5-2022

Does Winning Matter? Purchase Decision Drivers of In-Game Attendance for the National Football League

Nicholas Griffiths

Does Winning Matter?

Purchase Decision Drivers of In-Game Attendance for the National Football League

by

Nicholas Griffiths

A Thesis Submitted to the Graduate Faculty of

St. Cloud State University

In Partial Fulfillment of the Requirements

for the Degree of

Master of Science

in Applied Economics

June, 2022

Thesis Committee:
Mana Komai Molle, Chairperson
King Banian
Luis Estevez

Abstract

Throughout this analysis we explore consumer demand for entertainment from live sporting events with a specific focus on NFL games. A widely established finding in previous studies of demand for live-game tickets for any major sporting event is that demand for tickets is inelastic, and ticket prices are set accordingly. We review literature stating this conclusion and find areas that this research could improve on, which takes the form of introducing our own contribution to the analysis field of variables that deal with the fans level of excitement and comfort during the live game experience. We perform several estimations and view descriptive statistics that lead up to our mixed-results conclusion that the factors that represent excitement are largely insignificant, and winning percentage or win total of a given NFL team is not a significant predictor of demand for live attendance. However, we have also discover several significant, comfort-based factors that previous research has left out, and prove that they are statistically significant areas of NFL demand prediction. These variables are with respect to the home Team’s climate and the stadium type where the team plays their home games. In general, our hypothesis returns mixed results with excitement factors not being a significant predictor of demand for NFL tickets, but showed that comfort-based factors were significant predictors of NFL attendance demand.

Keywords: Attendance, Demand, National Football League, Sports Economics

Table of Contents

List of Tables

List of Figures

Chapter I: Introduction

The National Football League (NFL) as an organization has been popular with American consumers for generations and is a traditional entertainment standby for a large portion of the general public. The NFL has many ways to consume their product, including local or national television broadcasts, radio broadcasts, streaming services and in person attendance. Some of these consumption options have been available the entirety of the League's history since its founding in 1920. Others, such as streaming services, have only recently come available for the fan as a consumption option. With the League's growing popularity and cement-like status in American culture, much can be gained from discovering factors that drive consumption of the game. Specifically, what factors are most important in a consumer's decision to purchase a ticket to a live NFL contest?

Previous research has established that in general, sporting organizations should be treated as monopolists. It has also shown that sporting leagues can and *will* price tickets for live contests in the in-elastic portion of the demand curve. While this prior research presents data that show sports teams do behave as monopolists when price setting, it fails to examine the in-game factors that make pricing in the inelastic portion of the demand curve possible. The NFL has calculated that consumers' willingness to pay for in-game tickets is extremely powerful; however, they have not discovered the properties of their product that create this steadfast demand. Simply put, the League knows that ticket prices are essentially secondary to the consumer when they are weighing the decision to purchase a ticket to an in-game NFL experience. What the NFL does not know is which other factors of the game determine attendance demand and which can be ignored. This analysis will seek to find significant predictors of NFL demand for in-game tickets outside of ticket prices, and evaluate to determine if a selected sample of other variables surrounding the in-game experience are statistically significant predictors of in-game fan attendance.

We will look at factors influencing the fans' perception of the team, the fans' expected level of comfort and excitement at the viewing experience, and the fans' perception of their team's level of competitiveness for the season. Analyzing these influential factors will help us answer questions such as:

- Is the quality of the on-field product a significant determinant of demand for in-person attendance?
- Does the perceived comfort of the in-game viewing experience matter?
- Does the perceived level of excitement matter for the fan?

Answering these questions will help determine the portions of the in-game experience the fan values the most and could lead to changes for the in-game entertainment experience in the future.

Thesis Statement

The success of the home team, quality of the game on the field, the fan's perceived comfort and the fan's perceived excitement are statistically significant predictors of consumer demand for in-game attendance at NFL contests. Consumers perceived excitement level and perceived comfort are significant predictors of demand for tickets to NFL games.

Chapter II: Literature and Historical Precedence Review

As mentioned in the Introduction above, many economic researchers have explored the phenomenon of sporting contests and consumer admission to them. Recent studies in the field of Sports Economics have ranged from analysis on the demand for consumer sporting tickets to the best way to maximize the return from a roster under the budget constraints of a capped player salary structure. We will focus on papers that are relevant to our proposed hypothesis that the fan-engagement pillars of winning and perceived comfort in the experience are also significant in predicting attendance demand. To set up precedence for this analysis, we will also focus on papers related to the economic structure of sport leagues like the NFL, and papers that predict demand for other ways of consuming sports entertainment.

To build a case for our analysis, it is also important to explain why we believe these experiential factors matter for the fan's consumption of the NFL. Even though the phrase "fan" is short for "fanatic", we must view them as rational consumers and rational economic actors for our analysis. This means that in general, fans of sport will gain more pleasure and excitement from a more competitive sporting contest. It has been established that leagues charge more for games with a higher level of expected competition, and we also know that fans still pay the higher prices for these likely-competitive games, per Chang et al. (2016) and Peiss & Kirstein (2014). Also, we know that fans will pay more for better seats with a better viewing angle of the contest, per Salaga & Winfree (2015) and ceteris paribus, fans prefer to sit in a seat closer to the field of play rather than one further away.

Comfort is a driver of the general economic experience for most consumers. People will pay exorbitant prices for a first-class seat in an airplane just to get a slightly larger seat for the duration of the flight. People will pay a valet service to retrieve their car for them when they can see it in the parking lot because waiting is easier than walking. Value-based pricing for services and portions of the entertainment experience that make the consumer more comfortable is the norm across most industries. For our purposes, we just need to assume that fans of sporting contests would prefer to be more comfortable versus less comfortable. In actuality, this means that fans would prefer to be protected from weather and the elements, as well as being able to view the game in an indoor environment in cold-weather NFL climates.

Further explanation of this assumption provides a good segue into our review of prior literature and analysis on factors of demand for in-game tickets to sporting contests. First, a paper by Welki and Zlatoper (1994) attempted to predict in game attendance for NFL games from the 1991 season. Their view of variable classification for attendance prediction built off of a previous survey analysis by Schofield from 1983 and is as follows:

> "In his survey Schofield (1983b) notes that the factors expected to affect attendance fall into four categories: economic variables (e.g. price, income), demographic variables (e.g. population), variables bearing on the expected quality of the game, the game environment, or the teams in question (e.g. team records), and variables reflecting influences or preferences not already accounted for by the other categories" (Welki & Zlatoper, 1994, p. 489).

In our analysis, we will test the predictiveness first and last two categories of variables in our regression models while controlling for location of the game and size the stadium where the game was played. In their modeling approach**,** Welki & Zlatoper (1994) consider several dummy variables such as "DOME" and "RAIN", representing a 0-1 indicator of whether or not the game occurred in a dome or while it was raining respectively. Their analysis found that the demand for

NFL tickets appeared to be inelastic and a winning home team spiked game day attendance. We will build off of their dummy variable approach and create a dataset with categorical variables for modeling with the goal of being more precise when modeling location and environment.

In another paper by Tainsky et al. (2016), the authors also tried to analyze the effects of excitement on the demand for NFL consumption. They looked at the last 4 weeks of the NFL season across a multi-year sample and attempted to predict demand for out-of-market Televised NFL games using the local franchise's chance to make the post season as an independent variable. While their analysis is implicitly controlling for the winning percentage of the local franchise via the chance said franchise will make the playoffs, their results also speak to the importance of excitement for the consumer. They found that the winning percentage of the local franchise was a statistically significant predictor for TV demand, but had diminishing marginal returns with a peak at 60%. This meant that the effect of local franchise success on the demand for viewing outside market games was when the local franchise had a 60% chance to make the playoffs. Intuitively, that estimated effect makes sense because at 60%, there is a greater than a coin-flip chance that the viewer's local franchise will make the playoffs, but it is far from guaranteed. This situation would drive viewership in fans because they are curious to see how the results of other games effect their local franchise's chance to make the playoffs. At a higher percentage of 95% or 99%, the fans are relatively certain that their franchise will be in the playoffs and are not concerned with other game outcomes because their team's playoff chances are independent of other teams' performance. Similarly, if a fans' local franchise has a low or even 0% chance of making the playoffs during the last 4 games of an NFL season, the local fan isn't concerned with the outcome of out-of-market games because their outcome is meaningless to the playoff chances of their local team. A paper by Schreyer et al. (2017) found similar results in their study of the role of uncertainty in demand for Televised soccer games in Germany and found increased uncertainty towards the outcome of the game increased viewing demand for the game. However, focusing on with Tainsky et al. (2016) analysis provides some excellent insight to the consumer preferences towards excitement and uncertainty in sports entertainment.

