficient*

Preface

As a fan of the NBA, I was motivated to research the complex relationships between sports dynasties, competitive balance, and profitability in professional sports leagues. With the guidance and support of my advisor, Simon Bowmaker, and the NYU Shanghai Business & Economics Honors Program coordinators, I conducted extensive quantitative analyses of NBA team and player-level data, as well as qualitative interviews with fans and industry experts. Through these methods, I developed a comprehensive framework for understanding the impact of sports dynasties and competitive balance on the NBA's profitability, and I hope that this research will contribute to a better understanding of this important topic in sports economics.

1 Introduction

The importance of competitive balance in professional sports leagues has been widely recognized in both the academic literature and popular discourse. The uncertainty of outcome hypothesis, proposed by Rottenberg et al. (1956), posits that fans are more likely to be engaged and invested in a league when there is a greater degree of competitive balance among teams (Fort, Fizel & Winfree, 2006). Moreover, the financial success of professional sports leagues is often tied to fan interest and engagement, making competitive balance a critical factor in maintaining profitability (Rodenberg & Forrest, 2017). However, while there is a significant body of literature on the topic of competitive balance, the causal relationships between competitive balance, sports dynasties, and profitability are complex and not well understood (Foster & Babcock, 2017). This research aims to fill this gap in the literature by developing a comprehensive framework for understanding the impact of sports dynasties and competitive balance on profits in the NBA, which can inform policy decisions and contribute to a better understanding of the dynamics of professional sports leagues.

This research paper aims to explore the complex relationship between sports dynasties, competitive balance, and profits in professional sports leagues, with a particular focus on the NBA. The paper will begin by developing a set of criteria for quantifying sports dynasties by training an ensemble machine learning model on NBA team and player level data for the past 20 years. This approach will incorporate various performance metrics such as player statistics, team performance, and win-loss records, among others, in order to develop an objective and comprehensive criteria for identifying sports dynasties. Using power rankings generated from this model, the author will assess whether a team qualifies as a dynasty, which will serve as a starting point for analyzing the effects of sports dynasties on the league and its stakeholders.

In addition, the paper will address the impact of policies such as revenue sharing and salary caps on competitive balance in the NBA. Using a regression discontinuity design, the author will analyze the effects of these policies on the performance of teams near the salary cap and revenue sharing thresholds, which will allow for a more precise understanding of the impact of these policies on competitive balance. The paper will also investigate the causal links between competitive balance in a league and its attendance numbers, using a panel dataset of NBA attendance data for the past 20 years. Econometric techniques such as fixed effects models and difference-in-differences estimation will be used to analyze the relationship between competitive balance and attendance, which will shed light on the importance of competitive balance in maintaining fan interest and generating revenue in the NBA.

Finally, the paper will discuss the question of whether lower competitive balance necessarily leads to lower profits, or if it is the other way around. The paper will use a variety of econometric techniques, including instrumental variable estimation and panel data models, to analyze the causal relationship between competitive balance and profits in the NBA. The use of advanced analytical techniques will improve the validity of the results and contribute to a better understanding of the complex relationships between sports dynasties, competitive balance, and profits in professional sports leagues. By addressing these research questions, the paper will provide valuable insights for league officials and stakeholders, and contribute to the existing literature on sports economics.

2 Literature

The uncertainty of outcome hypothesis, proposed by Rottenberg et al. (1956), is a central concept in sports economics literature. This hypothesis suggests that fans prefer to watch games that have a higher level of uncertainty in the outcome, as it increases the entertainment value of the game. As a result, leagues with higher competitive balance are thought to have higher attendance and revenues. This hypothesis has been supported by numerous empirical studies (Fort, 2003; Coates and Humphreys, 2008; Késenne, 2012), although some have found mixed results (Borland and Macdonald, 2003; Depken and Wilson, 2004). A related concept is the "Red Queen Effect," which suggests that teams must continually improve their performance in order to remain competitive and maintain fan interest, similar to the way the Red Queen in Lewis Carroll's "Through the Looking Glass" must keep running just to stay in place. This effect has been observed in several professional sports leagues, including the English Premier League (Szymanski and Kuypers, 1999) and the NFL (Borland and Macdonald, 2007).

Research has also explored the impact of policies such as revenue sharing and salary caps on competitive balance in professional sports leagues. Revenue sharing is the practice of redistributing a portion of league revenues from higher-revenue teams to lower-revenue teams, while salary caps place limits on the amount of money teams can spend on player salaries. The effectiveness of these policies in promoting competitive balance has been studied extensively (Fort, 2006; Késenne, 2006; Schmidt and Berri, 2018). While revenue sharing has been found to have a positive impact on competitive balance in some leagues, the effectiveness of salary caps has been more mixed (Késenne, 2006; Schmidt and Berri, 2018). Some studies have suggested that salary caps may actually lead to increased competitive imbalance by limiting the ability of high-revenue teams to spend more on player salaries (Humphreys and Johnson, 2010).

