Distribution Agreement:

Varun Saxena April 1, 2024

Determining the Economic Factors of American Professional Sports Franchise Valuations

by

Varun Saxena

Prof. Evan Saltzman
Adviser

Department of Economics

Prof. Evan Saltzman
Adviser

Prof. Shomu Banerjee
Committee Member

Prof. Rachel Hall-Clifford
Committee Member

2024

Determining the Economic Factors of American Professional Sports Franchise Valuations

by

Varun Saxena

Prof. Evan Saltzman
Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Arts with Honors

Department of Economics

2024

Abstract

Determining the Economic Factors of American Professional Sports Franchise Valuations
By Varun Saxena

This study analyzes the effects of American professional sports teams' location, in-game success, star players, fan engagement, and business operations on their financial valuation. Metropolitan area population, metropolitan area median personal income, average game attendance, number of team all-stars, regular season wins, franchise championships, and recent championship are the independent variables that capture these effects in this study. These measures are analyzed for the National Football League (NFL) and National Basketball Association (NBA) sports leagues. The methods for these analyses are an Ordinary Least Squares (OLS) regression and Random Forest regression. Across both leagues, population, income, and operating income are the most significant estimators of franchise value, indicating that location and business operations are the most relevant factors in determining franchise value. The in-game success and fan engagement metrics vary in significance based on the model and league. The presence of star players is largely insignificant in determining franchise value across both leagues. The results of this study explain the economic benefits of franchise relocation to larger markets and growth of sponsorship deals, among other phenomena in the NFL and NBA.

Determining the Economic Factors of American Professional Sports Franchise Valuations

by

Varun Saxena

Prof. Evan Saltzman
Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences of Emory University in partial fulfillment of the requirements of the degree of Bachelor of Arts with Honors

Department of Economics

2024

Acknowledgements

I would like to express my deepest gratitude to my primary thesis advisor, Prof. Evan Saltzman for his patience and guidance throughout this process. I am also thankful for Prof. Shomu Banerjee and Prof. Rachel Hall-Clifford for being helpful and supportive members of my thesis committee. Special thanks to Prof. Stephen O'Connell for his active mentorship role as head of the Economics honors program. I would also like to extend my thanks to Prof. Christina DePasquale from the Johns Hopkins Economics Department for helping me develop the genesis of this project.

I would also like to recognize the Economics graduate teaching assistants for providing resources and direction for this thesis. Additional thanks to my friends and family for our brainstorming sessions and editing during this project.

Table of Contents

1 Abstract

This study analyzes the effects of American professional sports teams' location, in-game success, star players, fan engagement, and business operations on their financial valuation. Metropolitan area population, metropolitan area median personal income, average game attendance, number of team all-stars, regular season wins, franchise championships, and recent championship are the independent variables that capture these effects in this study. These measures are analyzed for the National Football League (NFL) and National Basketball Association (NBA) sports leagues. The methods for these analyses are an Ordinary Least Squares (OLS) regression and Random Forest regression. Across both leagues, population, income, and operating income are the most significant estimators of franchise value, indicating that location and business operations are the most relevant factors in determining franchise value. The in-game success and fan engagement metrics vary in significance based on the model and league. The presence of star players is largely insignificant in determining franchise value across both leagues. The results of this study explain the economic benefits of franchise relocation to larger markets and growth of sponsorship deals, among other phenomena in the NFL and NBA.

2 Background

2.1 American Team Sports

Sports have been a cornerstone of the American entertainment industry for well over a century, beginning with baseball serving as "America's national pastime". As baseball endured through trying times in modern American history, other sports like football began gaining popularity. The Super Bowl era of football began in the mid-1960s and has since grown into the most watched television event in American history. During this expansion of profession sports popularity, basketball icons like Michael Jordan and LeBron James have grown into international celebrities with the popularity of major music stars and actors. Today, professional sports have never been more popular and continue to grow in popularity year after year.

As sports have continued to expand, they have become more corporatized. Initially, professional sports teams served as representatives of the city in their team's name as they played against regional competitors. Sports were treated as solely an entertainment product, no different than a theater performance. Now, professional sports is a multi-billion-dollar industry that uses sports teams as a means for increasing profit and growing the business[1]. This shift is exemplified by teams changing their stadium name from "Boston Garden" to "TD Garden" or "Giants Stadium" to "MetLife Stadium" as a means of increasing the revenue of the business.

2.2 American Sports Markets

In modern American sports, every team operates as a business under the umbrella of its league, with the employees being the players, coaches, and staff, and the product being the games. The business that encompasses each sports team is known as a sports franchise[1]. Sports franchises are similar in structure with traditional businesses in that they have an ownership group, led by the majority shareholder, who is widely considered the owner of the team. They differ from traditional businesses, however, in that the positions underneath ownership consist of a general manager, president of sports operations, and the head coach of the team.

Through an enterprise model of professional sports leagues, they can be viewed as near-monopolies due to the lack of direct competitors offering their product. To use the National Football League (NFL) as an example, collegiate football and leagues like the Canadian Football League (CFL) or the United Football League (UFL) exist as alternative sources of the football product. However, the diminished quality of this product combined with

[1] Vault. (2023)

less popularity across these leagues render the NFL as the sole distributor of high-level professional football. In other words, these other leagues offer a similar product, but cannot directly compete with the NFL because it occupies the entire market demand for high-level professional football. This market saturation renders entry or exit into this market nearly impossible[2].

Each individual sports league functions as an oligopoly, as a finite number of teams compete for the same players, championships, and third-party sponsorship and media deals. Similar to the monopolistic structure of the sports market, there is a very high barrier to entry or exit within a sports league. Though some franchises have relocated, no franchise has ever been eliminated from a sports league, leaving no precedent for exit from an American sports league. The only precedents for entry into a sports league are league mergers, such as the American Basketball Association-National Basketball Association (ABA-NBA) merger in 1976, or league expansion, in which new teams are created and assigned to cities, typically ones without a current team[2]. More specifically, sports leagues operate through the cartel model of oligopolies, where the few businesses in a given market collude together to ensure collective success. In sports leagues, teams engage in group sponsorship deals and revenue sharing models to help each other out.

These revenue sharing practices consist of a portion of the profits of each franchise are pooled together and given to the league office for reallocation. Typically, the league office combines this pool with its own revenue and distributes the funds equally between all teams to ensure the viability poorer performing teams[3]. The specific process for revenue sharing varies by sports league.

Non-sports businesses fall under the oversight of organizations like the Federal Trade Commission (FTC) and the Securities and Exchange Commission (SEC) and experience limits on their business operations through these agencies and tax code. However, the business operations of a sports franchise are also tightly regulated by the league in which the

[2] Li, W. (n.d.)
[3] Thomas, T. (2023, October 25)

franchise exists. The most prominent example of this concept is the Salary Cap, a mechanism used by major sports leagues to standardize the amount of money spent on player salary between teams. Salary Caps are used across most major professional sports leagues to control for fluctuations in franchise profitability during the process of team building[4]. The National Basketball Association (NBA) uses a "soft cap", which allows for teams to exceed the Salary Cap in specific situations, such as to retain a current player on a maximum-salary contract extension. The NFL uses a "hard cap", where teams are not allowed to exceed the Salary Cap under any circumstances.

2.3 Franchise Values

The valuation of a sports franchise is crucial for determining its current financial well-being, outlook, and price should a transfer of franchise ownership take place. This distinction between value and price is especially important for sports franchises due to the infrequency of transactions [5]. In other business sectors where purchases are regularly made, a large sample of purchasing data is available to precisely determine the value of a product, so the gap between price and value is practically nonexistent. With sports franchises, values must be determined through an alternative cash flow approach, similar to how private businesses are valued.

This cash flow approach incorporates revenue, expenditures, and secondary factors to determine how much a franchise is worth[6]. Revenue sources primarily include television and radio deals, corporate sponsorships, ticket sales, and merchandise. Expenditures primarily include player and coaching contracts, facility costs, staffing, and business operation costs. Secondary factors can include branding, stadium lease agreements, or fan base composition, although these factors play a smaller role in the valuation. The output of this cash flow model is then scaled by a market multiplier, which gives value boosts to franchises in larger

[4]Brandt, A. (2019, June 25)
[5]Damodaran, A. (2024, March)
[6]Kaushik, V. (2022, January 14)

cities, such as New York or Los Angeles.

While this approach has been mostly reliable, there are some underlying issues with the outlined methodology. Firstly, the cash flow approach fails to properly account for singular major changes in the composition of a sports franchise, such as a major sponsorship, a new television contract, or the signing of a major free agent. Secondly, much of the information used to determine the factors behind the valuation is speculative or unverified by the franchise itself[7]. Outside of improved data collection methods, there exists no solution to the second issue with a cash flow approach. However, there exist alternatives that address the first concern and provide more predictive power to a valuation model. Two such alternatives are Ordinary Least Squares (OLS) and Random Forest regression. Both of these methods also grant more predictive power, allowing for future valuations to be calculated as well.

2.4 OLS Regression

Ordinary Least Squares regression is an econometric technique for developing linear regression equations that describe the relationship between one or a set of independent variables and a dependent variable. This style of regression yields a set of coefficients and standard errors that estimate the effect of a particular independent variable on the dependent variable. Some previous work has used OLS regression to estimate franchise value, although this method has not been used extensively in sports finance. One such study was conducted in 2004 by Donald L. Alexander and William Kern. This paper details the effect of several proposed factors that could alter the value of professional sports franchises in the National Football League (NFL), National Basketball Association (NBA), National Hockey League (NHL), and Major League Baseball (MLB)[8]. At the time of its publication, few studies had examined the economics of sports franchises. As such, the novelty of this paper fueled much research in the field of sports economics.

This paper focuses on the non-sports-related factors of a sports franchise, analyzing

[7] Appraisal Economics. (2021, February 17)

[8] Alexander, D. L., & Kern, W. (2004)

mostly business-related variables that could reasonably affect any large business. The identified economic determinants used in this study include per-capita income, city population, team standing in the previous season, new team (recent expansion team), new facility (recently built new stadium), new location (recent franchise relocation), team identity change, regional identity, and regional population. The per-capita income variable was used to control franchise-specific effects that are inherently unchangeable from franchise to franchise. The city population and regional population variables evaluate the market size of the franchise. The rest of the variables are meant to evaluate if fan or investor engagement would increase or decrease based on structural changes within the organization[8]

Even though there are noticeable differences between the four different professional sports evaluated in this paper, the same variables are used for all four leagues. Data for certain variables, such as identity change, new location, and new team for the NBA, was not available, so those variables were excluded from the analysis of the NBA. All franchise values were taken from Financial World (FW) which uses financial models to estimate the value of businesses, including professional sports franchises[8].

All of these data analyses are conducted through an unrestricted OLS regression model with all available variables for a given league. The two unrestricted OLS regression models use Franchise Value and log(Franchise Value) for the regressands. The regression models use a combination of binary and continuous variables with listed parameters for the binary variables[8].

The result of this OLS regression analysis is that city population and team standing are the two most significant determinants of a franchise's value. For the NFL, the city population does not have much of an effect on franchise value. The researchers propose that the rationale for this result is that the revenue-sharing model in the NFL diminishes the effect of franchise location on its value. The study also finds that a new stadium had a positive effect on franchise value. The rationale for this result is that new facilities offer more revenue streams

[8]Alexander, D. L., & Kern, W. (2004)

(more seats, more concessions options, better facilities). However, expansion teams do not have a noticeably different value when compared to other teams in a given league. The study shows that having a strong regional identity also increases franchise value by proxy of increased fan engagement. The other variables had an insignificant effect on franchise values. From this study, the conclusion is that the market in which a team resides and its success are the predominant factors that determine a sports franchise's value[8].