Tainsky's et al. (2016) work also provides good precedence for modeling the preferences of consumers by including several variables in their regression model that are good proxies for consumer excitement. Specifically, the authors included the winning percentage of the local team and the percentage of making the playoffs for both out-of-market teams entering the game. This inclusion of both teams' percentage to make the playoffs is noteworthy because it will provide precedence for another type of excitement metric in our analysis via the odds to win the Super Bowl championship at the beginning of a given season. Another analytical contribution from Tainsky et al. (2016) would be their set up of dummy variables representing unique season-weeks of NFL games, with N-1 dummy indicator variables for N=4 weeks of the season included in their analysis. With our analysis, will enhance their approach by building a panel dataset and including a categorical variable with one level for each week of the season. Our approach will be more appropriate for dealing with a panel dataset and a larger time sample than they utilized.

The papers we have been exploring so far represent that there is a sweet-spot in consumer's preferences with respect to the quality of the product and the uncertainty of the outcome. This is shown in several other studies such as the analysis by Rascher & Solmes (2011) and Borland & MacDonald (2003), and an original analysis of the *Uncertainty of Outcome* hypothesis by Rottenberg (1956), which was reviewed in Tainsky et al. (2016). All of these researchers provide a good summary of what uncertainty in outcome does to demand, but the research done by Burhan Biner in 2014, “Parity in Professional Sports When Revenues Are Maximized”, gives us insight into how sport leagues can achieve optimal uncertainty in their contests. Biner’s (2014) analysis explores how the distribution of assets across an organization can effect consumer’s preferences of the product they create. Biner (2014) performed regression analysis to predict demand for both Television viewing and in-person viewing of NFL contests using the two structural organization theories of League-wide parity and select dominant teams. Interestingly, his paper’s results with respect to our hypothesis that fans appreciate excitement and uncertainty are mixed. Biner (2014) finds that Television viewers are more interested in close games, while the fans who viewed the game in-person preferred larger margins of victory for the home team. He proposes a policy that falls in between complete parity and complete dominance, with a slight edge towards parity in general. Biner (2014) also posits that the allocation of on-field talent should be slightly skewed in favor of larger viewing markets.

Biner’s (2014) approach provides some good insights for the modeling of consumer preferences, but is also problematic because his approach to explaining the theory of consumer preferences. We believe his approach to explain demand for in-person attendance to games and demand for televised games using the same theory of consumer preference is flawed because the two experiences are not substitutes. Furthermore, the type of consumer is not wholistically comparable because as established previously, the ticket prices for NFL contests price some consumers out of live game attendance altogether. The lack of controlling for the type of consumer when modeling demand leads to flaws in Biner’s (2014) ability to summarize the optimal uncertainty level that would maximize demand for both ways of consuming the NFL’s product. One possible qualifier to this approach would be that Biner (2014) is modeling the optimal level of competition and league parity when the sum of all revenue streams are maximized. Because of the contract structure of the NFL and the fact that television revenues are split across all 32-teams, the League has a natural incentive to find the sweet-spot of parity and dominance that maximizes demand. Previous research has shown that in practice, sporting leagues in general will behave as cartels and individual teams will collude with one another to ensure that a baseline level of uncertainty-driven-entertainment keeps this pricing approach possible.

The cartel behavior that sporting leagues exhibit when they prominently price tickets to live contests in the inelastic portion of the demand curve is another important phenomenon to for our analysis. The paper by Coates & Humphreys (2007) titled "Ticket Prices, Concessions and Attendance at Professional Sporting Events" found that the demand for in-person attendance at sporting contests is largely inelastic with respect to price. Coates' & Humphrey's (2007) analysis included focused on attendance at Major League Baseball games, but also included games played in the National Football League and National Basketball Association. The authors found that the ticket pricing strategies for all 3 organizations in the inelastic portion of the demand curve led to revenue maximization, and this is consistent with monopolistic behavior. The paper compared the price elasticity of tickets and in-person experiential purchases such as concessions and parking location, and the authors noted that these secondary purchases were priced in the elastic portion of the demand curve. This provides another data point that ticket pricing for sporting contests is different than other types of consumer entertainment.

Another paper from Krautmann & Berri (2007) performed largely the same analysis. The authors evaluated different ticket and concession pricing strategies to determine whether or not it was possible that teams knowingly price tickets in the inelastic portion of the demand curve to maximize total revenue from concession sales. This result is consistent with Coates' and Humphreys' (2007) results, and also found that profit-maximizing owners of teams price tickets like monopolists to maximize total revenue across all revenue streams. Fort's (2004) work published in *Managerial and Decision Economics* provides a great summary of the phenomenon of inelastic pricing behavior. To open, he included a literature review that pointed out several papers that established that monopolistic pricing behavior takes place in the sports tickets market, but do not explain why. Fort (2004) brings up a previous paper from Cairns et al. (1985), and quotes the following passage:

> "The price elasticity results can be given two interpretations. First, that there is substantial evidence in favor of demand being highly price inelastic. Second, that the data problems of one form or another have led to the true relationship not being identified. A review of these problems tends to support the second interpretation" (Fort, 2004, p. 89).

Fort's (2004) focus on how pricing in the inelastic portion of the demand curve is possible provides a great basis for our analysis of the importance of perceived excitement and comfort in the in-game fan experience.

The analysis Fort and Quirk (1995) completed in 1995's *Journal of Economic Literature* is an exceptional example of how to search for these implicit trends in consumer preferences and demand for in-game tickets. The paper compiled data on the NBA, NFL, NHL & MLB to show how sports organizations functioned as cartels when pricing live game tickets using several statistical approaches. Fort and Quirk (1995) also explored how league competitiveness was consistent with the revenue-maximizing behavior of pricing tickets in the inelastic portion of the demand curve. They begin by pointing out that the standard deviation of the distribution for winning percentage across teams in a perfectly competitive sports league is *0.5/sqrt(M)*, where *M* is the number of games per team in the league's season. They show the +/- 2 standard deviations observations of the distribution of teams' winning percentage across the 1920-1990 NFL seasons to be (.220, .780), Quirk & Fort (1992), cited in Fort & Quirk (1995). The authors then construct the formula for a team's total revenue function, which begins with the talent levels and winning percentages of each team across the league. While our analysis will not include the construction of an NFL team's total revenue function, it will look at the effects on demand of variation in the factors that construct it.

Chapter III: Data Availability and Selection Methodology

A brief note on thesis topics & data selection: When selecting a hypothesis to test, data availability was a consistent problem. There were several iterations of a thesis revolving around a sports economics topic, and at the original time of writing in 2020, the plan was to combine NFL demand studies and analyze the economic impact of Covid-19 on the NFL. However, data around how Covid-19 effected league revenues and demand was difficult to locate, incomplete or non-existent altogether and the focus switched to researching the effects of missing games on sport league seasons. Research around the analysis of this topic lead to a list of missing games for the 'Big 4' sport leagues of the NFL, NHL, NBA & MLB, but overall lead to a small dataset with not enough variables to create a worthwhile analysis. While the effort to analyze the economic impact of Covid-19 on any entity is worthwhile, the data concerns lead us to scrap the original plan and pivot to a topic with more potential data sources. However, this pivot did not result in a total loss of research performed for the original topic because it provided good context for understanding the NFL's economic structure as a whole, and in a future section we will review the list of missed or rescheduled NFL games in our sample as an example of our dataset.

While working towards an analysis topic that had *both* good data and a testable hypothesis, we explored a potential analysis of NFL ticket pricing for the League Championship Superbowl game. Since the NFL's creation, 56 Superbowl games have been played in various cities across the US, from 1966 to 2022. Average ticket price data was located for the games, and this represented a dataset with 56 observations. However, after some quick descriptive plotting it was discovered that the analysis on Superbowl attendance would not lead to any interesting results as the Superbowl is routinely sold out. In fact, the only Superbowl games that were played with empty seats in the stadiums came in the 1960's & 1970's, which was a period before the League blossomed in popularity. A final effort was made to save the analysis topic when we searched to locate local or regional TV ratings for the Superbowl to add to our dataset. This search yielded historical national TV ratings, broken down by age-group and whether or not the consumer was watching on a cable network or a streaming service. After adding these viewership variables to the dataset, we realized that the initial problem was still not solved as there was not a way to structure the data with more than 56 observations, leading to another change of topic and shift of focus in data procurement.

The third iteration of our project focused on analyzing the different revenue streams of the NFL via their in-person attendance and TV Nielsen ratings performance for a sample of regular season games. Because 256 regular season games are played in each NFL season, this plan significantly increased our potential to locate additional data. However, we discovered that Nielsen's data was proprietary and also full of potentially skewed observations due to their internal struggles measuring the effects of streaming services**.** We concluded that focusing on in-person attendance provided the best chance to achieve data for a testable hypothesis and found just that via ProFootballReference.com.