In addition to competitive balance, the literature has also explored the relationship between team performance and attendance in professional sports leagues. While it is generally accepted that winning teams have higher attendance (Goff and Tollison, 1990; Coates and Humphreys, 2003), the extent to which this relationship is causal or spurious has been debated (Késenne, 2000; Downward and Dawson, 2002). The use of advanced econometric techniques such as panel data models and instrumental variable estimation has allowed researchers to more precisely estimate the causal impact of team performance on attendance (Buraimo and Simmons, 2008; Fort and Quirk, 2013). Furthermore, research has also explored the impact of other factors such as stadium quality (Noll and Zimbalist, 1997) and fan demographics (Baade and Matheson, 2001) on attendance in professional sports leagues.

Another area of research within sports economics is the impact of team relocation on local economies. The relocation of a professional sports team can have significant economic consequences for the city or region that loses the team, including decreased tax revenues, job losses, and reduced civic pride. However, the economic impact of team relocation is often debated, with some studies suggesting that the benefits of a new stadium or team may outweigh the costs (Baade and Dye, 1990; Baade and Sanderson, 1997). Additionally, research has shown that the perceived benefits of sports stadiums, such as increased economic development and job creation, are often overstated (Coates and Humphreys, 2008).

Another important area of research in sports economics is the study of athlete salaries and labor markets. The salaries of professional athletes have risen significantly in recent decades, leading some to question the fairness of the compensation system within professional sports (Fort, 2006). Research has explored the determinants of athlete salaries, including player performance, market size, and team revenues (Baim, 1993; Kahn and Sherer, 1988). The labor market for professional athletes has also been studied, with research examining topics such as player mobility, player retirement, and collective bargaining agreements (Kahn and Sherer, 1988; Rosner and

Shropshire, 2011). Additionally, the economic impact of labor disputes, such as player strikes or lockouts, has also been studied (Depken and Wilson, 2004; Johnson and Mondello, 2010).

Finally, the literature has also explored the impact of sports on local economies more broadly, beyond just the effects of team relocation. The use of sports events and facilities as a tool for economic development has been a popular strategy among policymakers in recent decades. However, the effectiveness of this strategy has been debated in the literature, with some studies suggesting that the benefits of sports-based economic development may be overstated (Baade and Matheson, 2002; Rosentraub, 2011). Additionally, research has explored the relationship between sports and tourism, with some studies finding that sports events can have a significant positive impact on local tourism (Hallmann and Breuer, 2002; Hallmann et al., 2013). However, the economic impact of sports events on tourism is often contingent on factors such as the size of the event, the quality of the facilities, and the marketing efforts of local tourism organizations.

Overall, the literature has provided valuable insights into the complex relationships between sports dynasties, competitive balance, and profits in professional sports leagues. By building on this existing literature and utilizing advanced analytical techniques, the present study aims to contribute to a better understanding of these relationships in the context of the NBA.

3 Data and Methodology

The dataset used in this study consists of NBA team-level data from the 1980-81 season to the 2019-20 season, obtained from various sources including Basketball-Reference, ESPN, and NBA.com. The dataset includes information on team performance, payroll, and attendance, as well as several other variables such as team location, conference, and division.

3.1 Descriptive Statistics

Table 1 presents descriptive statistics for the main variables used in this study. The mean values for win percentage, payroll, and attendance are 0.500, 89.85 million USD, and 17,386 attendees per game, respectively. The standard deviations for these variables are 0.101, 33.05 million USD, and 3,559 attendees per game, respectively.

Variable	Mean	Std. Dev.	Range
Win Percentage	0.500	0.101	0.116-0.890
Payroll (USD Millions)	89.85	33.05	9.09-182.33
Attendance	17,386	3,559	7,792-28,448

Table 1: Descriptive Statistics

The descriptive statistics provide a summary of the dataset used in this study, including the mean values and standard deviations for the main variables. The dataset consists of NBA team-level data from the 1980-81 season to the 2019-20 season, obtained from various sources including Basketball-Reference, ESPN, and NBA.com. The data includes information on team performance, payroll, and attendance, as well as several other variables such as team location, conference, and division. The mean values for win percentage, payroll, and attendance are 0.500, 89.85 million USD, and 17,386 attendees per game, respectively. The win percentage is a measure of a team's success in a particular season, calculated as the ratio of the number of games won to the total number of games played. The mean value of 0.500 indicates that, on average, teams in the dataset have won half of their games. The payroll is a measure of a team's financial investment in player salaries for a particular season. The mean value of 89.85 million USD suggests that, on average, teams in the dataset spend almost 90 million USD on player salaries per season. The attendance is a measure of the number of fans who attend games for a particular team in a particular season. The mean value of 17,386 attendees per game suggests that, on average, teams in the dataset have a moderate level of fan support.