Another similar study was conducted by Don Grage in 2022 measuring the impact of social media following, attendance, and city population on Major League Baseball (MLB) and NFL franchise values[9]. This study utilizes simple linear regression with each of the explanatory variables and identifies each individual team in the plots for deeper analysis. The result of this study is that attendance and social media following have significant impacts on franchise value in both the MLB and NFL, but city population is only significant in the NFL. The study also states that these differences only account for 70-80% of the factors behind franchise value, and additional measures like championships, playoff appearances, and year of existence need to be incorporated to gain a fuller understanding of the determinants of franchise value.

OLS regression is a critical tool for unpacking the relative importance of a set of independent variables on a dependent variable, but OLS models sometimes need modification to increase accuracy. One of the most common modifications done to OLS models is a time series correction, which siphons out the effects of year-over-year changes from the model. In the context of estimating the value of a business, a time series correction reduces the impact of market fluctuations and inflation from the model, allowing it to focus more on the independent variables of interest. The first way of implementing a time-series correction into an OLS model is to introduce an independent variable into the regression that represents time, known as a time trend. This variable increases by one for each successive year increase within the dataset. If the dependent variable changes by a nonlinear pattern over time,

[8]Alexander, D. L., & Kern, W. (2004)
[9]Grage, D. (2022, February 2)

another "time" variable can be created in the form of the logarithm of time ($log(time)$) or time-squared ($time^2$). The second way of implementing a time-series correction into an OLS model is to introduce a set of dummy variables into the model, each of which corresponds to a year in the dataset. This way, each data entry will be analyzed in the context of the year it existed. This form of time-series correction is known as a fixed-effects model.

2.5 Random Forest

Random Forest is a decision-tree-based machine learning algorithm that has gained popularity over the last few years for creating heart disease risk models, determining agricultural land valuations, and guiding financial trading [10][11]. Random Forest models use a multivariate data set with a singular dependent variable, similar to an ordinary least squares linear regression model. From this input data set, a typical decision tree model would analyze patterns in the independent variables to create a decision tree that could be followed to obtain the proper dependent variable value[12]. Each node can be categorically, binarily, or continuously constrained to sort incoming data into separate branches on the tree. For example, a continuous node could sort all $x_1 > 4.0$ values down the left branch and all others down the right branch. The fully formed tree contains nodes and branches such that any inputted multivariate data set will yield a highly accurate dependent variable value. However, the issue with creating one decision tree with all variables and all input data is that it does not properly weigh the importance of the different variables and becomes highly specific to the input data set. With the goal of externalizing the model beyond the input data set, creating one decision tree may not yield the most accurate result. To address this issue, the Random Forest model creates hundreds or thousands of decision trees that only incorporate several input variables and a portion of the inputted data to create a decision tree. Through this method, each tree in this "forest" has a unique set of data and input variables, which means

[10] Hong, J. et al. (2019)
[11] Kim, J. et al. (2019)
[12] IBM. (n.d.)

that every dependent variable value that the tree puts out has a separate methodology from all other trees in the forest. With this "forest" of trees in place, a new data sample can be introduced and run through every tree in the forest, each of which will output its own dependent variable value. Through either a weighted average or a majority vote, the outputted values will be condensed into a singular output value for the dependent variable[13]. Although Random Forest models and OLS models have different methodologies, both are considered forms of regression analysis due to the nature of the models' inputs and outputs. The benefit of using a Random Forest model over an OLS model is that Random Forest models are less sensitive to multicollinearity issues, outliers, and heteroskedasticity as a result of its method of action[14]. They also can be more difficult to interpret due to the lack of estimator coefficients. The predictive power of machine learning algorithms can theoretically be better than an OLS model, but that varies greatly based on the dataset in question.

Random Forest algorithms have been used in relation to sports in the past, but not to quantify the economic determinants of professional sports franchises. One such study used Random Forest to create the ideal lineup for a team in the Indian Premier League (IPL), one of the biggest Cricket leagues in the world. Like many other sports, developing a roster requires balancing talent acquisition with roster balance, making sure that the player's skillsets, strengths, and weaknesses complement each other. The Random Forest algorithm uses statistical metrics to evaluate the strengths and weaknesses of a player in addition to their overall talent. Then, Random Forest uses its decision tree-based algorithm to test out a variety of different rosters to find the ones that theoretically perform the best[15].

2.6 Purpose

This senior thesis project seeks to determine the effect of a team's market, in-game success, fan engagement, star players, and business operations on their franchise valuation in the

[13]Biau, G et al. (2016)
[14]Sahota, H. (2022, January 31)
[15]Rodrigues, N et al. (2019)

National Basketball Association (NBA) and National Football League (NFL). Investigating the economic determinants of professional sports franchises across multiple leagues will allow for the isolation of the economic drivers for each league. By isolating these factors for each league, analysis can be done on factors intrinsic to one specific league. For example, the regular season record may impact NFL franchise values more than it would impact NBA franchise values due to there being 17 games instead of 82. Additionally, the number of all-stars may impact NBA franchise values for than the NFL due to there only being 5 players on the court at a given time, increasing the impact of a given star player on the franchise, both on and off the court.

3 Methods

3.1 Data Collection

The dependent variable for this project is Team Franchise Value, as determined by Forbes[16][17]. This data is sourced from Forbes.com through Statista, a large online database[18][19].

The independent variables measuring a team's market are metropolitan area population and metropolitan area median personal income. These data are sourced from the Bureau of Economic Analysis (BEA)[20][21], a bureaucratic agency in the United States federal government. Median Gross Domestic Product (GDP) was also gathered, but the sample size was deemed insufficient for inclusion into this study. The independent variables that measure a team's in-game success are regular season wins, total franchise championships, and a dummy variable representing a championship won in the last five years. These data are sourced from

[16] Ozanian, M., & Teitelbaum, J. (2023a, August 30)
[17] Ozanian, M., & Teitelbaum, J. (2023a, October 26)
[18] Statista. (2023a, August)
[19] Statista. (2023a, October)
[20] Bureau of Economic Analysis. (2023a, November 16)
[21] Bureau of Economic Analysis. (2023b, December 7)

Basketball Reference[22][23] and Pro Football Reference[24][25] for the NBA and NFL, respectively. The independent variable representing fan engagement is average regular season attendance. This data is sourced from the Entertainment and Sports Programming Network (ESPN) website[26][27]. The independent variable for this project representing star players is number of all-stars. This metric captures the number of high-level players on a team's roster. This data is sourced from Basketball Reference[28] and Pro Football Reference[29] for the NBA and NFL, respectively. The independent variable for this project representing business operations is operating income. Operating income measures the revenue of a business minus the operating expenses and depreciation values. This data is sourced from Forbes.com through Statista[30][31].

The reason for using Statista instead of pulling data directly from Forbes.com is that Forbes team-level data beyond the year 2023 is inaccessible without paying a sum far outside this project's budget for equivalent data directly from Forbes. Emory undergraduate students are granted free access to Statista's entire database, which sources the franchise value and operating income figures directly from Forbes.com. To verify this, Statista's 2023 values were cross-checked with the values on Forbes' website and found an exact match. Statista has been deemed reliable by credible academic institutions[32]

The data for this project is collected from their respective sources through the acquisition of spreadsheet files and web scraping algorithms through the Pandas package in the coding language Python. The reason for using web scraping algorithms as an alternative to downloaded spreadsheet files is that many of the sources have distinct web pages and/or

[22] Basketball Reference. (2024a)
[23] Basketball Reference. (2024b)
[24] Pro Football Reference . (2024a)
[25] Pro Football Reference . (2024c)
[26] ESPN. (2024a)
[27] ESPN. (2023)
[28] Basketball Reference. (2022)
[29] Pro Football Reference . (2024b)
[30] Statista. (2023b, August)
[31] Statista. (2023b, October)
[32] HBS Baker Library. (2024)

tables for each team in a given league and/or each year across the time span analyzed in this study. Web scraping algorithms provide a useful alternative to reading in Comma Separated Values (CSV) files or Excel files if there are a large number to download for a given measure. This alternative method extracts only the necessary information directly from the webpage and deposits it into a Python file in the form of a DataFrame. The algorithm is flexible in nature and can iterate over these multiple web pages and/or tables very easily, rendering it a much more efficient alternative to downloading individual spreadsheets in the outlined cases. Once the data is collected in a Python file, it is cleaned and condensed into a large DataFrame that is exported as a CSV file for further analysis.

For interpretation purposes, some variables in this study have been scaled. Franchise value and operating income are measured in millions of United States dollars. Population, personal income, and attendance have also been scaled to degrees specified in the regression tables. Unless otherwise specified, these measures are not scaled.

3.2 Independent Variables

Population as an captures both the size of the market and the size of the fanbase likely to support the local team. A larger market size and larger fanbase could lead to more lucrative television contracts, expanded media coverage, and increased fan engagement as a product of volume. However, many sports pundits hypothesize that larger markets have more alternative entertainment products, such as other sports leagues, theater, or social venues. This would hypothetically give increased fan engagement in smaller markets because the fans have less alternatives, increasing their commitment to the team. Empirical evidence for this theory can be seen in the NFL, where places like Pittsburgh, Green Bay, and Kansas City are known for their rabid fanbases, while the Los Angeles Rams and Chargers consistently struggle to fill their stadium with their own fans[33]. Median personal income captures the economic health of the individuals in a team's area. Areas with higher median personal

[33]Sullivan, K (2022, August 1)

income would be more economically viable, potentially leading to more merchandise profit and larger sponsorship deals.

The regular season wins and recent championship variables aim to capture the short-term success of a team. The practical distinction between these two variables is that regular season wins captures a strong individual season, disregarding postseason success. Recent championship captures exclusively postseason success but incorporates a potential multi-year spike in team popularity following a championship win. Championships measures the historical relevance of a team, weighing team success from decades ago equally with recent championships. Hypothetically, higher levels of all three of these in-game success variables should lead to higher franchise valuations across both sports. The rationale for this hypothesis is that more attention is given to teams with higher levels of success, giving them increased popularity. This increased popularity leads to higher levels of television streaming, more merchandise and ticket sales, and more media attention. These byproducts of winning are potential sources of revenue for owners.

The attendance statistic aims to capture fan engagement and the quality of the stadium in which a team plays. Due to league-wide popularity of the two sports selected for this study, the stadiums for most teams operate at nearly full capacity, making the attendance figure disproportionately reflect stadium size. However, there are enough teams that operate at below full capacity that this statistic is reliable for evaluating fan engagement with a team as well. Hypothetically, higher levels of regular season attendance should lead to higher franchise values across both sports. The rationale for this hypothesis is that higher levels of fan engagement and larger stadiums should correlate with increased popularity and a better fan experience. Like the effects of in-game success, the byproducts of these factors are potential sources of revenue for owners.

The number of all-stars on a team each year aims to capture the effect of star power and marketing campaigns. In the NBA, 12 players are selected as all-stars every season from each conference, 24 in total. In the NFL, 44 players are selected as pro bowlers, the NFL equivalent of all-stars, every season from each conference, 88 in total. Oftentimes, players get injured or opt out of the all-star game, resulting in more all-stars selected than the allotted amount. This number fluctuates year-to-year. While this metric does a good job of measuring the number of excellent players, it does not calculate the magnitude of each players' stardom. For example, Lebron James and Stephen Curry are two of the most prominent sports figures in the world, but are counted as equally as Rudy Gobert, an all-star caliber player only known by NBA fans. Additionally, the effect of having a former all-star player whose fame exceeds their playing ability is not captured by this metric. Hypothetically, having more all-stars should lead to a higher franchise value. This is because better players typically create more winning and fan engagement, leading to the winning-related effects detailed above. Additionally, the presence of a star player, especially if that player is of the highest caliber (e.g. Stephen Curry or LeBron James), leads to more local and national marketing campaigns for the player, team, and league. For example, a team may be discussed more on national sports media if they have more star players, leading to increased merchandise sales or television viewership for that team. Lastly, the celebrity status of a player could lead to the inclusion of non-sports fans into the fanbase of a team. A recent example of this phenomenon is the increase in Kansas City Chiefs merchandise sales and viewership following the public relationship of award-winning singer Taylor Swift and Chiefs star Travis Kelce.