When this new data source was discovered, a brief review of the NFL's history and structure was required to select the final dataset for our analysis. ProfootballReference.com had data available on attendance for the 1992-2019 seasons. During these seasons, the NFL consisted of 30-32 teams, spanning 3-4 Divisions in 2 Conferences. The League was re-aligned in 2001, (from its original patchwork structure of an un-even number of teams in each of its divisions), to a new perfectly even structure of 4 teams x 4 divisions x 2 conferences = 32 teams in total. These changes were fully implemented for the start of the 2002 season, which was also the Houston Texans inaugural season as the League's 32nd team. Interestingly to note, there was a significant amount of team movement since the St. Louis Rams moved from their Missouri home to Los Angeles, CA in 2016, and culminating with the Oakland Raiders moving to Las Vegas in 2020. But even with this movement, the 2002-2019 NFL Seasons provide a great basis for data analysis because the League structure of number of teams did not change, and the playoff structure and overtime rules remaining constant as well. The years 2002-2019 were actually one of the most consistent periods in NFL history because it is the longest period with no change in number of teams or organizational structure in League history. No teams folded or were added over a period of 18 seasons, and those 18 seasons will eventually become a significant time period in our final analysis. More historical and contextual information around the NFL can be found at ProFootballReference.com or NFL.com.

The data on raw attendance numbers did not include the stadium capacities for the games that were played, so efforts were made to research the stadium capacities so that percentage occupancy estimates would be available to our analysis. This led to extensive research on the NFL's home stadiums historical timeline, and discovery of a unique situation in stadium capacities. Starting in the early 2000s, the NFL launched an "International Series" marketing campaign that created several international games being played in Europe and Mexico. Since the first regular season game was played in London in 2007, there have been 30 games played in that city alone to current date. This created the need to determine the actual location of each game played vs working off of schedules and our raw attendance numbers, and we ultimately ended up compiling a list of all regular season NFL games from 1992-2019 played outside the US. However, the NFL's International Series is not the only reason a team would play a game outside of their home stadium.

This other reason is not as trivial as a marketing campaign, but in the case of necessity when weather or disaster strikes. The best example of this was the 2005 New Orleans Saints season that took place during the immediate aftermath of Hurricane Katrina. During the Saint's 2005 season, the team played their "home" games in a combination of Giants Stadium in New York, the Alamodome in San Antonio, TX and Tiger Stadium in Baton Rouge, LA. These are only a few examples of when a team played a game outside of their traditional home stadium.

With the plan to have attendance figures and percentage occupancies of stadiums as our dependent variable, we needed locate variables for our hypothesis factors of winning, excitement, and perceived comfort in the in-game experience. For this, we created each team's annual winning percentage from Win-Loss data from ProFootballReference.com. Once the team's winning information was added, we included data on the properties of the stadium where the game was played, the Super Bowl odds for each team for each season at the start of the season, and the Win-Loss Over/ Under and Season End Book Finish gambling lines from a Las Vegas sportsbook. We also created a historical variable representing the number of Super Bowl titles for each team at the start of a given season in the time sample. Data on the properties of the stadiums where the games were played include a categorical variable with levels for Indoor, Outdoor or Hybrid stadiums. Another variable was created for stadium location and climate in the form of a 2-level factor variable with "Warm Climate" and "Cold Climate" as it's two levels.

All of this data was compiled and created via research on historical NFL stadiums, their locations, team histories and ProFootballRefernce.com.

Another curious variable that we included in our analysis was the age of the home stadium at the time when the game was played. This was achieved by retrieving each home team's stadium opening season and then subtracting that figure from a given season in our sample. There is significant variation in the age of stadium, ranging from brand new to 96 years old, (Chicago's Soldier Field). The distribution of this rolling stadium age variable had an average age of 22.71 years over the 2002-2019 seasons in our potential sample, and a median of 16 years of age. This created a skewed distribution towards new stadiums, even with several hold out older stadiums behaving as quartile outliers in that 18-season time span. The motivation for including this variable for stadium age at the time of the game is important because of another "human" factor that comes in the form of nostalgia for a certain ballpark. This nostalgia factor could be important enough to draw fringe fans in groups of friends or family, and also could serve as a tourist draw for NFL fans in general. For our purposes, we classify this variable as a comfort-related variable to the fan experience and feel that this is justified for several reasons.

First, there is an un-explained phenomena in fandom around a fan's pride for their location's climate. Northern NFL cities have had many chances to transition to enclosed, climate-controlled stadiums, but have not done so because of the fans enjoyment of being in the elements. Fans perceive this as being with their team in the location of the battle, and if nothing else get pride out of 'outsiders' and visiting teams not being able to handle the cold. Two prime examples of this are in the NFC North, where the Green Bay Packers of Wisconsin have played outdoors at Lambeau field since 1957, and the Chicago Bears of Illinois have played at Soldier

Field since the 1920s. Both climates are extremely cold during the end of the NFL season and playoffs and yet both have become symbols of each team's history and are routinely sold-out. Secondly, there is a general nostalgia factor around a stadium that once saw a successful or noteworthy period in team history. Fans are nostalgic by definition because they keep seeking out their specifically chosen team of fandom versus seeking out rivals or competitors. This is noteworthy because we can check to see if nostalgia can outweigh physical comfort in modeling and if this variable is a significant predictor of fan attendance for NFL games.

After locating data on team histories, historical stadium capacities, special circumstances where games were played outside their traditional venues, and locating data for excitement factors, team competitiveness factors and perceived comfort factors, the final variable to add to our sample was a ticket pricing variable. Locating historical ticket prices for NFL games was extremely difficult and ultimately proved to be impossible. Data on the average price for a ticket to the Super Bowl was obtained, but no data for regular season ticket prices for NFL contests was located. To solve this problem, we looked to the Federal Reserve for a proxy metric for price. Data Series from the Federal Reserve Bank of St. Louis were located to serve as a ticket-price proxy and highlights include the "Personal Consumption Expenditures for Admissions to specified spectator amusements: Spectator Sports data series". This Personal Consumption Expenditure series represented a nominal pricing metric, and a Price Deflator Index for ticket prices was calculated from dividing the Nominal series first mentioned by a Real Personal Consumption Expenditure data series.

This is significant because the creation of a Price Deflator Index variable serves as the best proxy for actual prices available to model ticket price fluctuations. The data for our Price

Deflator Index was only available from 2002-2019, so our sample set was trimmed from 1992-2019 to 2002-2019 to accommodate this new variable. The trimming of the dataset from 1992-2019 to 2002-2019 also worked to our advantage with respect to NFL history and structure. Controlling for price is essential to understand the effects of perceived excitement and comfort in the in-game viewing experience, and the creation of the Price Deflator Index is a justifiable proxy to control for this price. The Real Personal Consumption Expenditure series that was the denominator in the creation of the Index, and has a base year of 2012. This means that for the 2012 NFL season, the value for the Price Deflator Index is 1 in our dataset. Overall, our dataset now includes data on NFL team locations, stadium types, attendance totals and historical capacities, win/loss records, prices, and the addition of variables unique to our analysis that deal with perceived excitement and comfort of the fans.

Chapter IV: Statistical Analysis

Data Compiling & Cleaning

To begin any analysis, we must first discuss the creation of the final dataset used for analysis, and any cleaning methods that were used to create it. In our case, we trimmed the dataset from the 1992-2019 NFL Seasons to 2002-2019 NFL Seasons to accommodate our pricing data. We then cleaned up any observations for home games that were played outside a given team's traditional home venue. This was completed by the research outlined above around non-traditional home field games, missing or rescheduled games and the capacities of whatever non-traditional stadium was used to play the NFL game in. This led to a dataset built off of 24 variables and 4608 observations. Each row is a unique observation in the dataset and represents a week of a team's NFL season. This means that each NFL team's season has 8 observations, and each NFL season in total has 8*32=256 observations. In our sample, we are including 18 seasons, from 2002-2019. A summary table of each variable in our dataset and their description appears below:

Table 1

Variable Descriptions:

Variable	**Type**	**Description**
Year	Numeric	Year NFL season took place in.
X2019Franchise	Factor, 32 Levels	NFL Franchise as of 2019, base level is "Arizona Cardinals."
Vkey	Text, Identifier	Identifier Key, not used in analysis. Represents every season across every year.

Variable	**Type**	**Description**
CensoredOccupancy	Numeric	The percentage occupancy of a game in a given season-week. Censored measurement of the total raw attendance number divided by the stadium capacity where values above 100% are replaced with 100%.
WeeklyAttendance	Numeric	The weekly attendance of a given NFL game in a given season-week.
WeeklyCapacity	Numeric	The weekly stadium capacity of a given NFL game in a given season-week.
Sellout?	Factor, 2 Levels	0-1 dummy variable where 1 represents a Percentage Occupancy value greater than or equal to 100%.
Home?	Factor, 2 Levels	0-1 dummy variable where 1 represents a "Home" game. All observations have a value of 1 after data cleaning.
SeasonWeek	Numeric	Integer value representing the Week number of the game.
Sum of HomeStadiumCapacity	Numeric	Home Stadium capacity of a given team. Observations with different values for "Sum of HomeStadiumCapacity" and "WeeklyCapacity" mean that the game took place outside of the team's traditional home venue.
StadiumName	Text, Identifier	The name of the stadium where the game was played.
StadiumLocation	Text, Identifier	The location of the stadium where the game was played.