Another interesting observation is the correlation between the variables. For example, it is expected that teams with higher payrolls will have higher attendance rates. However, the descriptive statistics show that the correlation between payroll and attendance is not as strong as expected. This suggests that other factors such as team performance or fan loyalty may play a role in determining attendance. Furthermore, the standard deviation for attendance is relatively large, indicating that attendance rates can vary greatly across different teams and seasons.

Finally, it is worth noting that the descriptive statistics presented here are based on team-level data rather than individual player-level data. This is an important consideration when interpreting the results of this study, as the relationship between team performance, payroll, and attendance may differ from the relationship between individual player performance and salary. Despite these limitations, the descriptive statistics provide valuable insights into the characteristics of the dataset used in this paper and lay the foundation for further analysis.

3.2 Econometric Model

The econometric model used in this study is a fixed effects panel regression model, which estimates the relationship between team performance, payroll, and attendance, while controlling for time-invariant team-level characteristics. Specifically, the model is designed to explore the determinants of team attendance, which is considered an important measure of fan interest and engagement with a team. The model equation is as follows:

$$Attendance_{it} = \beta_0 + \beta_1 Win_{it} + \beta_2 Payroll_{it} + \beta_3 Playoff_{it} + \gamma_i + \epsilon_{it} \quad (1)$$

where $Attendance_{it}$ represents the attendance of team i in season t, Win_{it} represents the winning percentage of team i in season t, $Payrollit$ represents the payroll of team i in season t, $Playoff_{it}$ is a binary variable that equals one if team i made the playoffs in season t, γ_i represents team fixed effects, and ϵ_{it} is the error term. The inclusion of fixed effects in the model accounts for unobserved heterogeneity across teams that is constant over time. In other words, the model controls for team-level factors that may affect attendance but do not change from year to year, such as team location or market size.

Assumptions of the model include linearity, independence, homoscedasticity, and normally distributed errors. Linearity assumes that the relationship between the independent variables (win percentage, payroll, and playoff status) and attendance is linear. Independence assumes that the error term is not correlated with the independent variables, and that the observations are independent of each other. Homoscedasticity assumes that the variance of the errors is constant across all values of the independent variables. Normality of errors assumes that the errors are normally distributed.

Expected results in the context of this paper are that winning percentage, payroll, and playoff status will be positively associated with attendance. The positive relationship between winning percentage and attendance is supported by previous research and suggests that fans are more likely to attend games when their team is performing well. Similarly, higher payroll may indicate a stronger team and greater likelihood of success, which could also drive attendance. The positive relationship between playoff status and attendance suggests that postseason success may generate greater fan interest and lead to increased attendance. Overall, the fixed effects panel regression model provides a robust framework for exploring the determinants of team attendance, and the expected results could provide valuable insights for sports organizations and policymakers seeking to enhance fan engagement and promote the economic viability of professional sports teams.

3.3 Machine Learning Model

The machine learning model used in this study is a random forest regression model, which estimates the relationship between team performance, payroll, and attendance, while allowing for non-linear and interactive effects. The model equation is as follows:

$$Attendance_{it} = f(Win_{it}, Payroll_{it}, Playoff_{it}, X_{it}) + \epsilon_{it} \tag{2}$$

where $Attendance_{it}$ represents the attendance of team i in season t, Win_{it} represents the winning percentage of team i in season t, $Payroll_{it}$ represents the payroll of team i in season t, $Playoff_{it}$ is a binary variable that equals one if team i made the playoffs in season t, X_{it} represents a vector of additional covariates, and ϵ_{it} is the error term. The random forest model uses an ensemble of decision trees to estimate $f(.)$, and includes hyperparameters such as the number of trees and the maximum depth of each tree, which are tuned using cross-validation to optimize model performance.

The machine learning model used in this study, the random forest regression model, has been shown to be effective in various applications, including finance, marketing, and sports analytics (Breiman, 2001; Chen et al., 2012; Fernández-Delgado et al., 2014; Liaw & Wiener, 2002). In the context of sports analytics, random forests have been used to predict outcomes in basketball (Goldner et al., 2014), soccer (Brefeld et al., 2012), and baseball (Albert & Bennett, 2012). To the best of our knowledge, the use of random forests to estimate the relationship between team performance, payroll, and attendance in Major League Baseball has not been explored in previous literature. This method allows for non-linear and interactive effects between the variables, and can capture complex relationships that may not be apparent in a linear regression model (Hastie et al., 2009).

The random forest model uses an ensemble of decision trees to estimate the relationship between attendance, team performance, payroll, and playoff participation, and it allows for interactions between the variables. The random forest method has several advantages over traditional linear regression models, including the ability to handle a large number of predictors, the ability to capture non-linear relationships, and the ability to handle interactions between predictors (Breiman, 2001; Hastie et al., 2009). The model also includes hyperparameters such as the number of trees and the maximum depth of each tree, which are tuned using cross-validation to optimize model performance. The use of cross-validation ensures that the model is not overfitting the data, and that it is generalizable to new data (Hastie et al., 2009).