The operating income metric is critical for evaluating a business' viability and health. In the context of sports franchises, operating income evaluates the quality of the business operations of a franchise. Better quality of operations includes increased merchandise sales, ticket sales, television deals, and other sources of revenue. While operating income captures much of the effects that this study hopes to evaluate, this metric is also incorporated into the cash flow approach that calculates franchise value[5]. For this reason, operating income is only selectively used in the regression equations in this study. From the design of the cash

[5]Damodaran, A. (2024, March)

flow model used to calculate franchise values, more operating income is established to lead to a higher franchise value. The rationale for including operating income into select regression equations is to compare the other estimators with and without operating income. This comparison will allow for the determination of the factors that may be absorbed by operating income. The distinction between statistically significant factors that are impacted by the omission of operating income and those that are not impacted is important in deciphering how exactly that measure impacts franchise value.

Among these explanatory variables, Population, Income, Operating Income, and Championships are hypothesized to have the biggest positive impact on franchise value. Wins, Recent Championship, Attendance, and All-Stars are hypothesized to have smaller positive impacts of franchise value as compared to these other variables. Between the two leagues, All-Stars is hypothesized to have a bigger impact in the NBA than the NFL due to the player-centric marketing and fewer players on a team.

3.3 Regression Models

There are three datasets used in this study: an NBA dataset, an NFL dataset, and an NBA-NFL combined dataset. This combined dataset contains a dummy variable that controls for the sports league of each data entry. Inclusion of this dummy variable is important due to the discrepancies in regular season games, league-wide revenue and popularity, and stadium size between these two leagues. The NBA dataset ranges from the year 2012 to 2022, while the NFL dataset ranges from the year 2002 to 2023. Population and income data for the Toronto Raptors within the NBA dataset are excluded due to its location in Canada. Additionally, population and income data is not available for the year 2023, so data entries for that year are excluded as well.

The variables included in all of the restricted regression equations are population, income, attendance, and all-stars. To avoid multicollinearity issues, the independent variables measuring on-court success are split into three separate restricted regression equations. Additionally, due to operating income having a significant role in the calculation of franchise value, each restricted regression model is run with and without operating income as an independent variable.

$$FranchiseValue = \beta_0 + \beta_1 Population_{it} + \beta_2 Income_{it} + \beta_3 Attendance_{it} + \beta_4 AllStars_{it} + \beta_5 InGameSuccess^{*}_{it} + \beta_6 OperatingIncome^{**}_{it} + \lambda_t + u_{it}$$

Equation 1. Time fixed-effects model for the NBA and NFL individually

*This estimator consists of either Wins, Championships, or Recent Championship

**This estimator is not included in every reported regression model

$$FranchiseValue = \beta_0 + \beta_1 Population_{it} + \beta_2 Income_{it} + \beta_3 Attendance_{it} + \beta_4 AllStars_{it} + \beta_5 InGameSuccess^{*}_{it} + \beta_6 OperatingIncome^{**}_{it} + \beta_7 D_{league} + \lambda_t + u_{it}$$

Equation 2. Time fixed-effects model for the NBA and NFL combined

*This estimator consists of either Wins, Championships, or Recent Championship

**This estimator is not included in every reported regression model

Each of these regression equations has a corresponding OLS, time variable, and time fixed-effects model. However, the time fixed-effects model has the highest validity of the three models across these regression equations for each dataset, so that is the only model reported in this study.

The Random Forest regression model includes population, income, wins, championships, recent championship, attendance, and all-stars as the independent variables. Operating income is omitted from this model due to its integral role in calculating franchise value. Random Forest models do not process dummy variables as well as OLS regression models, so a quadratic time trend is incorporated into the Random Forest model instead of fixed-effects parameters. Additionally, Random Forest models cannot tolerate N/A values, so any data entry with an N/A value is excluded from the Random Forest model. These specific exclusions are listed above in the OLS regression model design.

$$FranchiseValue = \beta_0+\beta_1 Population_i+\beta_2 Income_i+\beta_3 Attendance_i+\beta_4 AllStars_i+\beta_5 Championships_i+ \beta_6 OperatingIncome_i^* + \lambda_{team} + u_i$$

Equation 3. Team fixed-effects model for the NFL

*This estimator is not included in every reported regression model

This team fixed-effects model aims to capture the league-wide effects of these explanatory variables on franchise value over two distinct time periods. By controlling for team, the explanatory variables will instead measure how these variables affect the league as a whole over two distinct time periods: 2002-2012 and 2013-2023. The reason for breaking up the date range into two 11-year periods is that there is a steep change in growth rate that occurs around the early 2010s, so assessing the differences in each explanatory variables between these date ranges is instrumental in unpacking this difference.

$$FranchiseValue = \beta_0 + \beta_1 Population_{it} + \beta_2 Income_{it} + \beta_3 Attendance_{it} + \beta_4 AllStars_{it} + \beta_5 Wins_{it} + \beta_6 Championships_{it} + \beta_7 D_{RecentChampionship} + \beta_8 t + \beta_9 t^2 + u_{it}$$

Equation 4. Equivalent regression model for the Random Forest model. β coefficients are not used in random forest models, this equation serves as a visualization for the setup of the Random Forest model.

Random Forest regression models are designed by separating the data into two groups: a training set and a testing set. The training set, which consists of approximately 75% of the total data, serves as the input for the designing of the forest of decision trees. The testing set, which consists of approximately 25% of the total data, is run through the decision trees created by the training set to determine if the established dependent variable value matches the predicted value determined by the Random Forest regression model. The mean absolute error is then calculated based on the discrepancy between the predicted and actual values, which determines the validity of the model.

4 Results

4.1 Summary Statistics

Table 1: NFL Summary Statistics

	count	mean	std	min	median	max
Franchise Value	703	1855.31	1380.01	374.0	1173.0	9000.0
Population	703	4738330.70	4400332.09	288747.0	3290730.0	19774386.0
Income	671	51062.37	14109.26	25332.0	47958.0	124398.0
Attendance	689	66152.62	12851.08	375.0	67672.0	105149.0
All Stars	703	3.46	2.39	0.0	3.0	13.0
Wins	703	8.02	3.09	0.0	8.0	16.0
Championships	703	1.45	1.73	0.0	1.0	6.0
Recent Championship	703	0.15	0.36	0.0	0.0	1.0
Operating Income	703	58.02	61.01	-49.0	40.6	504.0

Note:
count = number of observations; std = standard deviation
Income data is not available for the year 2023, leading to less income observations
Due to the COVID-19 pandemic, certain stadiums did not report or measure attendance data for the years 2020 and 2021, leading to less attendance observations

Table 2: NBA Summary Statistics

	count	mean	std	min	median	max
Franchise Value	330	1950.33	1250.93	405.0	1600.0	7700.0
Population	319	5542590.34	4894550.20	1123950.0	4317016.0	19774386.0
Income	319	57542.92	13221.35	35389.0	55105.0	124398.0
Attendance	330	16321.84	4754.28	0.0	17469.0	21876.0
All Stars	330	0.88	0.85	0.0	1.0	4.0
Wins	330	40.02	12.05	10.0	41.5	73.0
Championships	330	2.33	4.16	0.0	1.0	17.0
Recent Championship	330	0.12	0.33	0.0	0.0	1.0
Operating Income	330	41.74	40.24	-99.0	34.5	200.0

Note:
count = number of observations; std = standard deviation
Population and Income data is not available for the Toronto Raptors, leading to less observations in those two statistics

Table 3: NBA and NFL Combined Summary Statistics

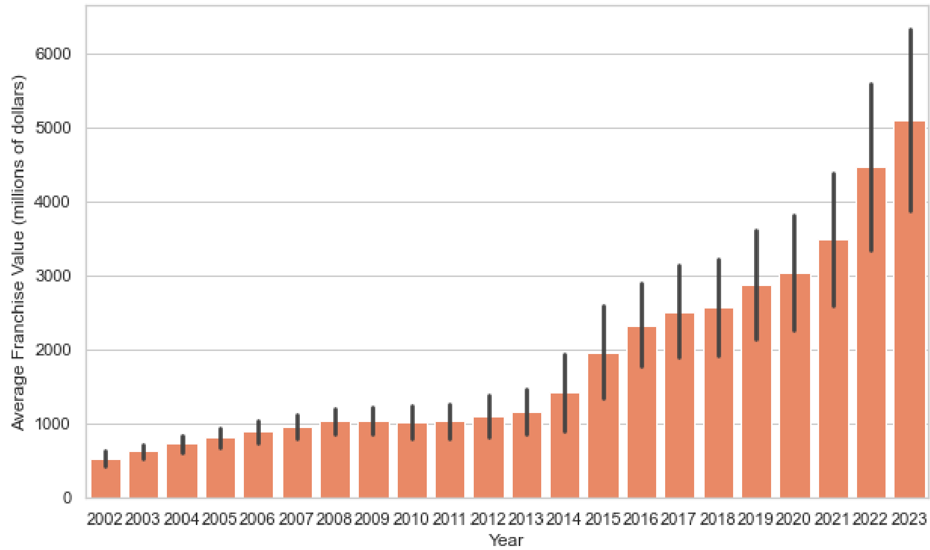
	count	mean	std	min	median	max
Franchise Value	1033	1885.66	1340.26	374.0	1400.0	9000.0
Population	1022	4989366.73	4573152.46	288747.0	3500666.5	19774386.0
Income	990	53150.55	14150.92	25332.0	50831.0	124398.0
Attendance	1019	50015.08	25752.27	0.0	62746.0	105149.0
All Stars	1033	2.64	2.36	0.0	2.0	13.0
Wins	1033	18.25	16.60	0.0	10.0	73.0
Championships	1033	1.73	2.78	0.0	1.0	17.0
Recent Championship	1033	0.14	0.35	0.0	0.0	1.0
Operating Income	1033	52.82	55.73	-99.0	38.8	504.0

Note:
count = number of observations; std = standard deviation
Income data is not available for the year 2023, leading to less income observations
Due to the COVID-19 pandemic, certain NFL stadiums did not report or measure attendance data for the years 2020 and 2021, leading to less attendance observations
Population and Income data is not available for the Toronto Raptors, leading to less observations in those two statistics

The summary statistics across the NFL, NBA, and the combined dataset in Tables 1, 3, and 5 indicate a strong right skew for franchise value, as the mean is consistently higher than the median. This suggests that Similar trends are observed with the population statistic as well, indicating that the rate of change for increases in population may also be increasing. Additionally, the summary statistics indicate abnormally low attendance numbers in the distribution. These low attendance numbers can be attributed to the reduced stadium capacities instituted during the COVID-19 pandemic, which impacted both leagues during the 2020 and 2021 seasons.

The average NFL franchise values by year, as seen in Figure 1, demonstrate a consistent year-over-year increase in franchise value across all franchises. The pattern of this increase appears to follow a somewhat quadratic trend, indicating that the rate of growth of the NFL as a whole is also increasing over time. As it relates to the data, this increase over time

Figure 1: Average NFL Franchise Value by Year



stresses the importance of controlling for time in the regression analyses, whether that be through the use of a fixed-effects model or a time trend. The error bars reflect an increase in variance for the franchise values of teams over time, indicating that the differentiating factors between these teams may have grown in importance over time as it relates to franchise value.