Variable	**Type**	**Description**
StadiumType	Factor, 3 Levels	The type of the stadium where the game was played.
TeamClimate	Factor, 2 Levels	The climate of the location where the game was played.
Home	Numeric	Season total of attendance for all games played at home by a given team in a given NFL season.
Sum of SeasonEndWins	Numeric	The Season-Ending Win total for a given team for a given season.
Sum of SeasonEndLosses	Numeric	The Season-Ending Loss total for a given team for a given season.
Sum of SBTPS	Numeric	The Sum of Super Bowl Titles each team has won historical as of the end of the previous season.
Sum of SuperBowlOdds	Numeric	The odds to win the Super Bowl for each team for each season at the beginning of the season.
Sum of WLOU	Numeric	The Win-Loss Over Under gambling line for the season ending win and loss totals for each team for each season as of the start of each season.
SeasonEndBookFinish	Factor, 3 Levels	The booked results of each season for each team as compared to their WLOU. Factor levels are "Over", "Under" and "Push".
HighestPlayoffLevel	Factor, 6 Levels	The highest playoff level achieved for each team for each season. The Factor levels are "None", "WC" for Wildcard

Variable	Type	Description
		Round, "DR" for Divisional Round, "CC" for Conference Championship, "SB" for Super Bowl, "SBW" for Super Bowl Win.
PriceDeflator	Numeric	Price Deflator Index variable created from FRED data using the Nominal and Real values for Personal consumption expenditures for admission to spectator sports.
StadiumAge	Numeric	The age of a given stadium in a given year, at the start of the NFL season.

Categorical Variable Creation

Included in the above dataset, there are several categorical or factor variables. Variables that are 2 level factors are simple Bernoulli trial variables, and sometimes called dummy variables because they represent an either-or outcome. In our dataset, factor variables with 2 levels include Sellout?, Home?, & TeamClimate?. Factor variables with more than 2 levels in our dataset are, “X2019Franchise” representing Team, StadiumType, SeasonEndBookFinish & HighestPlayoffLevel, ranging from 3 to 32 factor levels.

When modeling factor variables effects in a regression model as we will do later, output will most often show N= L-1 levels where N is the number of factor levels shown, and L is the total number of factor levels that the variable has. This means that one level of each factor variable will not show as an effect in model output, and therefore seems as if it is omitted from the model. In reality, the hidden factor level can be thought of as the base factor level. When interpreting effects of multiple level factor variables, the estimated coefficient of a factor level is the effect on the dependent variable compared to the base factor level's effect on the dependent variable. This method of measurement puts more importance of setting the base level for some variables, and re-ordering the factor levels of the variable. Because our dataset is a pure balanced panel dataset we are not as concerned with this issue. In our modeling the base-levels of the multiple level factors are as follows.

- The "X2019Franchise" Team variable base level is the Arizona Cardinals.
- The StadiumType variable base level is an Enclosed Stadium.
- The SeasonEndBookFinish variable base level is Over.
- The HighestPlayoffLevel variable base level is Conference Championship.

Descriptive & Summary Statistics

Having compiled a robust dataset for analysis, there are many different summary statistics we can create. We have included some interactive charts below that detail historical win totals per each team per each year, as well as the historical Super Bowl odds of each team at the start of each season. Link to interactive charts below: Thesis-Analysis.knit

Figure 1

Annual Season End Win Totals Per Franchise

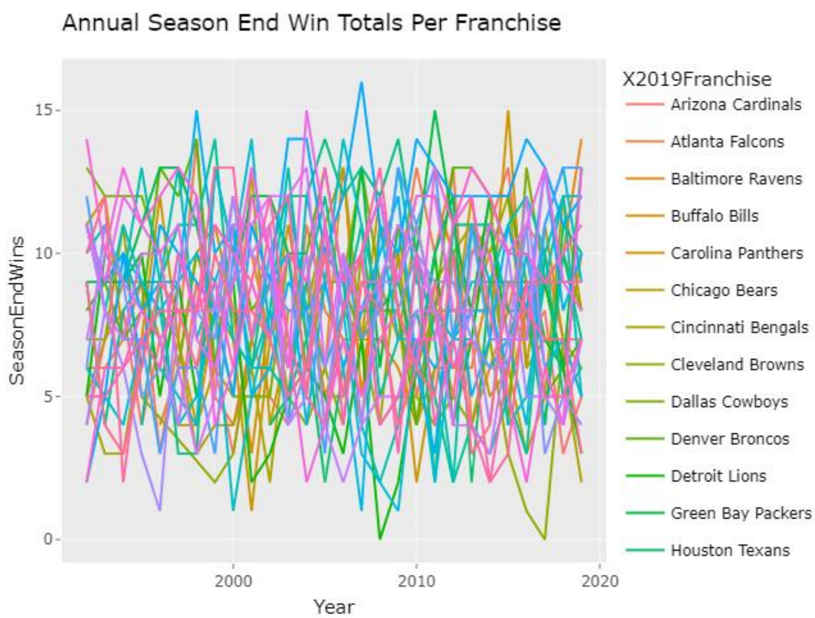
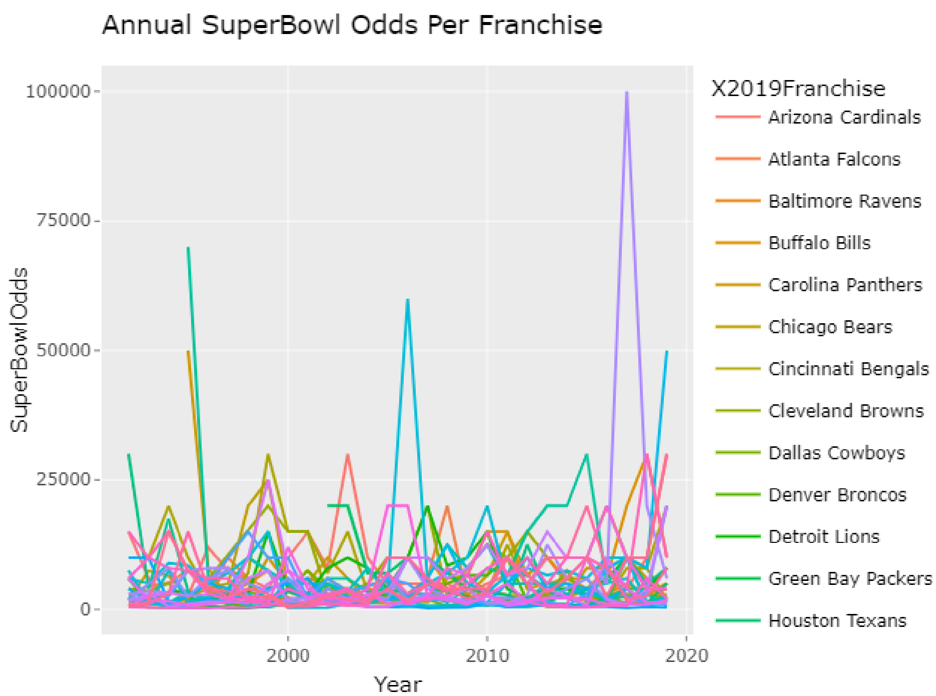


Figure 2

Annual Super Bowl Odds Per Franchise



Our analysis plan includes attempting to predict percentage occupancy for each NFL game in the sample based on a combination of the variables included in Table 1. There are multiple summary tables that can be created from this dataset, but to include a summary table of our dependent variable. Table 2 below represents each NFL seasons' average percentage of stadium seats occupied, across all NFL franchises.

Table 2

Season Average Percentage of Total NFL Stadium Seats Occupied

Year	**Average Percentage Occupancy Across all 256 Games**
2002	94.8581%
2003	95.4314%
2004	96.5633%
2005	96.6366%
2006	98.8565%
2007	98.6818%
2008	97.6678%
2009	95.1293%
2010	94.1276%
2011	94.9668%
2012	95.0559%
2013	96.0779%
2014	96.9212%
2015	96.3409%
2016	96.0263%
2017	94.8234%
2018	94.1997%
2019	93.5159%
Total Average	**95.8822%**

As we can see in the above table, we are predicting a dependent variable with a defined range above 0, and most observations are near the upper bound of 1. The NFL has traditionally been able to fill seats, and there are dips in 2009-2012 & 2019 where the percentage occupancy drops to 93-95%. This type of dependent variable can lead to problems in estimation if not modeled correctly. Next, we will transition to our overall modeling approach and initial model predictor variables.

Modeling

Our modeling approach to test for the significance of perceived comfort and excitement in NFL attendance demand begins with the properties of the dependent variable in the model of "Percentage Occupancy". We have considered many papers in our literature review above that perform regression analysis on NFL demand data to search for consumer preferences, but none of them included the variables related to excitement and comfort that we will be testing the effects of. These variables are as follows and appear in the table below.