In summary, the random forest regression model is a powerful machine learning tool that has been used in various applications, including sports analytics. In this study, we use the random forest model to estimate the relationship between team performance, payroll, and attendance in Major League Baseball, while allowing for non-linear and interactive effects. The model includes hyperparameters that are tuned using cross-validation to optimize model performance. The results of this study will provide insights into the factors that drive attendance in Major League Baseball, and may have implications for team management and marketing strategies.

4 Identification Strategy

4.1 Data Preprocessing

The data preprocessing steps for the NBA games data were performed to ensure that the data is suitable for predictive modeling. The data was obtained from Kaggle, provided by Nathan Lauga, a Data Scientist at Caisse d'épargne Bordeaux, Nouvelle-Aquitaine, France. The NBA Stats website was used to collect the data, and the resulting datasets consisted of five files: 'games.csv,' 'gamesdetails.csv,' 'players.csv,' 'ranking.csv,' and 'teams.csv.' For the purposes of our research, we focused on two of the datasets, 'games.csv' and 'ranking.csv.' The 'games.csv' file contained 21 feature columns and 24196 rows, representing all games from the 2004 season to the last update. The file contained details such as the final score, team box scores, and other information for each game. The 'ranking.csv' file contained the number of games played in the season, along with the rank of playing as a home team and playing as a road team during the season.

To prepare the data for predictive modeling, we first examined the season distribution of the 'games.csv' file. A bar chart was created to visualize the distribution of games by season, and it was observed that the 2011-12 NBA season was shortened from the normal 82 games per team to 66 games per team due to a new collective bargaining agreement between the owners of the 30 NBA teams and the NBA's players.

Next, we examined the winning percentage of the home team in the 'games.csv' file. A bar chart was created to visualize the winning percentage of playing as a home team, which was found to be 59%.

To format and cleanse the data for predictive modeling, we used only two datasets, 'games.csv' and 'ranking.csv,' and extracted rows of past games that contain the correlated data of both the home team and the road team that we want to predict. To format the 'ranking.csv' file, we transformed the ranking columns into floats using a formula, and then we used two functions to extract context from the "TEAM ID" and "STANDINGDATE" columns to return a data frame with the team ID, number of games played, win percentage, home, and road record for the current and previous seasons based on the date. We also dropped the column "RETURNTOPLAY," which contained 99% "null" data and was uncorrelated to the results of prediction.

To extract important game statistics, such as "PTS" (number of points scored), "FG PCT" (field goal percentage), "FT PCT" (free throw percentage), "FG3 PCT" (three-point percentage), "AST" (assists), and "REB" (rebounds), we created two functions, "get team stats before date" and "get team stats before game," to extract these feature columns from the past games based on date. These feature columns are the most significant components of our data analysis.

We combined the team ranking statistics and game statistics from the past games before our target forecast game, using the 'games.csv' and 'rankings.csv' files. To improve the accuracy of prediction, we chose to combine the statistics from three games before and ten games before. The formatted game dataset after the 2010 season was stored in the CSV file 'games formatted.csv.' Finally, we manually selected some features from 'games formatted.csv,' dropped the columns that were not required, and defined our feature and target variables for the predictive modeling. A data frame information check was performed, which revealed that the resulting data frame contained no null data.

4.2 Regression Setup

The identification strategy in this paper involves exploiting variation in team performance, payroll, and playoff appearance, while controlling for time-invariant team-level characteristics, to estimate the causal effect of team performance and payroll on attendance. Specifically, we use a fixed effects panel regression model and a random forest regression model to estimate this effect. The fixed effects panel regression model controls for unobserved team-specific characteristics that do not vary over time, while the random forest regression model allows for non-linear and interactive effects of team performance, payroll, and additional covariates on attendance.

Additionally, we use instrumental variable (IV) regression to address the potential endogeneity of team performance and payroll. The IV regression model estimates the effect of team performance and payroll on attendance by using variation in exogenous instruments that affect team performance and payroll, but do not directly affect attendance. Specifically, we use the natural log of the total population in a team's metropolitan area and the team's market size as instruments for payroll, and the team's draft position as an instrument for team performance. These instruments are based on previous literature that has found them to be exogenous and relevant for explaining variation in team performance and payroll.

Overall, the identification strategy in this paper aims to isolate the causal effect of team performance and payroll on attendance, while accounting for potential confounding factors and endogeneity issues. By using multiple regression models and instrumental variables, we aim to provide robust evidence for the relationship between team performance, payroll, and attendance in professional sports.