The NFL franchise value histogram, as seen in Figure 2, shows a considerable right skew in the data. Based on the mean and median values from Table 1, this skew is to be expected. The quadratic trend seen in Figure 1 may explain the strong magnitude of this right skew, as there are many more years of relatively low average franchise values before notable league-wide growth begins around the mid-2010s.

The average NBA franchise values by year seen in Figure 3 demonstrate consistent year-over-year increases, similar to the NFL franchise values by year in Figure 1. However, the rate of growth appears to be more consistent than the NFL rate of growth, suggesting a more linear trend as opposed to a quadratic trend. For the purposes of this study, a quadratic time

Figure 2: NFL Franchise Value Histogram

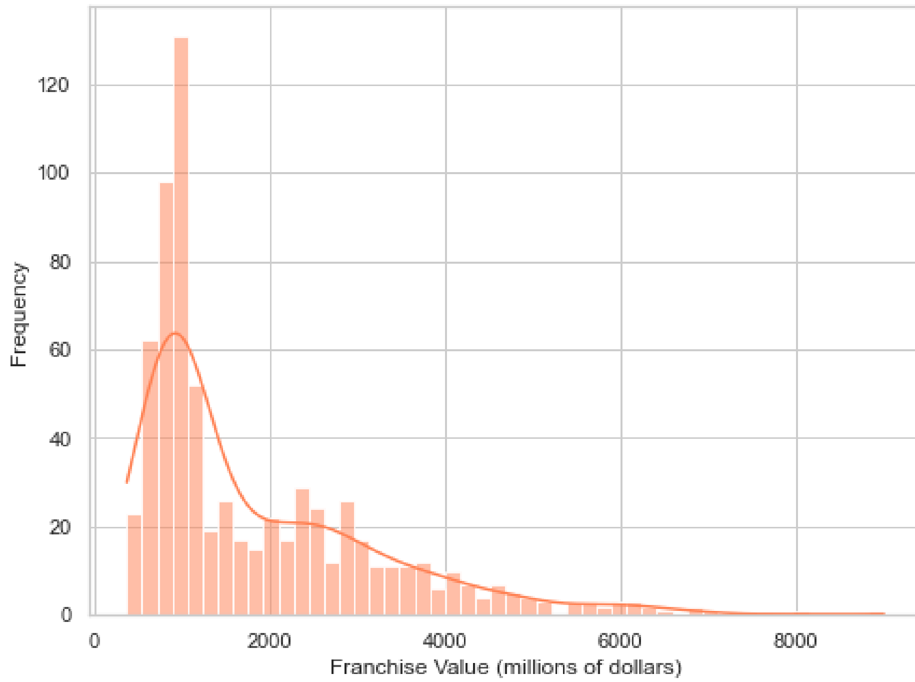
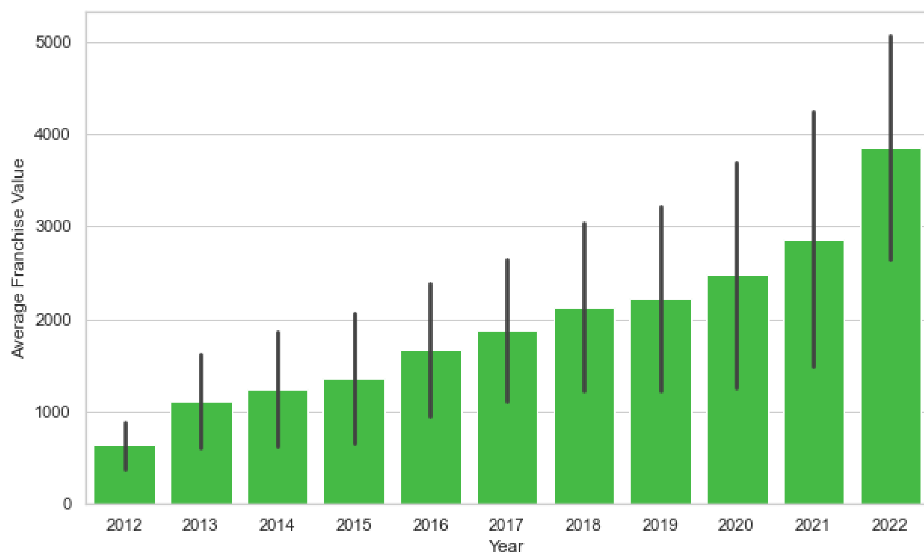
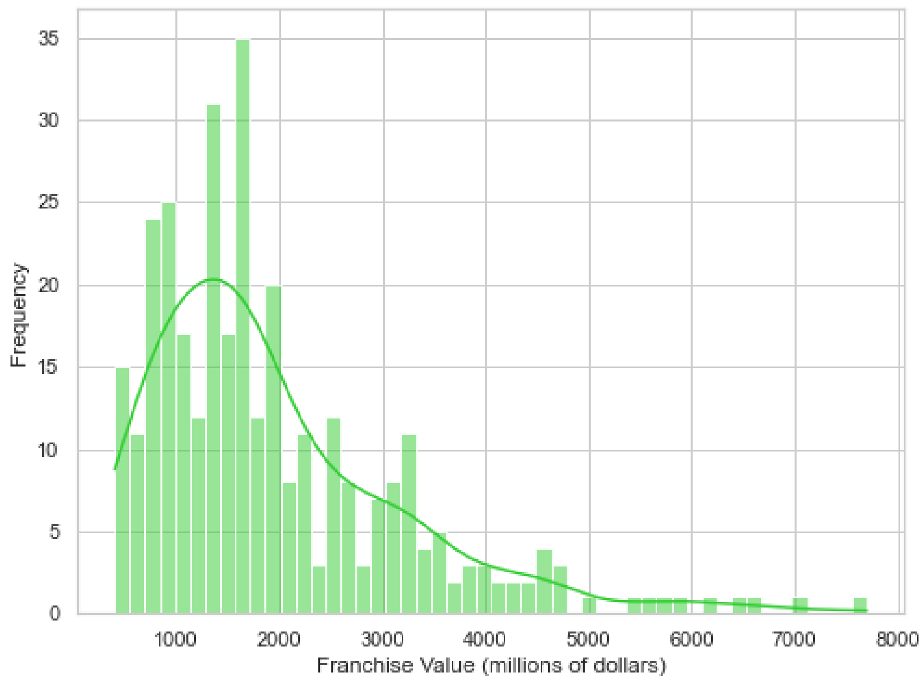


Figure 3: Average NBA Franchise Value by Year



trend will still be applied to the NBA dataset in the Random Forest model since slightly larger jumps in franchise value are observed in the latter years in this dataset. The error bars indicate higher variance in franchise values as time progresses, similar to the NFL franchise values.

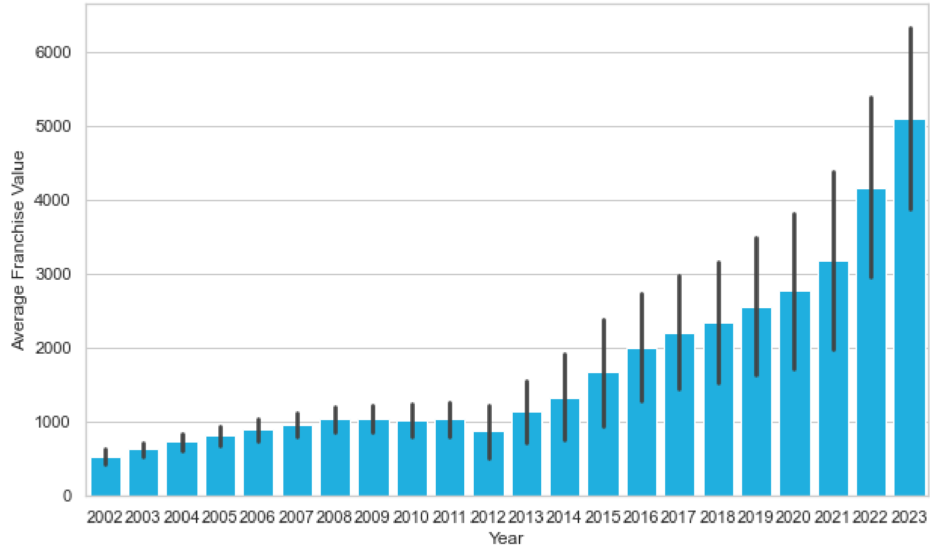
Figure 4: NBA Franchise Value Histogram



The NBA franchise value histogram in Figure 4 demonstrates a strong right skew, just like the NFL franchise value histogram in Figure 2. The gravity of this right skew is not quite as strong as what is depicted in Figure 2, but that is to be expended due to the more linear league-wide growth in franchise value over time. However, the right skew is too strong for a linear growth pattern, as a more even distribution would be expected. The error bars showing an increase in franchise value variance could possible explain why the distribution is so skewed despite a linear growth pattern.

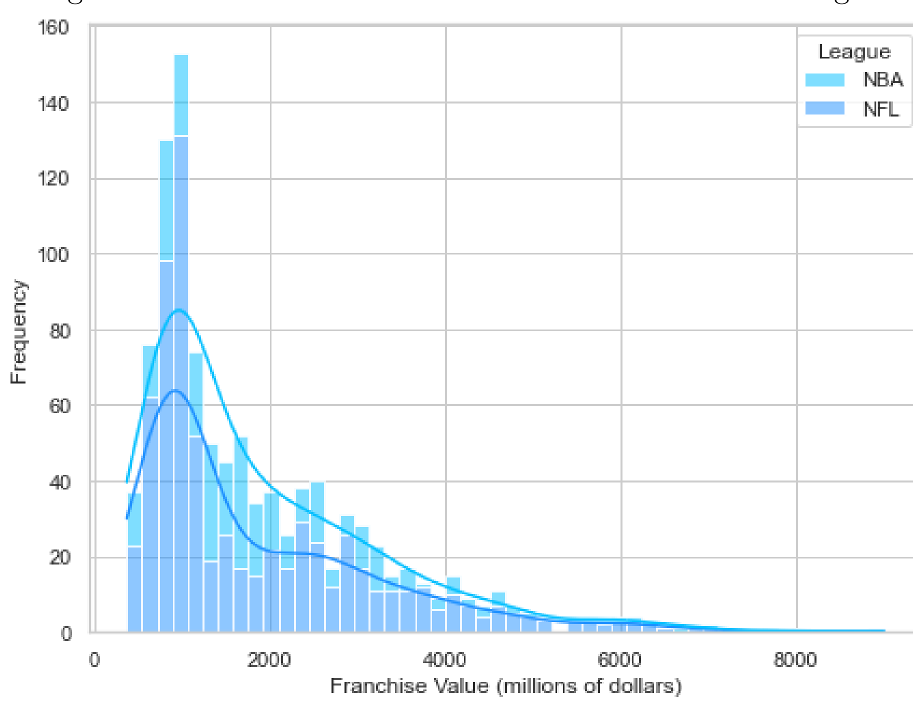
The average NBA and NFL combined franchise values by year in Figure 5 show consistent year-over-year growth in value, with the exception of a brief stint in the early-2010s where a slight decrease is observed. The explanation for this observation is that the years before 2012 do not contain NBA data. Since the NBA is a smaller league than the NFL, it is to be expected that the introduction of NBA teams into the dataset would bring down

Figure 5: Average NBA and NFL Combined Franchise Value by Year



the average value of the constituent teams. Other than this observation, the trend largely resembles the NFL quadratic growth trend in Figure 1 with increasing standard deviations over time.