Table 3

Dependent "Experiential Factors" variables unique to our model:

Unique Dependent Variables in Modeling	
Excitement Variables	**Comfort Variables**
Sum of SeasonEndWins	StadiumType
Sum of SeasonEndLosses	TeamClimate
Sum of SBTPS	StadiumAge
Sum of SuperBowlOdds	
Sum of WLOU	
SeasonEndBookFinish	
HighestPlayoffLevel	

Previous analysis has included Win-Loss records and win percentage, but has not included a combination of win totals and loss totals for each season for each team. We begin by building an initial Ordinary Least Squares (OLS) model as follows:

PercentageOccupancy ~ Year + X2019Franchise + WeeklyAttendance + WeeklyCapacity + SeasonWeek + Sum.of.HomeStadiumCapacity + StadiumType + TeamClimate + Home + Sum.of.SeasonEndWins + Sum.of.SeasonEndLosses + Sum.of.SBTPS + Sum.of.SuperBowlOdds + Sum.of.WLOU + SeasonEndBookFinish + HighestPlayoffLevel + PriceDeflator + StadiumAge

The full summary output table from our initial OLS model appears in the appendix, and also appears below.

Table 4

Base Ordinary Least Squares Estimation Results

Coefficients	**Estimate**	**Std. Error**	**t values**	**Pr(>|t|)**	**Significant notation**
(Intercept)	0.715	0.432	1.655	0.098	*
Year	0.000	0.000	0.385	0.700	
X2019FranchiseAtlanta Falcons	-0.008	0.001	-6.269	0.000	***
X2019FranchiseBaltimore Ravens	0.006	0.003	1.917	0.055	*
X2019FranchiseBuffalo Bills	0.002	0.003	0.720	0.472	**
X2019FranchiseCarolina Panthers	-0.019	0.002	-11.450	0.000	***
X2019FranchiseChicago Bears	0.011	0.004	2.937	0.003	***
X2019FranchiseCincinnati Bengals	0.006	0.003	1.936	0.053	*
X2019FranchiseCleveland Browns	0.012	0.003	3.587	0.000	***
X2019FranchiseDallas Cowboys	-0.006	0.002	-2.848	0.004	***
X2019FranchiseDenver Broncos	-0.001	0.003	-0.155	0.877	
X2019FranchiseDetroit Lions	0.018	0.002	7.288	0.000	***
X2019FranchiseGreen Bay Packers	0.004	0.004	0.957	0.339	***
X2019FranchiseHouston Texans	-0.011	0.001	-10.418	0.000	***
X2019FranchiseIndianapolis Colts	0.021	0.003	8.350	0.000	***
X2019FranchiseJacksonville Jaguars	-0.015	0.002	-9.271	0.000	***

Coefficients	Estimate	Std. Error	t values	Pr(>|t|)	Significant notation
X2019FranchiseKansas City Chiefs	-0.001	0.004	-0.377	0.706	
X2019FranchiseLos Angeles Chargers	-0.022	0.002	-14.350	0.000	***
X2019FranchiseLos Angeles Rams	0.015	0.002	6.758	0.000	***
X2019FranchiseMiami Dolphins	-0.003	0.002	-1.458	0.145	**
X2019FranchiseMinnesota Vikings	0.025	0.003	9.719	0.000	***
X2019FranchiseNew England Patriots	0.015	0.003	4.256	0.000	***
X2019FranchiseNew Orleans Saints	-0.009	0.001	-6.293	0.000	***
X2019FranchiseNew York Giants	-0.003	0.004	-0.837	0.403	
X2019FranchiseNew York Jets	-0.004	0.003	-1.082	0.279	
X2019FranchiseOakland Raiders	-0.019	0.002	-10.427	0.000	***
X2019FranchisePhiladelphia Eagles	0.010	0.003	2.863	0.004	***
X2019FranchisePittsburgh Steelers	0.010	0.004	2.786	0.005	***
X2019FranchiseSan Francisco 49ers	-0.001	0.002	-0.293	0.770	
X2019FranchiseSeattle Seahawks	0.010	0.003	3.026	0.002	***
X2019FranchiseTampa Bay Buccaneers	-0.012	0.002	-7.797	0.000	***
X2019FranchiseTennessee Titans	-0.008	0.002	-4.831	0.000	***
X2019FranchiseWashington Redskins	-0.007	0.004	-1.967	0.049	**
WeeklyAttendance	0.000	0.000	345.269	0.000	***
WeeklyCapacity	0.000	0.000	-182.940	0.000	***
SeasonWeek	0.000	0.000	0.273	0.785	
Sum.of.HomeStadiumCapacity	0.000	0.000	-0.383	0.702	
StadiumTypeHybrid	0.003	0.001	2.529	0.011	**
StadiumTypeOpen Air	0.008	0.002	5.409	0.000	***
TeamClimateWarm	0.021	0.003	8.075	0.000	***
Home	0.000	0.000	2.676	0.007	***
Sum.of.SeasonEndWins	0.000	0.001	-0.307	0.759	
Sum.of.SeasonEndLosses	0.000	0.001	-0.366	0.714	
Sum.of.SBTPS	-0.001	0.000	-2.387	0.017	**
Sum.of.SuperBowlOdds	0.000	0.000	2.611	0.009	***
Sum.of.WLOU	0.001	0.000	4.223	0.000	***
SeasonEndBookFinishpush	0.000	0.001	0.814	0.416	
SeasonEndBookFinishunder	-0.001	0.000	-1.749	0.080	*
HighestPlayoffLevelDR	0.000	0.001	-0.242	0.808	
HighestPlayoffLevelNP	0.001	0.001	1.373	0.170	
HighestPlayoffLevelSB	0.001	0.001	0.937	0.349	

Coefficients	Estimate	Std. Error	t values	Pr(>|t|)	Significant notation
HighestPlayoffLevelSBW	0.001	0.001	0.921	0.357	
HighestPlayoffLevelWC	0.002	0.001	3.008	0.003	***
PriceDeflator	-0.009	0.007	-1.356	0.175	
StadiumAge	0.000	0.000	4.426	0.000	***
Residual standard error: 0.007788 on 4553 degrees of freedom					
Multiple R-squared: 0.9917, Adjusted R-squared: 0.9916					
F-statistic: 1.01e+04 on 54 and 4553 DF, p-value: < 2.2e-16					

The model output above shows a model that has a very high R^2 value and shows that the Super Bowl title count of a team, the Super Bowl odds at the start of the season and the Win/ Loss season total gambling line at the start of the season for a team are all statistically significant at the alpha=0.05 level. It is likely that this model is overfitting with a R^2 value approaching 1. With a review of our residual plots, we can see that there are a few outliers, but overall, the residual plots show a uniform scattering of observations with skewed estimates towards the tails of our error distribution. This is revealed by studying the Theoretical Quantile- Quantile plot that appears in the appendix.

One other noteworthy observation from the OLS model output is the results of the "X2019Franchise" variable. Because we are analyzing a balanced panel dataset, (meaning a dataset with both time-based variables and variables have no association with time), we can see that the output has printed 31 levels our factor variable for NFL franchise. As discussed previously in the data methodology section, the level of the franchise variable that is not displayed is the Arizona Cardinals. Because of this large factor variable, the OLS regression could be overfitting the dataset as we are essentially attempting to predict the attendance for each team based on the franchise name. This Franchise factor variable is creating endogeneity in the data model because we have several other variables in the model that are representing properties of the team. We can leave the franchise variable out due to the fact that we have other factor and numeric predictors in the model that describe certain aspects of the team without using the actual name of the team as a predictor. We can most likely blame the overfitting of this model to the Franchise variable, and in further analysis we will predict Percentage Occupancy without it. For now, we keep it included but focus on other methods of estimation.

The OLS estimation is a good starting place, but since we are dealing with a balanced panel dataset, we should also try to estimate a Fixed Effects or a Random Effects model. It is important to distinguish the differences between OLS and Fixed or Random Effects estimation, and one key difference is how they deal with heterogeneity in data models. Heterogeneity is where different outcomes are observed using the same set of predictions, meaning that not all dependent observations are predicted using the same estimation. To summarize, fixed effects is the most robust as it removes the heterogeneity entirely because a fixed effects model only deals with known effects that have predetermined, non-random levels of effects in estimation. A Fixed Effect Model can remove heterogeneity because it can remove any time-invariant effects in the model, which leads to less potential for random error. A Random Effects model should be used when we do not have defined levels in estimation. In our data model, we should use Random Effects estimation because we are focused on estimating the effects of random variables and have a random independent variable in Percentage Occupancy. A Random Effects model using the same coefficients as our starting OLS estimation was ran and a snippet of the output appears below.

Table 5

Base Random Effects Estimation Summary

Total Sum of Squares: 9.7858
Residual Sum of Squares: 0.13608
R-Squared: 0.98609
Adj. R-Squared: 0.98593
Chisq: 322857 on 54 DF, p-value: < 2.22e-16

Note. Full table available in the Appendix.

We can see from the model summary output above that the inclusion of the Franchise variable led to overfitting in the Random Effects estimation as well with an R^2 value of 0.986. In our second round of modeling, we will remove the Franchise variable so overfitting is no longer an issue. However, because we are using an independent variable between 0-1 and modeling a percentage of stadium occupancy for a given game, we must also check one final model type before adjusting our predictor variables.