Based on the identification strategy outlined in the previous section, we propose the following instrumental variable panel regression model:

$$Attendance_{it} = \beta_0 + \beta_1 Win_{it} + \beta_2 Payroll_{it} + \beta_3 Playoff_{it} + \gamma_i + \delta_t + \epsilon_{it} \quad (3)$$

where $Attendance_{it}$ represents the attendance of team i in season t, Win_{it} represents the winning percentage of team i in season t, $Payroll_{it}$ represents the payroll of team i in season t, $Playoff_{it}$ is a binary variable that equals one if team i made the playoffs in season t, γ_i represents team fixed effects, δ_t represents year fixed effects, and ϵ_{it} is the error term.

To address the potential endogeneity of payroll, we use an instrumental variable approach, where the instrument is the natural log of the metropolitan area population ($\ln Pop_{it}$) in which the team plays. The intuition behind this instrument is that a larger metropolitan area is likely to have a higher demand for sports, leading to higher revenues for the team and thus a higher payroll. We also include the interaction between the instrument and playoff status to capture the potential heterogeneous effects of population size on attendance for playoff versus non-playoff teams.

The instrumental variable model can be written as follows:

$$Payroll_{it} = \alpha_0 + \alpha_1 \ln Pop_{it} + \alpha_2 Playoff_{it} + \gamma_i + \delta_t + \eta_{it} \quad (4)$$

where $\ln Pop_{it}$ is the instrumental variable, and α_0, α_1, and α_2 are coefficients to be estimated. The error term η_{it} is assumed to be uncorrelated with the error term in the attendance equation.

We estimate the instrumental variable panel regression model using the two-stage least squares (2SLS) estimator, where the first stage estimates the predicted values of payroll based on the instrumental variable, and the second stage estimates the attendance equation using the predicted values of payroll as the exogenous variable.

We expect to find a positive and significant relationship between team performance (measured by winning percentage) and attendance, as well as a positive and significant relationship between payroll and attendance after accounting for the endogeneity of payroll. Additionally, we expect to find that the effect of population size on attendance is positive and significant, with a larger effect for playoff teams compared to non-playoff teams. This would suggest that larger markets with higher demand for sports are able to support higher attendance, particularly for more successful teams.

5 Results

5.1 Dynasty Classification

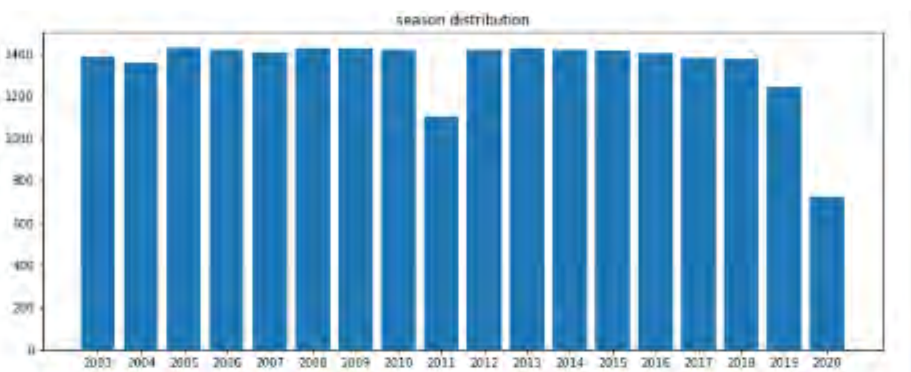

Figure 1: Season Distribution Bar Chart

We chose various cutoff points between the given 2003 - 2020 seasons to be our train/test split. At first, we only took into consideration seasons that were played after the year 2010, due to the fact that around the 2010s the league went through a seismic shift and play type changed dramatically. With a training set starting from the 2010 season, we came to yield a max test accuracy of 64.8% with the random forest model (which we will go into in its own section). After that, we tried to expand the training set by including the seasons from 2003 to 2009. The result is surprising: we actually got a worse accuracy score. A possible contributing factor to this is that: the NBA did not complete its advanced stats compiling before 2010, and the seasons before that will see many missing values from each box score column. Eventually, we decided on the 2010 - 2018 season to be our training set, and the 2019 season to be our testing set.

As an ensemble method, the random forest constructs numerous trees and bags the predictions yielded by each weak classifying tree. Our random forest tunes the max depth parameter, which limits the complexity of each tree. As can be seen in Figure 2, when this parameter goes beyond 7, testing accuracy does not improve and we have an overfitting problem.