Figure 6: NBA and NFL Combined Franchise Value Histogram



This histogram shown in Figure 6 overlays the distribution of NBA and NFL franchise values. This overlay allows for the direct comparison of NBA and NFL franchise value distributions. A right skew is observed in this distribution, which aligns with the distribution patterns seen in Figures 2 and 4. Additionally, high amounts of very low franchise values are observed in the NFL, while the NBA contains a relatively even amount of franchises in each bin upto 2 billion dollars, where it starts to drop off. The NFL, however has more franchises valued in the extremely high range compared to the NBA.

4.2 OLS Regression Results

4.2.1 NFL

Table 4: NFL OLS Regression Results

In-Game Success Metric:	Wins	Chips	RctChip
Intercept	-1449.75 (222.84**)	-961.16 (173.16**)	-1397.31 (228.18**)
Population	22.27 (4.94**)	24.78 (4.66**)	22.5 (4.91**)
Income	149.82 (20.51**)	62.68 (22.44**)	146.27 (19.57**)
Attendance	194.85 (35.02**)	158.26 (28.17**)	192.9 (35.58**)
All Stars	4.16 (9.46)	0.26 (6.09)	8.27 (7.06)
Wins	7.27 (7.05)	–	–
Championships	–	106.39 (13.86**)	–
Recent Championship	–	–	100.53 (46.25*)
B-P-G Test Statistic	5.44	7.11	5.42
Adjusted R-Squared	0.87	0.89	0.87

Note:
Chips = Championships; RctChip = Recent Championship; B-P-G = Breusch-Pagan-Godfrey
Scaled Statistics: Population (millions); Income (ten-thousands); Attendance (ten-thousands)
All standard errors are heteroskedasticity robust (HC1)
Breusch-Pagan-Godfrey test statistic determined before correction for heteroskedasticity
Standard errors are denoted in parenthesis below the regression coefficients
* Indicates statistical significance at the 5% level;
** Indicates statistical significance at the 1% level

Table 5: NFL OLS Regression Results with log(Franchise Value)

In-Game Success Metric:	Wins	Chips	RctChip
Intercept	5255.02 (73.88**)	5455.21 (62.7**)	5281.17 (74.85**)
Population	9.56 (1.95**)	10.6 (1.84**)	9.74 (1.93**)
Income	61.86 (7.76**)	25.98 (7.6**)	59.61 (7.43**)
Attendance	105.94 (10.78**)	90.83 (9.06**)	104.35 (11.01**)
All Stars	7.16 (4.10)	5.34 (2.5*)	8.11 (2.96**)
Wins	2.74 (2.87)	–	–
Championships	–	43.8 (4.46**)	–
Recent Championship	–	–	63.74 (15.66**)
B-P-G Test Statistic	3.59	5.37	4.25
Adjusted R-Squared	0.93	0.94	0.93

Note:
Chips = Championships; RctChip = Recent Championship; B-P-G = Breusch-Pagan-Godfrey
Scaled Statistics: Population (millions); Income (ten-thousands); Attendance (ten-thousands)
All standard errors are heteroskedasticity robust (HC1)
Breusch-Pagan-Godfrey test statistic determined before correction for heteroskedasticity
Standard errors are denoted in parenthesis below the regression coefficients
* Indicates statistical significance at the 5% level;
** Indicates statistical significance at the 1% level

In the NFL OLS regression models in Table 4, income, population, and operating income are the most significant contributors to franchise value. Regular season wins and number of all-stars appears to be insignificant in determining franchise value. Championships and recent championship are both significant contributors to franchise value as well, while the significant of attendance varies greatly depending on the inclusion of operating income into the model. The sign of all the coefficients in this model are positive, indicating that higher levels of each of these estimators leads to a higher franchise value. With the exception of regular season wins and all-stars, the results of these regressions align with the project's hypothesis.

In the NFL OLS regression models in Table 5 that regress the logarithm of franchise value on the explanatory variables, the results are largely the same as those in Table 4. The only main difference is the significance of the number of all-stars, which is insignificant when regressing with the raw franchise value data. Since the two tables do not align for this particular variable, the effect of all-stars remains inconclusive. The heteroskedasticity is much lower in these models, as indicated by the Breusch-Pagan-Godfrey test statistic. Additionally, the R-squared is higher across all six models, indicating that the logarithm of franchise value is the better indicator for explanatory purposes.

The regression results did not vary greatly between regular season wins and recent championship for both the operating income-inclusive regression and the operating income-omitted regression. The main distinction between these sets of regressions is that recent championship is statistically significant at the 5% level, while wins is not significant. Additionally, all-stars becomes statistically significant at the 5% level in the operating income-inclusive regression with recent championship. A potential reason for this difference is that the regular season component of wins takes away from the relative importance of all-stars, but when number of all-stars is the only measure of regular season success, it becomes significant.

However, the regression results varied greatly when comparing these short-term success measures to number of championships, a long-term success measure. When number of championships is used as the measure of in-game success, the value of income and attendance drop off precipitously, more so in the operating income-omitted regression. An explanation for this observation is that some of the market and fan-engagement effects on franchise value captured by these two variables are also explainable through sustained success at a high level. In the case of attendance, perhaps the association between franchise value and attendance is partly a byproduct of historical success, which impacts both measures separately.

In comparing the operating income-inclusive regression to the operating income-omitted regression, major changes are observed in almost every variable. When removing operating income, the absolute value of the intercept increases by almost a full order of magnitude. This shift indicates that much of the baseline assumptions about franchise value are captured by operating income, and removal of this term shifts much of that weight to the intercept. There are positive shifts in income and population, but these shifts are minor. These shifts would indicate that the economic health and market of a team's city are not captured by operating income but are still highly significant in determining franchise value.

The number of all-stars metric also experiences a shift and loses a great deal of statistical significance when operating income is omitted. This shift would indicate that operating income explains a substantially different effect than number of all-stars does as it relates to franchise value. The opposite effect is observed with attendance, which is insignificant with operating income but highly significant without operating income. The strong significance of attendance in the absence of operating income suggests that much of attendance's effect on franchise value is also captured by operating income.

Based on these factors, these regressions are great for unpacking the explanatory variables and have great predictive power as well.. The models using the logarithm of franchise value are excellent for this purpose as well, but may not be the best for predictive purposes due to the manipulation of the dependent variable.

4.2.2 NBA

Table 6: NBA OLS Regression Results

In-Game Success Metric:	Wins	Chips	RctChip
Intercept	-3771.68 (386.08**)	-3019.3 (432.34**)	-3486.26 (372.41**)
Population	73.8 (12.34**)	74.48 (11.25**)	80.14 (11.61**)
Income	406.44 (63.95**)	324.26 (74.24**)	374.66 (55.74**)
Attendance	1349.2 (185.7**)	918.71 (158.58**)	1089.46 (185.3**)
All Stars	26.65 (47.24)	-29.58 (37.34)	-55.12 (40.12)
Wins	-7.11 (3.36*)	–	–
Championships	–	62.16 (10.63**)	–
Recent Championship	–	–	438.55 (137.76**)
B-P-G Test Statistic	9.38	14.85	12.79
Adjusted R-Squared	0.82	0.85	0.82

Note:
Chips = Championships; RctChip = Recent Championship; B-P-G = Breusch-Pagan-Godfrey
Scaled Statistics: Population (millions); Income (ten-thousands); Attendance (ten-thousands)
All standard errors are heteroskedasticity robust (HC1)
Breusch-Pagan-Godfrey test statistic determined before correction for heteroskedasticity
Standard errors are denoted in parenthesis below the regression coefficients
* Indicates statistical significance at the 5% level
** Indicates statistical significance at the 1% level

Table 7: NBA OLS Regression Results with log(Franchise Value)

In-Game Success Metric:	Wins	Chips	RctChip
Intercept	4247.0 (129.64**)	4590.07 (125.9**)	4345.88 (127.94**)
Population	38.23 (3.28**)	38.28 (2.71**)	40.41 (3.14**)
Income	120.5 (14.94**)	84.29 (16.97**)	109.54 (13.49**)
Attendance	835.34 (66.61**)	653.44 (54.15**)	746.08 (64.01**)
All Stars	26.03 (18.49)	7.06 (12.99)	-1.94 (14.59)
Wins	-2.42 (1.38)	–	–
Championships	–	27.6 (2.31**)	–
Recent Championship	–	–	151.36 (40.66**)
B-P-G Test Statistic	4.04	7.79	3.50
Adjusted R-Squared	0.91	0.94	0.91

Note:
Chips = Championships; RctChip = Recent Championship; B-P-G = Breusch-Pagan-Godfrey
Scaled Statistics: Population (millions); Income (ten-thousands); Attendance (ten-thousands)
All standard errors are heteroskedasticity robust (HC1)
Breusch-Pagan-Godfrey test statistic determined before correction for heteroskedasticity
Standard errors are denoted in parenthesis below the regression coefficients
* Indicates statistical significance at the 5% level
** Indicates statistical significance at the 1% level

The NBA OLS regression models in Table 6 demonstrate very similar trends to the NFL regression models. Population and income are both strongly significant in determining franchise value, indicating that city and market factors play a very large role in determining franchise value for the NBA. Income appears to be a much larger effect than population in the NBA models compared to the NFL models. A possible explanation for this trend is that the NFL has a much larger reach and average viewership than the NBA, so the effect of a team's metro area population would be exaggerated for NFL teams. The greater effect of income could be attributed to the glamor element of the NBA, where dozens of courtside seats and luxury suites are available for celebrities, attracting more wealthy consumers towards the sport. By contrast, the NFL is a more rugged sport with the playoffs taking place in the wintertime and many games played in extreme temperatures and inclement conditions. Additionally, attendance matters much more in the NBA than the NFL, with coefficients across all the models demonstrating strong statistical significance. A reason for this difference is that the NFL only has 17 regular season games with a much larger fan base, so the demand for game attendance is regularly high enough to meet the supply, which in this case would be stadium capacity. By contrast, the NBA has 82 regular season games, increasing the supply of games available for consumers to attend. This increased supply may not be met by fan demand under certain circumstances, leading to a higher variance in NBA game attendance. For this reason, consistently sold-out stadiums have a much greater impact on franchise value in the NBA than the NFL, which sells out much more regularly.

The biggest surprise in the NBA model is the lack of significance of all-stars on franchise value. The NBA is known as a "star-driven league" with strong pushes by NBA commissioner Adam Silver to market the players instead of the teams. This fact is demonstrated through social media following, where the two biggest NBA stars, LeBron James and Stephen Curry have 159 million and 56.3 million followers on Instagram, respectively. The NFL's biggest star, Patrick Mahomes, only has 6.8 million followers on Instagram[34]. This large discrepancy illustrates how the NBA is a star-driven league, but the results of these NBA models show that the number of stars does not have much of an impact on franchise value.

[34] Meta. (n.d.)

Perhaps the reason for the lack of impact of this measure is that only the biggest stars drive franchise value, while borderline all-stars have little impact on franchise value. Among the success-based metrics, championships and recent championship have the biggest impact on franchise value. The reason for this may be because the regular season is diluted over 82 games, decreasing the value of the regular season. Additionally, more players are engaging in "load management", a practice where players strategically rest games to prepare for the post-season. All of these trends place increased importance on the post-season, which as demonstrated through these models, has a much higher impact on franchise value. According to ESPN, regular season national television broadcasts average 1.7 million viewers in the 2023 season, while the 2023 NBA playoffs averaged 5.47 million viewers per game, capped off by the NBA finals averaging 11.64 million viewers[35]. This sharp increase serves as evidence for the increased importance of the post-season compared to the regular season, as these models indicate.