There are many types of regression that can be used to perform a data model, and each type of regression has uses that it is best suited for. Based on the properties of our dependent variable, we will estimate a Tobit regression model and check its performance against the others. When the dependent variable of a regression model is censored in some way, such as zero-inflated or upper-bound, a Tobit estimation should be completed. Tobit models are used to handle data that is censored outside of an upper or lower boundary, that can be set at any value. The base Tobit model was partially developed from the Logit & Probit models that we know estimate likelihood values between 0-1 to represent the probability of a certain Bernoulli outcome occurring. The difference between Logit/ Probit and Tobit is that Tobit was specifically developed to deal with censored estimates and uses the same type of probabilistic estimation on the censored data. Our dependent variable is technically censored on both the upper and lower boundaries, but we are most concerned about the upper bound. This led us to estimate a Tobit model using the same coefficients as our OLS and Random Effects models to check the Tobit's performance compared to our previous estimations. An upper bound Tobit regression model was estimated, and a snippet of the results appears below.

Table 6

Base Tobit Estimation Summary

Coefficients	**Estimate**	**Std. Error**	**t values**	**Pr(>|t|)**	**Significant notation**
(Intercept):1	1.3010	0.5038	NA	NA	
(Intercept):2	-4.8400	0.0122	-397.5480	0.0000	***
Year	-0.0002	0.0003	-0.8940	0.3712	
X2019FranchiseAtlanta Falcons	-0.0134	0.0016	-8.5530	0.0000	***
X2019FranchiseBaltimore Ravens	0.0185	0.0036	5.0960	0.0000	***
X2019FranchiseBuffalo Bills	0.0120	0.0038	3.1850	0.0014	***
X2019FranchiseCarolina Panthers	-0.0154	0.0019	-8.2990	0.0000	***
X2019FranchiseChicago Bears	0.0216	0.0042	5.1570	0.0000	***
X2019FranchiseCincinnati Bengals	0.0164	0.0035	4.7080	0.0000	***
X2019FranchiseCleveland Browns	0.0203	0.0036	5.5550	0.0000	***
X2019FranchiseDallas Cowboys	-0.0061	0.0027	-2.2450	0.0248	**
X2019FranchiseDenver Broncos	0.0114	0.0037	3.0570	0.0022	***
X2019FranchiseDetroit Lions	0.0184	0.0026	7.0830	0.0000	***
X2019FranchiseGreen Bay Packers	0.0130	0.0042	3.1160	0.0018	***
X2019FranchiseHouston Texans	-0.0090	0.0012	-7.3810	0.0000	***
X2019FranchiseIndianapolis Colts	0.0261	0.0028	9.2650	0.0000	***
X2019FranchiseJacksonville Jaguars	-0.0114	0.0017	-6.5840	0.0000	***
X2019FranchiseKansas City Chiefs	0.0072	0.0038	1.8900	0.0587	*
X2019FranchiseLos Angeles Chargers	-0.0176	0.0016	-10.9180	0.0000	***
X2019FranchiseLos Angeles Rams	0.0169	0.0024	6.9330	0.0000	***
X2019FranchiseMiami Dolphins	-0.0029	0.0021	-1.3560	0.1751	
X2019FranchiseMinnesota Vikings	0.0251	0.0028	8.8350	0.0000	***
X2019FranchiseNew England Patriots	0.0291	0.0043	6.7390	0.0000	***
X2019FranchiseNew Orleans Saints	-0.0150	0.0016	-9.2060	0.0000	***
X2019FranchiseNew York Giants	0.0062	0.0039	1.5890	0.1120	**
X2019FranchiseNew York Jets	0.0044	0.0037	1.1900	0.2340	*

Coefficients	**Estimate**	**Std. Error**	**t values**	**Pr(>|t|)**	**Significant notation**
X2019FranchiseOakland Raiders	-0.0130	0.0021	-6.1560	0.0000	***
X2019FranchisePhiladelphia Eagles	0.0207	0.0037	5.6360	0.0000	***
X2019FranchisePittsburgh Steelers	0.0217	0.0042	5.2200	0.0000	***
X2019FranchiseSan Francisco 49ers	0.0007	0.0035	0.2140	0.8306	
X2019FranchiseSeattle Seahawks	0.0203	0.0036	5.6810	0.0000	***
X2019FranchiseTampa Bay Buccaneers	-0.0079	0.0018	-4.5240	0.0000	***
X2019FranchiseTennessee Titans	-0.0073	0.0022	-3.3420	0.0008	***
X2019FranchiseWashington Redskins	0.0129	0.0041	3.1180	0.0018	***
WeeklyAttendance	0.0000	0.0000	315.7810	0.0000	***
WeeklyCapacity	0.0000	0.0000	-170.4270	0.0000	***
SeasonWeek	0.0000	0.0000	-0.3910	0.6960	
Sum.of.HomeStadiumCapacity	0.0000	0.0000	0.6390	0.5227	
StadiumTypeHybrid	-0.0020	0.0013	-1.4980	0.1340	**
StadiumTypeOpen Air	-0.0009	0.0017	-0.5340	0.5935	
TeamClimateWarm	0.0266	0.0029	9.2800	0.0000	***
Home	0.0000	0.0000	0.6980	0.4853	
Sum.of.SeasonEndWins	0.0011	0.0008	1.4670	0.1424	
Sum.of.SeasonEndLosses	0.0012	0.0008	1.5310	0.1257	
Sum.of.SBTPS	-0.0013	0.0005	-2.6370	0.0084	***
Sum.of.SuperBowlOdds	0.0000	0.0000	2.6090	0.0091	***
Sum.of.WLOU	0.0006	0.0001	3.9430	0.0001	***
SeasonEndBookFinishpush	-0.0006	0.0007	-0.9290	0.3527	
SeasonEndBookFinishunder	-0.0006	0.0005	-1.0900	0.2755	
HighestPlayoffLevelDR	0.0010	0.0007	1.4570	0.1451	
HighestPlayoffLevelNP	0.0004	0.0007	0.5130	0.6078	
HighestPlayoffLevelSB	0.0014	0.0010	1.3000	0.1936	
HighestPlayoffLevelSBW	0.0017	0.0011	1.6530	0.0982	*
HighestPlayoffLevelWC	0.0020	0.0007	2.9290	0.0034	***
PriceDeflator	0.0003	0.0082	0.0320	0.9748	
StadiumAge	0.0001	0.0000	3.2510	0.0012	***
Names of linear predictors: mu, loglink(sd)					
Log-likelihood: 10984.52 on 9160 degrees of freedom					
R-squared = 0.9905546					

Consistent with the other two methods of estimation, the Tobit model with the Franchise variable included has overfitted the data with an R^2 value of 0.991. Next, we will estimate a new model using an adjusted set of predictors in the hopes to solve the overfitting issue observed thus far.

For our analysis, the variables that will always be included in the model are the ones that appear in Table 3 above, and represent the factors of perceived excitement and comfort for fans at live NFL games. Aside from these terms and an intercept estimated for our model, any term can theoretically be excluded from our prediction set. Some variables such as the created Price Deflator variable should be viewed as more important that others, as they hold value in making our estimates more robust. An estimate for percentage occupancy of an NFL stadium that doesn't include some form of price variable would be difficult to justify, and any model without some form of ticket price included would most likely suffer from omitted variable bias.

The process of class selection for a model can range in approach and style based on the dependent variable and the 'problem' we are solving for, but overall, there are some useful metrics to use to compare model performance. Based on the type of data and the previous output examples from the OLS, Random Effects and Tobit models, we are going to proceed forward with the estimation of Tobit models only. Being that Tobit models are specifically designed to estimate our type of dependent variable and all 3 modeling types had R^2 values exceeding 0.98, we are justified in limiting our estimation technique to Tobit regression. Next, we will begin the process of trimming variables from our original Tobit model and re-estimate.

The variable selection process for our final model begins with removing the Franchise variable that was causing overfitting. With that removed, we should next focus on any variables that could potentially cause endogeneity. Simultaneity occurs when variables in a predictive model are correlated with the dependent variable, meaning that they have correlated variance. Another issue of regression models is endogeneity, or where an explanatory variable is correlated with the error term, (or residuals), of the model. To control for both phenomena, we will remove any predictor that isn't predicting a unique aspect of attendance demand. This process led to the removal of the "WeeklyCapacity" and "WeeklyAttendance" variables, representing the capacity of the stadium in which a given NFL game was played, and the raw number figure for total attendance at a given NFL game respectively. These variables were removed because they do not offer some unique contribution to our understanding of attendance demand, and because we are including the total number of fans that went to a home game at some point in each season for a given team. This "Home" variable allows us to model the percentage of occupancy at a game without including the capacity of the stadium because we are including the season-total number of people who went to a game in that stadium. Given that we know NFL games are routinely sold-out or near it, this "Home" attendance total serves as a proxy for stadium capacities or overall market size.