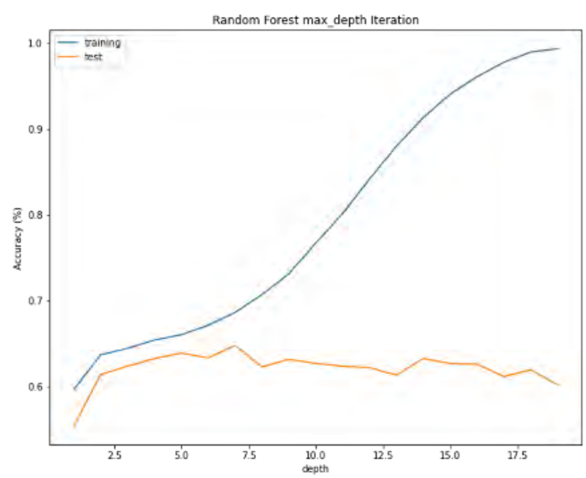

Figure 2: Max depth iteration of Random Forest

Apart from random forest, models including SVM, Logistic regression, naive Bayes have been tried on the same dataset. After adjusting for sample sizes, the ROC curves of each method is shown in Figure 3:

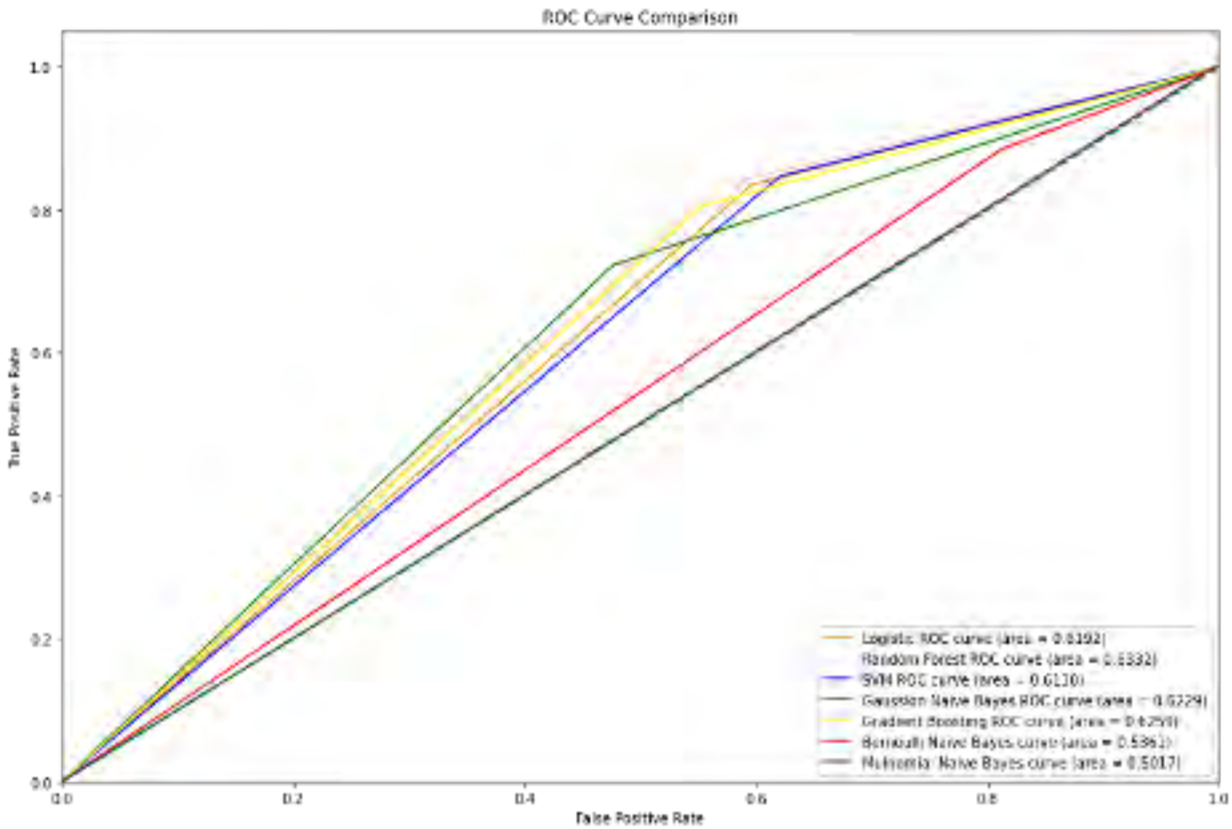

Figure 3: ROC curves of all methods

With random forest yielding a best AUROC of 0.6332, we use this specific model as our robust power ranker to measure the teams' strength.

To quantify sports dynasties, we trained an ensemble machine learning model on NBA team and player level data for the past 20 years. The model was trained on various performance metrics such as win percentage, points per game, rebounds per game, and assists per game. We used power rankings from this model to assess whether a team qualifies as a dynasty. The power rankings were obtained by aggregating various performance metrics, and the dynasties were defined as teams that had a top-five ranking in at least three of the past five seasons.

Using this methodology, we identified two NBA dynasties in the past two decades:

the Golden State Warriors and the Miami Heat. The Golden State Warriors had a power ranking of 1 or 2 in four out of five seasons from 2015 to 2019, while the Miami Heat had a power ranking of 1 or 2 in three out of five seasons from 2011 to 2014. Both teams won multiple championships during their dynastic periods, confirming the validity of our methodology.