In the NBA OLS regression models in Table 7 that regress the logarithm of franchise value on the explanatory variables, the results are practically identical to those in Table 6. The heteroskedasticity is much lower in these models, as indicated by the Breusch-Pagan-Godfrey test statistic. Additionally, the R-squared is higher across all three models, indicating that the logarithm of franchise value is the better indicator for explanatory purposes.

4.2.3 NBA and NFL Combined

[35]Sim, J. (2023, June 14)

Table 8: NBA and NFL Combined OLS Regression Results

In-Game Success Metric:	Wins	Chips	RctChip
Intercept	-742.32 (198.01**)	-333.54 (180.27)	-597.79 (161.06**)
Population	50.75 (6.94**)	48.87 (6.5**)	51.38 (6.84**)
Income	254.35 (37.23**)	164.3 (39.86**)	243.89 (33.71**)
Attendance	75.38 (26.9**)	62.64 (23.45**)	72.78 (26.86**)
All Stars	18.46 (8.18*)	11.01 (6.67)	15.68 (7.76*)
Wins	3.07 (2.77)	–	–
Championships	–	89.99 (8.56**)	–
Recent Championship	–	–	251.28 (56.12**)
B-P-G Test Statistic	8.77	8.79	8.25
Adjusted R-Squared	0.80	0.84	0.81

Note:
Chips = Championships; RctChip = Recent Championship; B-P-G = Breusch-Pagan-Godfrey
Scaled Statistics: Population (millions); Income (ten-thousands); Attendance (ten-thousands)
All standard errors are heteroskedasticity robust (HC1)
Breusch-Pagan-Godfrey test statistic determined before correction for heteroskedasticity
Standard errors are denoted in parenthesis below the regression coefficients
* Indicates statistical significance at the 5% level
** Indicates statistical significance at the 1% level

Table 9: NBA and NFL Combined OLS Regression Results with log(Franchise Value)

In-Game Success Metric:	Wins	Chips	RctChip
Intercept	6173.58 (76.32**)	6407.61 (55.09**)	6289.46 (53.42**)
Population	24.75 (2.56**)	23.71 (2.34**)	24.87 (2.52**)
Income	89.21 (10.67**)	49.73 (10.81**)	85.26 (9.58**)
Attendance	48.87 (8.61**)	43.57 (7.17**)	48.05 (8.64**)
All Stars	13.01 (3.59**)	11.53 (2.78**)	13.44 (3.22**)
Wins	2.8 (1.34*)	–	–
Championships	–	40.6 (2.53**)	–
Recent Championship	–	–	120.71 (19.46**)
B-P-G Test Statistic	12.82	14.28	12.15
Adjusted R-Squared	0.88	0.91	0.88

Note:
Chips = Championships; RctChip = Recent Championship; B-P-G = Breusch-Pagan-Godfrey
Scaled Statistics: Population (millions); Income (ten-thousands); Attendance (ten-thousands)
All standard errors are heteroskedasticity robust (HC1)
Breusch-Pagan-Godfrey test statistic determined before correction for heteroskedasticity
Standard errors are denoted in parenthesis below the regression coefficients
* Indicates statistical significance at the 5% level
** Indicates statistical significance at the 1% level

The NBA and NFL combined OLS regressions, as seen in Table 8, in large part mirror the results of the NFL regressions. Population, Income, and Operating income are the most significant variables, with small changes in their coefficients between the operating income-inclusive and operating income-omitted models. Oddly enough, the number of all-stars is statistically significant at the 1% level for most of the models and at the 5% level in two of the operating income-omitted models. This observation is particularly unusual because in the NBA and NFL individually, the number of all-stars is not significant. The inconsistency between the league-isolated models and the combined model for this particular metric is noteworthy and very difficult to explain. Similar to the NFL model, attendance gains large degrees of significance once operating income is omitted, probably for similar reasons as the NFL-specific model. Similar to the NBA model, the success-based metrics remain relatively constant in significance between the operating income-inclusive and operating income-omitted models. The coefficient magnitudes do increase upon the omission of operating income, but this increase is proportional to the change in intercept value, as observed in the NBA-specific model. Like the NFL-specific model, operating income adds large degrees of heteroskedasticity to the model, making the operating income-omitted regression more efficient. The adjusted R-squared and intercept values indicate, however, that operating income does give predictive power to the model.

In the NBA and NFL combined OLS regression models in Table 9 that regress the logarithm of franchise value on the explanatory variables, the results are largely the same as those in Table 8. The only main difference is the insignificance of attendance, which is significant when regressing with the raw franchise value data. However, attendance is significant in the operating income-omitted models and in the league-specific models, so this change does not impact the outcome of this study. The heteroskedasticity is the same in these, as indicated by the Breusch-Pagan-Godfrey test statistic. Additionally, the R-squared is a little higher across all six models, indicating that the logarithm of franchise value is the better indicator for explanatory purposes.

4.2.4 NFL Team Fixed-Effects

Table 10: NFL Team Fixed-Effects Regression Results

Year Range	2002-2012	2013-2023
Intercept	-9079.17 (1465.26**)	-1723.54 (186.96**)
Population	1301.19 (333.25**)	284.75 (53.95**)
Income	930.17 (54.91**)	275.93 (16.89**)
Attendance	60.11 (13.54**)	70.36 (15.31**)
All Stars	21.59 (9.89*)	-0.67 (2.27)
Championships	212.33 (101.7*)	95.13 (16.33**)
B-P-G Test Statistic	2.10	2.66
Adjusted R-Squared	0.89	0.88

Note:
B-P-G = Breusch-Pagan-Godfrey
Scaled Statistics: Population (millions); Income (ten-thousands); Attendance (ten-thousands)
All standard errors are heteroskedasticity robust (HC1)
Breusch-Pagan-Godfrey test statistic determined before correction for heteroskedasticity
Standard errors are denoted in parenthesis below the regression coefficients
* Indicates statistical significance at the 5% level
** Indicates statistical significance at the 1% level

These NFL team fixed-effects models, as seen in Table 10, indicate a number of notable differences between the time period. Firstly, the intercept is much larger in magnitude in the earlier time period, which indicates that the coefficients for these models must be assessed accordingly to compensate for this large negative number. While statistically significant in most models, the estimator coefficient for championships and attendance is nearly identical between the early and late date ranges. This would indicate that these measures are much more important in the later time period due to the difference in the intercept. Population and Income have much higher estimator coefficients for the earlier time period, indicating that these factors were more important during that time than in present day. The estimator coefficient for all-stars also sharply decreased from the early to late time period and lost statistical significance, indicating that this factor mattered much more in the early time period as well. In addition, these models are much less heteroskedastic than the models which are not separated by year, even those ones have a time control built into the model. This decrease in heteroskedasticity is emphasized in the earlier time period, likely due to the linear growth pattern observed over these years in Figure 1.

4.3 Random Forest Regression Results

Table 11: Random Forest Relative Importances

	NFL	NBA	Combined
Population	0.05	0.12	0.07
Income	0.09	0.63	0.18
Attendance	0.05	0.03	0.06
All Stars	0.01	0.01	0.01
Wins	0.01	0.02	0.03
Championships	0.04	0.07	0.06
Recent Championship	0.0	0.02	0.01
Year	0.39	0.06	0.29
$Year^2$	0.36	0.06	0.29
Accuracy	92.08%	84.07%	89.5%
Training Set Size	503	239	742
Testing Set Size	168	80	248

Note:
All variable importance values for a given dataset sum 1.00
Accuracy is calculated as a product of mean absolute error between model predicted values and actual values