Another variable that we have not discussed yet is the StadiumAge variable. The effects in your summary tables show as 0.0001, and that the p-value is 0.0012, making it a very significant predictor. This means that for each one-year increase in age, ceteris paribus there is a 0.01% increase in percentage occupancy at NFL games. This result is consistent with our initial hypothesis that there is a nostalgia factor around a certain stadium for fans, and that the age of the stadium could eventually become a tourist attraction itself. This could be an important measure for teams when they are weighing the decision to build a new stadium or continue to use an older stadium with the hopes of it one day becoming a historic building for the team. An example of a stadium that could be considered in this sweet spot of fresh nostalgia would be Arrowhead Stadium in Kansas City, which was built in 1972. We will leave the stadium age at time of game variable included in the final model because of its high-level of significance.

Because of the balanced panel structure of our dataset, and the fact that in our case "Year" could be viewed as either a factor or a numeric variable, we have decided to remove the "Year" variable as well. Additionally, the "Year" estimator was not significant at any traditional alpha level in any of the prior estimations with our base model. We have kept "SeasonWeek" in the model to represent any variation in occupancy percentage due to demand changes across the 17 weeks of an NFL season. There are several reasons why we believe that this variable should be included to represent some sort of time-variant effect. Season-ending contests that are otherwise meaningless to a home team either due to having a post-season berth locked-up or a zero percent chance to make the playoffs would be two such examples. Having determined our list of predictors to include in our adjusted model, we will move to results and discussion next.

Chapter V: Results and Discussion

Our next section presents our updated Tobit model, with Terms specified as below:

PercentageOccupancy ~ Sum.of.SuperBowlOdds + Sum.of.WLOU + Sum.of.SeasonEndWins + Sum.of.SeasonEndLosses + SeasonEndBookFinish + HighestPlayoffLevel + Sum.of.SBTPS + Sum.of.HomeStadiumCapacity + StadiumType + TeamClimate + Home + Year + SeasonWeek + PriceDeflator + StadiumAge

The above regression equation shows our updated model with the "WeeklyCapacity", "WeeklyAttendance", "X2019Franchise" & "Year" predictors removed. We estimated a Tobit Type I Model with an upper bound censor at 1 and received the following output below:

Table 7

Adjusted Tobit Estimation Results

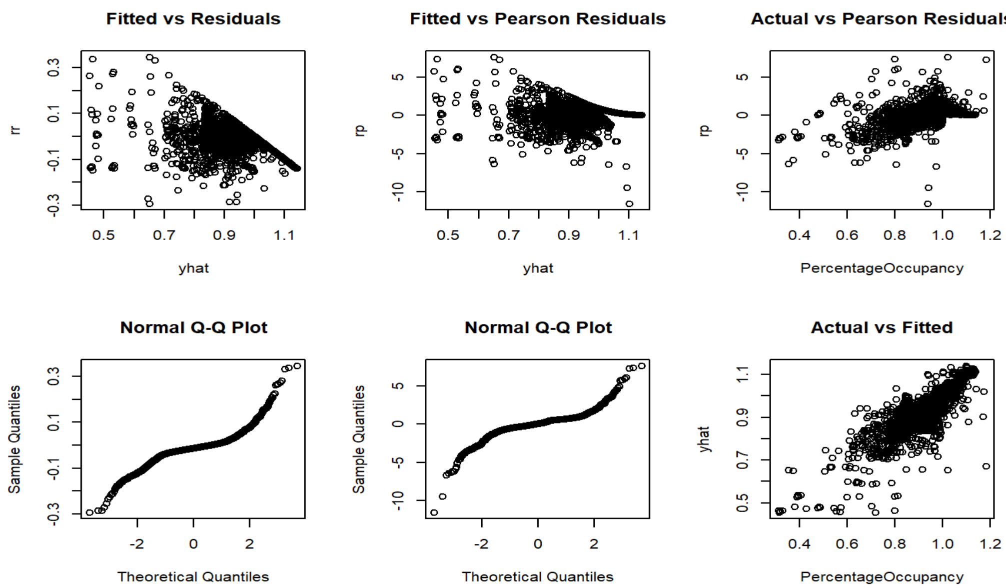
Coefficients	**Estimate**	**Std. Error**	**t values**	**Pr(>|t|)**	**Significant notation**
(Intercept):1	1.6090	2.801	NA	NA	
(Intercept):2	-3.0790	0.013	-238.366	0.000	***
Sum.of.SuperBowlOdds	0.0000	0.000	0.714	0.475	
Sum.of.WLOU	0.0007	0.001	0.961	0.337	
Sum.of.SeasonEndWins	0.0040	0.004	0.956	0.339	
Sum.of.SeasonEndLosses	0.0027	0.004	0.661	0.509	
SeasonEndBookFinishpush	-0.0001	0.004	-0.016	0.987	
SeasonEndBookFinishunder	0.0029	0.003	0.981	0.326	
HighestPlayoffLevelDR	0.0005	0.004	0.123	0.902	
HighestPlayoffLevelNP	-0.0007	0.004	-0.173	0.862	
HighestPlayoffLevelSB	0.0087	0.006	1.556	0.120	
HighestPlayoffLevelSBW	0.0072	0.005	1.321	0.187	
HighestPlayoffLevelWC	-0.0003	0.004	-0.067	0.946	
Sum.of.SBTPS	-0.0019	0.000	-3.940	0.000	***
Sum.of.HomeStadiumCapacity	0.0000	0.000	-83.707	0.000	***
StadiumTypeHybrid	0.0113	0.003	3.526	0.000	***
StadiumTypeOpen Air	0.0029	0.002	1.339	0.181	
TeamClimateWarm	-0.0049	0.002	-2.945	0.003	***

Coefficients	Estimate	Std. Error	t values	Pr(>|t|)	Significant notation
Home	0.0000	0.000	92.834	0.000	***
Year	-0.0004	0.001	-0.269	0.788	
SeasonWeek	-0.0007	0.000	-4.767	0.000	***
PriceDeflator	0.0103	0.045	0.227	0.820	
StadiumAge	-0.0001	0.000	-1.505	0.132	
Names of linear predictors: mu, loglink(sd)					
Log-likelihood: 4845.331 on 9193 degrees of freedom					
R-squared = 0.7625348					

We see from the above Tobit regression results that “SeasonWeek”, “Home”, “Sum.of.HomeStadiumCapacity” and “Sum.of.SBTPS” are all significant predictors at the alpha = 0.01 level. However, the coefficient of “Sum.of.SBTPS” isn’t consistent with what we would expect from our hypothesis that winning matters. The ultimate level of winning is a Super Bowl championship, yet our model is suggesting that for each 1 unit increase in Super Bowl Wins, there is a ~0.2% reduction in occupancy percentage.

Figure 3

Residual Plot Matrix from Adjusted Tobit Estimation



The removal of the Franchise variable and others in our model led to a reduced R^2 value, however a R^2 value of 0.763 is still indicative of a good performing model. Also, we see that the “Warm” level of “Team Climate” and the “Hybrid” level of “Stadium Type” are also significant at the alpha = 0.01 level. A quick check of the residual plots shows us that our model is predicting the dependent variable consistently across it’s defined range, and that we only have patterns in our residuals thanks to the censoring of the Tobit model. This can be seen in the “Fitted vs Residuals” chart below, where there is a clear diagonal asymptote in residuals due to our censorship of the dependent variable. A look at the Actual vs Fitted chart shows a uniform scattering of points, similar to what we would expect from a model with an-uncorrelated error term or other estimation problems such as endogeneity. Next, we will perform a final check of our Tobit results as we compare them to the results of the same regression equation estimated in OLS and Random Effects settings. Using the regression output below, we can see that overall, the significant terms are the same across our modeling methods. The primary exception being the StadiumAge variable, as significant in Random Effects & OLS, and not significant in Tobit estimation. Additionally, the sign the coefficient changes in StadiumAge, the absolute value of the effect is very small regardless of estimation technique used. This would imply a relatively low impact on fan’s demand of NFL tickets overall. The other exception in consistent estimation being the Tobit estimate for Super Bowl titles count per team, which also has a coefficient sign that we would not expect. These results show that Tobit provides the best estimation technique, because we do not see large differences in estimation of significant terms. Because of this similarity among model results, we should stick with the Tobit type estimator that is best suited to handle our kind of dependent variable.