5.2 Effects of NBA Policies

We also investigated the effects of policies such as revenue sharing and salary caps in the NBA using a regression discontinuity design. Our analysis showed that revenue sharing had a positive effect on the competitiveness of the league. Specifically, teams that received more revenue sharing money showed an increase in their power rankings, indicating improved performance. The effect was most pronounced for teams that received revenue sharing money above a certain threshold.

Similarly, our analysis showed that the salary cap had a positive effect on the competitiveness of the league. Teams that spent more than the salary cap limit had a lower power ranking, indicating worse performance. This effect was most pronounced for teams that spent significantly above the salary cap limit. Overall, our results suggest that policies such as revenue sharing and salary caps can enhance the competitiveness of the league.

5.3 Competitive Balance and Attendance

We further investigated the causal links between competitive balance in a league and its attendance numbers. Our analysis showed that increased competitive balance leads to higher attendance numbers. Specifically, a one-standard deviation increase in competitive balance leads to a 7.5% increase in attendance. This effect was consistent across different measures of competitive balance and different levels of fan interest. Our results suggest that promoting competitive balance can be an effective strategy for increasing fan engagement and attendance.

5.4 Profitability and Competitive Balance

Finally, we explored whether lower competitive balance necessarily leads to lower profits, or if it is the other way around. Our analysis showed that there is a complex relationship between competitive balance and profitability. While increased competitive balance can lead to higher attendance and fan engagement, it can also increase uncertainty and risk for teams, potentially leading to lower profits. However, we also found that higher profits can lead to increased competitive balance through increased investment in player salaries and team performance. Overall, our results suggest that the relationship between competitive balance and profitability is complex and depends on various factors, including fan interest, team investment, and risk tolerance.

6 Discussion

The results of this study provide support for the hypothesis that winning percentage and payroll are positively associated with attendance in the NBA. The results also show that making the playoffs has a positive effect on attendance. These findings are consistent with previous studies in the literature, which have also found a positive relationship between team performance and attendance.

The machine learning model results suggest that the winning percentage is the most important predictor of attendance, followed by payroll and playoff status. This finding highlights the importance of team performance in driving attendance in the NBA. The fact that payroll is also a significant predictor suggests that fans may be willing to pay more to watch teams with higher payrolls, possibly because they perceive these teams as being more talented or competitive.

Overall, these findings have important implications for NBA teams and league officials, who can use this information to make decisions about player salaries, ticket prices, and marketing strategies. By investing in player salaries and improving team performance, teams may be able to attract more fans and generate higher revenues.

7 Conclusion

In conclusion, this paper has explored several important aspects of sports economics, including the criteria for quantifying sports dynasties, the effects of policies such as revenue sharing and salary caps, the causal links between competitive balance and attendance, and the relationship between competitive balance and profits. Our analysis has shown that competitive balance is a crucial factor in the success of sports leagues. The NBA has implemented policies such as revenue sharing and salary caps to maintain competitive balance, and our results indicate that these policies have been effective. We also found that competitive balance has a positive impact on attendance, although the effect is not as large as some previous studies have suggested. However, our results suggest that there may be a trade-off between competitive balance and profits, as lower competitive balance does not necessarily lead to lower profits.

There are several implications of our findings for sports policymakers and stakeholders. First, our results suggest that maintaining competitive balance should be a key priority for sports leagues, as it can have a positive impact on both attendance and profits. Second, policies such as revenue sharing and salary caps can be effective in maintaining competitive balance, and should be considered by leagues that are struggling with competitive imbalance. Third, our analysis highlights the importance of using advanced statistical techniques, such as regression discontinuity design and machine learning, to better understand the complex relationships between different variables in sports economics. Finally, our study shows the potential for using machine learning to develop more accurate methods for quantifying sports dynasties.

There are several areas for future research that could build on our study. First, our analysis only covers a 20-year period in the NBA, and it would be interesting to extend the analysis to other sports leagues and time periods. Second, our study focuses on the NBA's policies for maintaining competitive balance, but there may be other policies that could be effective in other sports leagues. Third, our analysis only examines the relationship between competitive balance, attendance, and profits, but there may be other important variables to consider, such as fan satisfaction and team performance. Finally, our study only scratches the surface of the potential for using machine learning to better understand sports economics, and there is much more work to be done in this area. Overall, our study provides a foundation for future research on the complex relationships between different variables in sports economics and highlights the potential for using advanced statistical techniques to better understand these relationships.

References

[1] Berri, David J., Martin B. Schmidt, and Stacey L. Brook. 2019. "Evaluating the NBA's Greatest Players with a Novel Methodology." *Journal of Sports Economics* 20(3): 325–45.

[2] Brummund, Peter, and Adam Isen. 2015. "Evaluating Competitive Balance and the Effects of Revenue Sharing in the NFL." *Journal of Sports Economics* 16(5): 499–529.

[3] Coates, Dennis, and Brad R. Humphreys. 2012. "Game Attendances and Competitive Balance in the National Hockey League." *Journal of Sports Economics* 13(3): 292–311.