Figure 7: NFL Relative Variable Importance

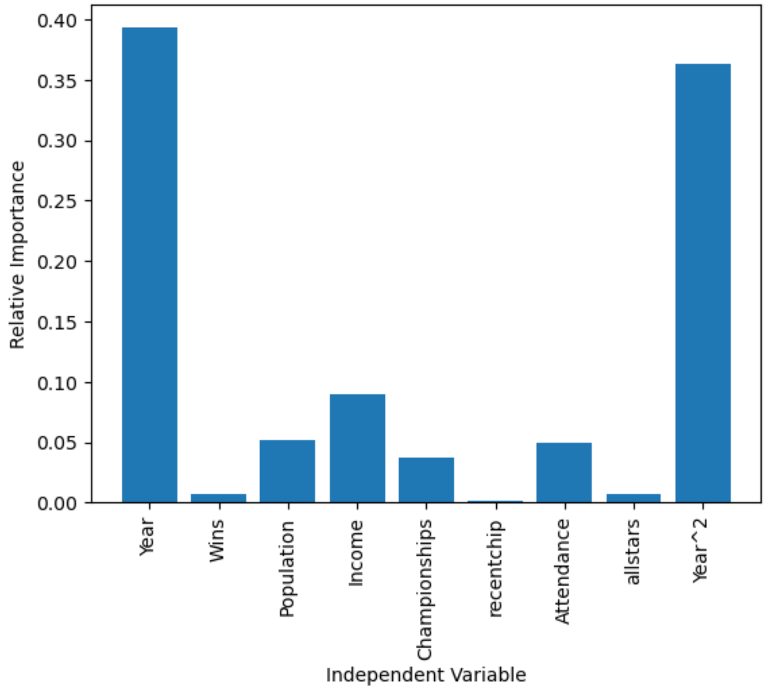


Figure 8: NBA Relative Variable Importance

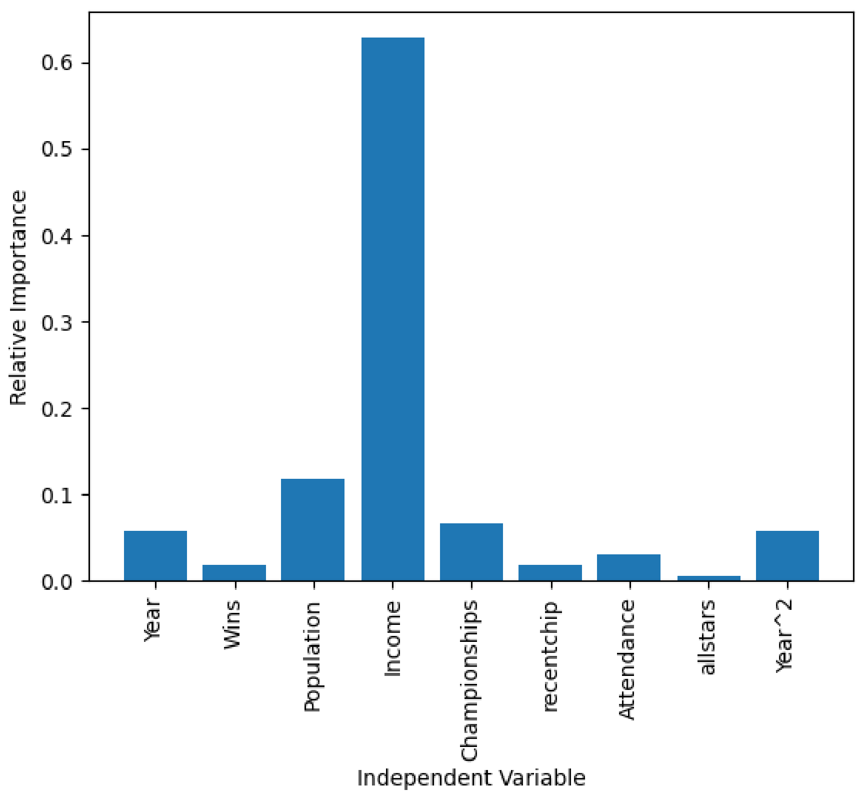


Figure 9: NFL and NBA Combined Relative Variable Importance

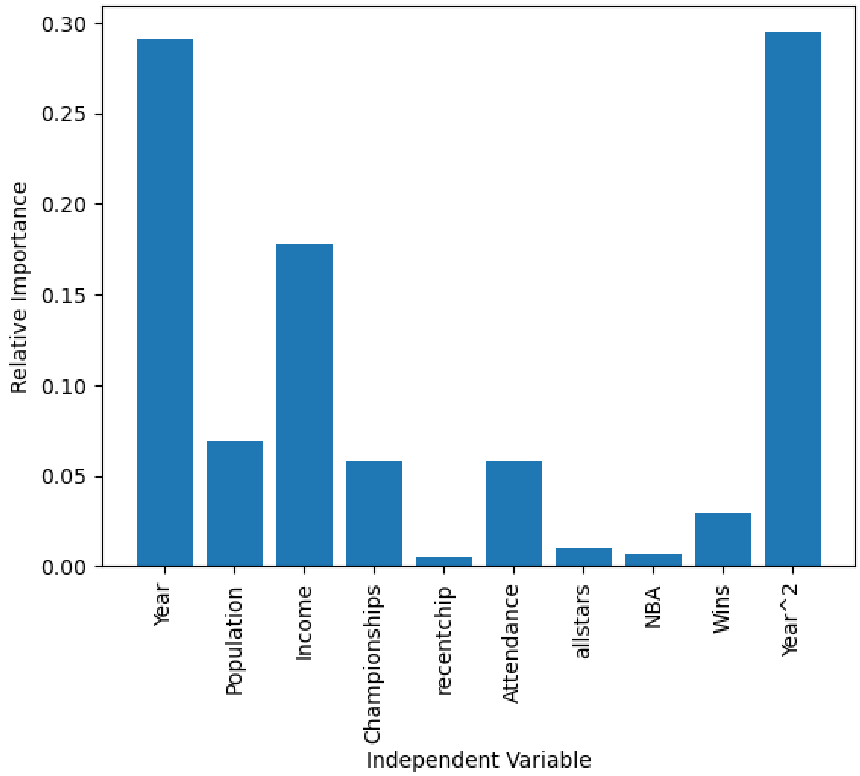
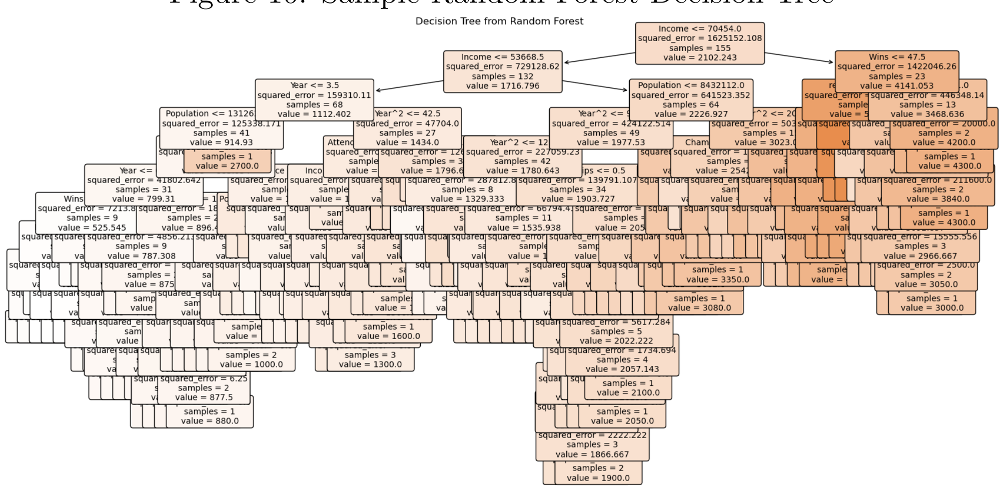


Figure 10: Sample Random Forest Decision Tree



Note:
This decision tree represents 1 of 1000 trees generated by the Random Forest model for the NBA. As a testing set is run through one of these trees, its value for a select independent variable determines which branch it travels down at each node. The terminal nodes at the bottom, or leaves, each contain an estimate for franchise value.

The theory behind using a Random Forest regression model as an alternative to OLS is that the machine learning component of the Random Forest algorithm gives more predictive power to the model. In analyzing the results of the Random Forest regression model for NBA and NFL franchise values in Table 11, similar results are found to OLS. In the NFL, population, income, attendance, and championships are the most important explanatory variables. In the NBA, population, income, and championships are the most important explanatory variables. In the combined dataset, population, income, attendance, and championships are the most important explanatory variables. The decreased importance of attendance in the NBA Random Forest model is noteworthy, since the opposite effect is observed in the OLS model, where the NBA places much more importance on attendance. However, the NBA Random Forest model only has an accuracy of 84.07%, which is much lower than the NFL and combined models. This lower accuracy measure is likely a byproduct of the low sample size of the NBA dataset compared to the other two datasets. With a dataset dating back to 2002, like the NFL, the NBA Random Forest model would likely have a higher accuracy. With that said, the accuracy of the other two models being near or above 90% is an indication that these models have very strong predictive power. Unlike an OLS regression, however, there is no mathematical equation that can take in a sample data entry and spit out a franchise value. This set must be inputted into the algorithm itself, where it is run through a set of decision trees, like the one generated for the NBA in Figure 10. While these models could be refined with higher sample sizes, these models have a high predictive power and could work very effectively for forecasting future franchise values as leagues add expansion teams and expect growth with current teams.

5 Discussion

5.1 Determinants of Franchise Value

5.1.1 Study Takeaways

The results of this study mostly align with the hypotheses outlined for each explanatory variable. In terms of statistical significance, only Wins, All-Stars, and Attendance are insignificant in select OLS models. The direction of each statistically significant explanatory variable aligns with the hypotheses outlined as well, as none of the measures have a negative impact on franchise value. The major takeaways from this study are that the team location and market size are paramount to the financial success of a sports franchise. The primary effects of locating in a larger city consist of more media attention, more lucrative sponsorship deals, and a larger fanbase. There also exist trickle-down effects of locating in a larger city, such as the appeal for notable players to sign with the franchise, more fame and lore associated with successful seasons, and a consistent fanbase during less successful seasons. All of these trickle-down effects are partially explained by other explanatory variables in this study but fall under the umbrella of metro area population and median personal income as well.

5.1.2 Franchise Relocation

The massive impact of location on the value of a franchise perhaps explains why franchises are so eager to relocate to bigger markets. In recent years, there have been several notable relocations within the NFL. The Oakland Raiders relocated to Las Vegas, the St. Louis Rams relocated to Los Angeles, and The San Diego Chargers also relocated to Los Angeles. Although the original location of these teams provide large enough markets for economic viability, there is no doubt that the newer locations for all three teams represents an upgrade from a market standpoint.

The recent NBA relocations tell a much different story, however. The only relocation from a smaller market to a bigger market was the New Jersey Nets moving to Brooklyn, but this hardly represents a market upgrade due to the proximity of Newark to the New York metro area. By many accounts, Newark actually exists in the New York metro area [36], rendering the market shift resulting from this relocation null and void. The other notable relocations have been the Charlotte Hornets moving to New Orleans and the Seattle Supersonics moving to Oklahoma City. Both relocations represent downgrades in market size. The aforementioned New Orleans Pelicans and Oklahoma City Thunder, as well as the Memphis Grizzlies and Milwaukee Bucks represent the four smallest markets among NBA teams. Perhaps these teams should take the lead from similar small market NFL teams and relocate to cities with larger populations and economic activity. Cities like Seattle, Las Vegas, or San Diego would be ideal locations for relocation, as their large market sizes are currently untapped by the NBA. As the NBA eyes possible expansion to 32 or 36 teams, the impact of market size must be taken into consideration, as the financial viability of a new NBA franchise is critical to its long-term survival and health.

The city of Las Vegas represents a very interesting location for new sports teams, as this gambling-centric city rapidly grows into the 21st century. The explosion of sports gambling represents a new age in professional sports and reach in Las Vegas provides a massive financial opportunity for sports leagues to partner with sports gambling companies. The best sure-fire way to acquire reach into the Las Vegas market is to relocate or expand a team there. The NFL held the most recent Super Bowl in Allegiant Stadium in Las Vegas, representing a turning point for the league's association with that city and sports gambling in general. As

[36] Bureau of Economic Analysis. (2023a, November 16)

seen in Figures 1 and 3, the growth of the NFL is far outpacing the growth of the NBA across the franchises in each league. Based on the results of this study, perhaps the NFL's willingness to tap into larger markets and specifically the Las Vegas market explains why its rate of growth and national presence dwarfs that of the NBA.

5.1.3 Star Power in the NBA

Where the NBA does have a considerable edge is in terms of star power. Some of the world's most famous athletes are NBA players, while the NFL's biggest stars are not as widely known both within the United States and internationally. The NFL's branding focuses on the teams and "the shield", which of course represents the NFL logo. Part of the reason for the NFL's focus on these larger entities is that players wear helmets and padding and share the field with many more players than in the NBA. However, the NBA still has a prime opportunity to leverage the star power and fame of their best athletes to grow the league financially. In recent years, NBA commissioner Adam Silver has done his best to lean into his players. Players of all calibers regularly receive fully guaranteed contracts, a rarity in the NFL. Television broadcasts regularly promote upcoming games with images and names of the players instead of the teams. For example, a recent game branded an upcoming Dallas Mavericks vs. Golden State Warriors with the caption "Steph x Luka" [37], referring to Steph Curry and Luka Doncic, two of the NBA's best players. Despite Adam Silver's best efforts to lean into his players, the continues to grow at a relatively slower rate compared to the NFL. The results of this study explain why this might be the case.

The metric for the number of All-Stars on a team did not significantly impact franchise value in the NBA-specific models. This finding would indicate that the marketing of stars does not result in financial gain for NBA teams nor the NBA league itself. With

[37] ESPN. (2024b)

the NBA being a "star driven league", this finding is very surprising and goes against the logical reasons for the player-centric marketing strategy. After all, the biggest marketing asset that the NBA has is its biggest stars. There could be two potential reasons for this finding: the emergence of "load management" and the growth of top-flight international star players. "Load management" is the practice of strategically benching a team's top players to ensure full health for the post-season[38]. Since many top players are sitting out key games, the marquee matchups between big NBA stars may not happen many nights, hurting the marketing efficacy of player-centric campaigns. The regularity by which key players miss games could be hurting the NBA's popularity among fans. Adam Silver has done his best to rectify this by instituting 65-game minimums for major awards, but the result of this change remains to be seen. The rise of international star players could also be hurting the league's player-centric campaigns since the vast majority of NBA fans are American. With the last five Most Valuable Player (MVP) awards, the most prestigious award in the NBA, going to international players, the impact of basketball's international reach is palpable. Though this phenomenon is difficult to quantify, having more international players may actually hurt domestic fan investment in the players since they are culturally so different from Americans. The contrast in fan investment between international and domestic players is evidenced by the recent playoff series between LeBron James' Lakers and Steph Curry's Warriors, which set a 30-year viewership record for second round playoff series. Both of these players are American and represent two of the sport's biggest icons.

5.1.4 Importance of Factors Over Time

Based on the franchise values over time in Figure 1, the NFL experiences a spike in league-wide growth around 2013, which not only led to increased franchise values, but also increased

[38] Load Management. (2021, November 27)

variance in franchise values. The results in Table 10 help to explain why that is the case. Evidently, there was a notable shift in the important factors relating to franchise value, with market factors like population and income becoming less important and team-specific factors like attendance and championships becoming more important. Even though nearly every variable is statistically significant in these models, it appears as if a team's valuation was merely a byproduct of their financial information, as indicated by operating income, combined with their location, as indicated by population and income from 2002-2012. The other factors played a minor role, if any. When taking operating income into account, attendance has a negative effect on franchise value during this early time period. An explanation for these measure having more importance is that the NFL media was very local during this time period, making each team's financial success more tied to their specific market. Nowadays, there are national television talk shows, social media, and other platforms that have made the NFL a more national product. This nationalization of the NFL has made the location of a franchise matter less than it did over a decade ago. The number of all-stars is a metric that loses significance in recent year, which is surprising considering the magnitude of this decline in importance. A reason for this surprising result may be that the talent in the NFL is more ubiquitous than it was in the earlier time period, making the difference in all-stars from the earlier to later time period not matter as much. Additionally, this increased depth in talent could have resulted in a smaller gap between Pro Bowlers and Non-Pro Bowlers (NFL all-stars), leading to this measure having less of an effect on franchise value. There has also been a decreased emphasis on the Pro Bowl in recent years, with many star players opting out of Pro Bowl selections due to nagging injuries or an apathy towards the event itself. This recent development has led to more mediocre players being selected to the Pro Bowl, which decreases the validity of this measure in capturing the presence on high-end talent on a team.

The increased emphasis on attendance and championships in recent years is indicative of a league-wide shift from the NFL being the primary determiner of franchise value to individual teams being the primary determiner of franchise value. Whereas before, teams were beholden to the revenue sharing of the NFL and their market, now, teams have more control over their own financial destiny. By putting a better product out on the field, their franchise value will go up more than an equivalent on-field difference would have made previously. This shift is also illustrated in Figure 1, with the standard deviation of franchise values by year increasing sharply as time progresses. Since these standard deviations capture the gap between the best and worst teams in terms of franchise value, this widening gap indicates that small differences between teams result in bigger differences in franchise value than ever before.

5.2 Predictive Capabilities of OLS and Random Forest

If the NFL or NBA ever decide to integrate expansion teams into their leagues, the financial viability of these new teams, as indicated by franchise value, would be a helpful tool to have. Additionally, teams attempting to relocate to a new city, sign a key player, or build a new stadium may want to weigh the economic benefits of these major changes before acting on them. For these reasons, molding the results of this study into a predictive model serves a practical purpose that would greatly benefit these two leagues. So what is the predictive power of the OLS and Random Forest models in this study?

Based on the OLS regressions, the operating income-inclusive models appear to have the best predictive power. However, operating income is nearly impossible to project for a new franchise or a franchise undergoing a singular major change. For this reason, the operating income-omitted models likely serve as better predictive models for this specific purpose. Among the three in-game success metrics, the models using championships appear to be the most significant and would be the most useful of the three for predicting franchise value. All three models are significant enough to work for this purpose. The only difficulty with using the OLS models for predictions is that they are fixed-effects models, so projecting franchise values out into the future may present some difficulties.

The Random Forest models may present a solution to this issue. The alternative approach of Random Forest in using a thousand distinct decision trees allows for the projection of franchise value to come with less variability and uncertainty. The reasons for including Random Forest into this study are that it provides an alternative regression tool to compare for analysis and it provides an alternative to OLS for predictive use. According to the accuracy scores of the model, the NFL and NBA-NFL combined models appear to have sufficient accuracy to function as effective predictive models. The NBA model likely requires a larger sample size to function as a predictive model. Additionally, the quadratic time trend may present issues with projecting far into the future, as league-wide growth will likely level off eventually.

5.3 Study Limitations

The restricted data availability for the NBA only allowed for the analysis of teams from 2012-2022, which is a much smaller sample than the NFL within this study. Having the equivalent 2002-2023 range for both leagues would have allowed for more significant and more accurate models across both methodologies. Additionally, television broadcast and social media data was not widely available across all teams for all the relevant years, so that data was not included in the study, even though it would have been helpful to explain other peripheral factors relating to fan engagement. Additionally, the speculative nature of many of the financial figures used to calculate franchise value in the first place makes it a suspect dependent variable to use. In an ideal world, a more robust measure would have been used instead of franchise value, but this was the best measure available to capture the desired aims of the study. Lastly, causality is very difficult to determine based on OLS regressions or Random Forest regressions and requires detailed analysis to prove. The inclusion of two separate regression methodologies allowed for causality to be inferred with increased confidence, but causality was not proven in this study.

5.4 Future Directions

In future studies, the inclusion of Major League Baseball (MLB) and the National Hockey League (NHL) would provide additional perspective into the nature of American professional teams sports as a whole. Expanding onto the global stage, exploring European soccer leagues, such as La Liga, the English Premier League, or Bundesliga would provide a compelling comparison between American professional sports franchises and international professional sports franchises. The unique characteristics of these leagues would also be an interesting theme to explore as it relates to the franchise values of teams within these leagues. Should more data become available, researching the impact of sports gambling and social media on franchise value could provide explanations for recent financial successes in both the NFL and NBA.

6 References

Alexander, D. L., & Kern, W. (2004). The Economic Determinants of Professional Sports Franchise Values. Journal of Sports Economics. https://doi.org/10.11771527002503251715

Appraisal Economics. (2021, February 17). Sports franchise valuation & appraisal services

- appraisal economics. Appraisal Economics - Valuation Services. Retrieved March 5, 2023, from https://www.appraisaleconomics.com/sports-team-valuation-appraisal/

Basketball Reference. (2022). 2022 NBA All-Star Game. Retrieved March 19, 2024,.

Basketball Reference. (2024a). 2022-23 NBA Season Summary. Retrieved March 19, 2024,.

Basketball Reference. (2024b). NBA & ABA Champions. Retrieved March 19, 2024,.

Biau, G., Scornet, E. (2016) A random forest guided tour. TEST 25, 197–227. https://doi.org/10.1007/s117 016-0481-7

Brandt, A. (2019, June 25). NFL salary cap is softer than the NBA's - sports illustrated. Sports Illustrated. https://www.si.com/nfl/2019/06/25/nfl-nba-salary-cap-guaranteed-contracts-money-franchise-tag

Bureau of Economic Analysis. (2023a, November 16). Personal Income by County, Metro, and Other Areas. Retrieved March 19, 2024,.

Bureau of Economic Analysis. (2023b, December 7). GDP by County, Metro, and Other Areas. Retrieved March 19, 2024,.

Burns, M. (2023, January 20). NBA Team Market Size Rankings. HoopSocial. https://hoop-social.com/nba-team-market-size-rankings/

Damodaran, A. (2024, March). THE SPORTING BUSINESS: VALUE AND PRICE. New York University - Stern.

ESPN. (2023). 2023 NFL Football Attendance. https://www.espn.com/nfl/attendance

ESPN. (2024). NBA Attendance Report - 2024. Retrieved March 19, 2024,.

ESPN. (2024b, March 25). Watch ESPN Upcoming Schedule.

Grage, D. (2022, February 2). Sports franchise values. 9WARE. Retrieved March 5, 2023, from https://www.9ware.com/AppliedMath/FranchiseValues

Grotkamp, L. (2024, February 5). NFL growth: Capturing the world's attention. SPORT-FIVE. https://sportfive.com/beyond-the-match/insights/nfl-international#: :text=This%20growing%20popularity%20can%20be,American%20football%20as%20a%20sport.

HBS Baker Library. (2024). Statista. Baker Library — Bloomberg Center. https://www.library.hbs.edu/find/databases/statista

Hong, J. et al. (2019). A House Price Valuation Based on the Random Forest Approach: The Mass Appraisal of Residential Property in South Korea. International Journal of Strategic Property Management 24(3), 140-152. https://doi.org/10.3846/ijspm.2020.11544

IBM. (n.d.). What is Random Forest? IBM. Retrieved March 5, 2023, from https://www.ibm.com/topics/random-forest

Kaushik, V. (2022, January 14). Sports franchise valuation considerations. EisnerAmper. Retrieved March 5, 2023, from https://www.eisneramper.com/sports-franchise-valuation-0122/#: :text=As%20is%20the%20case%20with,and%20operating%20expenses%20are%20unique

Kim J, Won J, Kim H, Heo J. (2021). Machine-Learning-Based Prediction of Land Prices in Seoul, South Korea. Sustainability; 13(23):13088. https://doi.org/10.3390/su132313088

Li, W. (n.d.). AN ANALYSIS ON THE MARKET STRUCTURE OF PROFESSIONAL

SPORT (thesis). European Association for Sport Management, Beijing. Retrieved March 19, 2024,.

Load Management. (2021, November 27). Physiopedia, . Retrieved 21:36, March 25, 2024 from https://www.physio-pedia.com/index.php?title=Load_Management&oldid=287698.

Meta. (n.d.). Instagram. https://www.instagram.com/

Ozanian, M., & Teitelbaum, J. (2023a, August 30). NFL team valuations 2023. Forbes. https://www.forbes.com/lists/nfl-valuations/?sh=17744d961738

Ozanian, M., & Teitelbaum, J. (2023, October 26). NBA valuations 2023. Forbes. https://www.forbes.com/lists/nba-valuations/?sh=2a211d386982

Pro Football Reference . (2024a). Super Bowl History. Retrieved March 19, 2024,.

Pro Football Reference . (2024b). 2023 NFL Pro Bowlers. Retrieved March 19, 2024,.

Pro Football Reference . (2024c). 2023 NFL Standings & Team Stats. Retrieved March 19, 2024,.

RealGM. (2024). NBA all-star game rosters. 2023-2024 NBA All-Star Game Rosters. https://basketball.realgm.com/nba/allstar/game/rosters/

Rodrigues, N.; Sequeira, N.; Rodrigues, S.; Shrivastava, V. (2019). "Cricket Squad Analysis Using Multiple Random Forest Regression," 2019 1st International Conference on Advances in Information Technology (ICAIT), Chikmagalur, India, pp. 104-108, doi: 10.1109/ICAIT47043.2019.8987367.

Sahota, H. (2022, January 31). Can't decide between a linear regression or a random forest? here, let me help. Medium. https://medium.com/artificialis/cant-decide-between-a-linear-regression-or-a-random-forest-here-let-me-help-ab941b94da4c

Schlepp, T. (2023, May 15). Lakers-warriors the most-watched NBA conference semifinals series in nearly 3 decades. KTLA. https://ktla.com/sports/lakers-warriors-the-most-watched-nba-conference-finals-series-in-nearly-3-decades/#: :text=The%20six%2Dgame%20series%2C%20which,results%20aren't%20entirely%20surprising.

Sim, J. (2023, June 14). NBA playoffs set new five-year ratings high with 5.47m average viewers per game tuning in. SportsPro. https://www.sportspromedia.com/news/nba-playoffs-finals-2023-tv-viewership-ratings-abc-espn-tnt-social-media/#: :text=2023%20NBA%20finals%2C%20which%20saw,peak%20audience%20of%2017.88%20million

Statista. (2023a, August). Franchise value of National Football League teams in 2023. Retrieved March 19, 2024,.

Statista. (2023b, August). Operating income of National Football League (NFL) teams in 2022. Retrieved March 19, 2024,.

Statista. (2023a, October). Average franchise value of NBA teams from 2001 to 2023. Retrieved March 19, 2024,.

Statista. (2023b, October). National Basketball Association teams ranked by operating income in 2022/23. Retrieved March 19, 2024,.

Sullivan, K. (2022, August 1). Die-hard or fair-weather fans? ranking all 32 NFL Fan Bases. Samford University. https://www.samford.edu/sports-analytics/fans/2022/Die-Hard-or-

Fair-Weather-Fans-Ranking-All-32-NFL-Fan-Bases

Thomas, T. (2023, October 25). Profitable partnerships: Athletes, nil and the evolution of TV revenue sharing. Sports Business Journal. https://www.sportsbusinessjournal.com/Articles/2023/10/25/oped-25-thomas

Vault. (2023). Sports. https://vault.com/industries/sports/structure