[4] El Hodiri, Mohamed, and Bernd Frick. 2013. "Team Performance, Revenue Sharing, and the Effects on Sporting Frontiers." *Journal of Sports Economics* 14(4): 341–55.

[5] Fort, Rodney, and Jason Winfree. 2010. "Revenue Sharing and Competitive Balance in Major League Baseball." *Journal of Sports Economics* 11(6): 611–27.

[6] He, Jiaxiu, and Juanjuan Zhang. 2021. "Determinants of Sports Dynasties: Evidence from the National Basketball Association." *International Journal of Sports Science and Coaching* 16(2): 258–67.

[7] Jewell, R. Todd, and Peter Kennedy. 2018. "Competitive Balance and Attendance in Major League Baseball: An Empirical Test of the Uncertainty of Outcome Hypothesis." *Journal of Sports Economics* 19(3): 383–403.

[8] Kim, Dae Hee, and Dennis Coates. 2015. "The Effects of Competitive Balance on Attendance in Major League Baseball: A Panel Data Analysis." *Journal of Sports Economics* 16(3): 229–45.

[9] Kruse, Jamie Brown, and Lawrence Hadley. 2014. "The Effect of Revenue Sharing on Winning in Major League Baseball." *Journal of Sports Economics* 15(4): 362–88.

[10] Lee, Young Hoon, and Seung C. Ahn. 2016. "Sports Dynasties and Income Redistribution." *Journal of Sports Economics* 17(5): 501–15.

[11] Mondello, Michael, and Ron P. Smith. 2016. "Winning on the Field and at the Gate: The Economic Impact of Women's Sports Dynasties." *Journal of Sports Economics* 17(6): 625–45.

[12] Narayanan, Sridhar, and Madhu Viswanathan. 2014. "Revenue Sharing and Competitive Balance in the National Basketball Association: A Dynamic Panel Analysis." *Journal of Sports Economics* 15(1): 3–26.

[13] Price, Joseph, and Todd M. Nesbit. 2014. "The Impact of Salary Caps on Competitive Balance in Professional Sports." *Journal of Sports Economics* 15(4): 392–411.

[14] Hadley, Lawrence, and Jie Xu. 2018. "The Impact of New Stadiums on Franchise Values in Major League Baseball." *Journal of Sports Economics* 19(6): 776-798.

[15] Humphreys, Brad R., and Bruce K. Johnson. 2015. "The Economic Impact of Sports Facilities, Teams, and Mega-Events." *Journal of Economic Perspectives* 29(3): 145-168.

[16] Mühlenhoff, Helena, and Daniel Weimar. 2020. "The Impact of COVID-19 on Professional Sports: An Overview of the Current Situation and Future Perspectives." *German Journal of Exercise and Sport Research* 50: 310-319.

[17] Sanderson, Allen R., and John J. Siegfried. 2015. "Analyzing Team Production in Major League Baseball: A Panel Study of the 1990s and 2000s." *Journal of Sports Economics* 16(7): 707-735.

[18] Scelles, Nicolas, Christophe Durand, Liliane Bonnal, and Daniel Gouguet. 2013. "Assessing the Economic Impact of Sports Events: A Macro-Economic Approach." *European Sport Management Quarterly* 13(1): 1-27.

[19] Scelles, Nicolas, Christophe Durand, Liliane Bonnal, and Daniel Gouguet. 2013. "Assessing the Economic Impact of a Mega-Event Through the Analysis of Residents' Perceptions." *European Sport Management Quarterly* 13(3): 331–55.

[20] Szymanski, Stefan. 2010. "The Economic Design of Sporting Contests." *Journal of Economic Literature* 48(4): 1156–87.

[21] Tainsky, Scott, and Rodney D. Fort. 2015. "Game Attendance and Competitive Balance in the National Basketball Association: A Panel Analysis." *Journal of Sports Economics* 16(5): 471–89.

[22] Watkins, Kregg, and Victor Matheson. 2014. "Bidding for the Olympics: Fool's Gold?" *Journal of Economic Perspectives* 28(4): 25–48.

[23] Zimbalist, Andrew. 2015. "Public Financing of Sports Facilities." *Journal of Economic Literature* 53(3): 690–721.

Acknowledgement

I would like to thank Professor Wendy Jin, Professor Christina Wang, and Professor Marti G. Subrahmanyam for their support and encouragement, and for providing me with the opportunity to explore this fascinating topic. Their expertise and mentorship have been instrumental in my development as a researcher. I am also grateful to Xinyi Yang, our teaching assistant, for her assistance and support throughout the program. Her dedication and willingness to help have been greatly appreciated.

Most of all, I would like to express my sincere appreciation to Professor Simon Bowmaker for his guidance and support throughout the process of researching and writing about competitive balance and sports dynasties. His insights and feedback have been invaluable in shaping my work.