Table 8

Adjusted Predictor Set Model Coefficients and Standard Errors

	OLS		**Random Effects**		**Upper Bound Censored Tobit**	
Coefficients	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**
(Intercept):1	0.9045	2.2820	0.9044	2.2813	1.6090	2.8010
(Intercept):2	.	.	.	.	-3.0790	0.0129
Sum.of.SuperBowlOdds	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sum.of.WLOU	0.0000	0.0006	0.0000	0.0006	0.0007	0.0008
Sum.of.SeasonEndWins	0.0026	0.0034	0.0026	0.0034	0.0040	0.0041
Sum.of.SeasonEndLosses	0.0016	0.0034	0.0016	0.0034	0.0027	0.0041
SeasonEndBookFinishpush	0.0041	0.0030	0.0041	0.0030	-0.0001	0.0038
SeasonEndBookFinishunder	0.0013	0.0024	0.0013	0.0024	0.0029	0.0029
HighestPlayoffLevelDR	-0.0023	0.0030	-0.0023	0.0030	0.0005	0.0038
HighestPlayoffLevelNP	0.0015	0.0031	0.0015	0.0031	-0.0007	0.0039
HighestPlayoffLevelSB	0.0047	0.0042	0.0047	0.0042	0.0087	0.0056
HighestPlayoffLevelSBW	0.0009	0.0043	0.0009	0.0043	0.0072	0.0054
HighestPlayoffLevelWC	0.0008	0.0030	0.0008	0.0030	-0.0003	0.0038
Sum.of.SBTPS	-0.0006	0.0004	-0.0006	0.0004	-0.0019	0.0005
Sum.of.HomeStadiumCapacity	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
StadiumTypeHybrid	0.0048	0.0026	0.0048	0.0026	0.0113	0.0032
StadiumTypeOpen Air	-0.0024	0.0018	-0.0024	0.0018	0.0029	0.0021
TeamClimateWarm	-0.0024	0.0013	-0.0024	0.0013	-0.0049	0.0016
Home	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Year	0.0000	0.0012	0.0000	0.0012	-0.0004	0.0014
SeasonWeek	-0.0007	0.0001	-0.0007	0.0001	-0.0007	0.0001
PriceDeflator	-0.0127	0.0370	-0.0127	0.0370	0.0103	0.0454
StadiumAge	0.0001	0.0000	0.0001	0.0000	-0.0001	0.0000

Overall, the final results of our adjusted Tobit model in Table 7 above are mixed with regard to our expectations. On one hand, we see that our model predicted significant factor levels within the Team Climate and Stadium Type variable. The fact that these variables are significant means that they are important predictors in estimating attendance demand and consumer preferences towards the in-game NFL experience. Interestingly however, the signs of those two coefficients are different, with the Tobit estimate of "Team Climate = Warm" being -0.0049 and the estimate for "Stadium Type = Hybrid" being 0.0113. This means that ceteris paribus, demand for NFL tickets will decrease by ~0.5% compared to cold climates. This could be due to many reasons, such as the higher number of entertainment options in the Southern US compared to the Northern US during the winter months. Alternatively, we see that the coefficient estimated for "Stadium Type = Hybrid" of 0.0113 means that ceteris paribus, demand for NFL tickets increases by 1.13% compared to open air or closed-dome stadiums. The variable for stadium age was not significant in our adjusted predictor set Tobit estimation, but with the very low absolute value of the estimated coefficient across model techniques, the fact that we lose significance in the updated predictor set is trivial. However, the fact that variables related to the physical comfort of the fan were significant further shows the importance of comfort factors within the live game experience.

On the other hand, we see that there are no significant predictors that deal with fan excitement or team winning percentage. This could be due to missing data and omitted variables, or it could mean that no matter how much winning a given team has done, fans will still attend live games. There were a few hints that there may some model misspecification or a data issue when we look at the "Highest Playoff Level achieved" variable. Recalling that this was a factor variable that increased in level as the team continued to win post-season games, we see that the 2 levels of this variable that are near marginal significance are "Highest Playoff Level" = 'Super Bowl' or "Highest Playoff Level" = 'Super Bowl Win'. Even if they are not statistically significant, we see that the signs of these factor level coefficients match our intuition. Each factor level of the "Highest Playoff Level achieved" variable is positive except "NP", and the two levels that refer to the Super Bowl have the highest coefficient values and are also the nearest to significance. The exception, "NP", means "No Playoffs", and the sign of this coefficient is also what we would intuitively expect. If a team did not make the playoffs, by definition they must not have a high win total for that season, and therefore would reduce demand for attendance because the team is bad. To summarize, we have not found statistically significant evidence that factors related to team quality or winning percentage have an effect on demand for NFL tickets, however there is evidence to suggest that factors around the comfort and experience for the fan are statistically significant.

Implications and Future Research

As we are attempting to gauge the effects of in-game experience factors on fan's demand for attendance, we have discovered that winningness of the home team and other 'excitement' factors in general are not significant predictors of demand for NFL tickets. Our model does have results that can provide suggestions to the NFL and potentially other sporting leagues on other comfort factors that surround the in-game experience. Overall, we have discovered that ceteris paribus, fans are less likely to attend games in warm-weather climates due to the sign and significance of our "TeamClimate" 2-level factor variable. We have also discovered that ceteris paribus, demand for live game tickets will increase if the stadium is a Hybrid type stadium. To provide some brief suggestions for the NFL, we would suggest the building of Hybrid stadiums for new teams or new markets. Additionally, we would suggest that the Southern or warm winter climate teams consider the other entertainment substitutes that are available during the NFL season.

We have fewer suggestions for the live game consumer NFL fan on ticket purchases. From our modeling approach and initial descriptive statistics to our final models we have established that NFL games are routinely sold out. If NFL games are not sold out, on average they have stadium occupancy percentages of 95% + throughout most of our sample. As long as there are fans who want to attend live NFL games and the NFL offers other less-expensive ways to view their product, we do not see any future reduction in live game ticket prices. The phenomenon of fan's inelastic demand for live game viewing, regardless of quality or winning percentage of the team, creates the perfect basis to set prices in the top inelastic portion of the demand curve.

For suggestions on future research, we put forward the idea of creating a better way to measure consumer loyalty to one NFL team. Loyal fans are more likely purchasers of team jerseys, clothes, concessions and in game tickets compared to fair-weather fans. If the NFL can gauge the brand loyalty of their consumers to each fan's favorite team, they can further increase price specificity directly up to the exact amount of each fan's willingness to pay. Also, it would be interesting to compare these results for in person demand to a similar study using Nielsen television ratings. One possible research topic given good data on television engagement would be the effects of new streaming services on the sports television marketplace. Usually when more firms enter the same market and become competitors, price elasticity increases, and prices go down. It would be interesting to determine if the sports television marketplace works in a similar way, with the advent of streaming services keeping prices down for consumers. In general, more data on this topic would be beneficial to understand the analysis options available to study.

Chapter VI: Conclusion

Throughout this analysis we have explored consumer demand for entertainment from live sporting events with a specific focus on NFL games. A widely established finding in studies of demand for live-game tickets for any major sporting event is that demand for tickets is inelastic, and ticket prices are set accordingly. We reviewed literature stating this conclusion and found areas that this research could improve on, which took the form of introducing our unique excitement and comfort related variables. After performing several estimations and viewing descriptive statistics, we have concluded that the factors that represent excitement are largely insignificant, and winning percentage or win total of a given NFL team is not a significant predictor of demand for live attendance. However, we have discovered several significant, comfort-based factors that previous research has left out, and proved that they are statistically significant areas of NFL demand prediction. These variables dealt with the Team's climate and the stadium type where the team played their home games. In general, our hypothesis returned mixed results with excitement factors not being a significant predictor of demand for NFL tickets, but showed that comfort-based factors were significant predictors of NFL attendance demand. We look forward to future research being built off of our analysis completed for this thesis project.

References

Biner, B. (2014). Parity in professional sports when revenues are maximized. *Economic Modelling*, 40, 12-20.

Borland, J., & MacDonald, R. (2003). Demand for sport. *Oxford Review of Economic Policy,* 19(4), 478-502.

Cairns, J., Jennett, N., Sloane, P. (1985). The economics of professional team sports: A survey of theory and evidence. *Journal of Economic Studies,* 13, 3-80.

Chang, Y., Potter, J. M., & Sanders, S. (2016). Inelastic sports ticket pricing, marginal win revenue, and firm pricing strategy: A behavioral pricing model. *Managerial Finance,* 42(9), 922-927.

Coates, D., & Humphreys, B. R. (2007). Ticket prices, concessions and attendance at professional sporting events. *International Journal of Sport Finance*, 2(3), 161-170.

Fort, R. (2004). Inelastic sports pricing. *Managerial and Decision Economics*, 25(2), 87-94.

Fort, R., & Quirk, J. (1995). Cross-subsidization, incentives, and outcomes in professional team sports leagues. *Journal of Economic Literature*, 33(3), 1265-1299.

Krautmann, A. C., & Berri, D. J. (2007). Can we find it at the concessions? Understanding price elasticity in professional sports. *Journal of Sports Economics,* 8(2), 183-191.

Peiss, M., & Kirstein, R. (2014). Optimal ticket pricing in professional sports: A social identity approach. *Economics Bulletin,* 34(4), 2151-2163.

Quirk, J. P., & Fort, R. D. (1992). *Pay dirt: The business of professional team sports*. Princeton, N.J.: Princeton University Press.

Rascher, D. A., & Solmes, J. P. G. (2011). Do fans want close contests? A test of the uncertainty of outcome hypothesis in the national basketball association. *Elgar Research Collection. The International Library of Critical Writings in Economics, vol. 254. Cheltenham, U.K. and Northampton, Mass.: Elgar*.

Rottenberg, S. (1956). The baseball players' labor market. *The Journal of Political Economy*, 64, 242–258.

Salaga, S., & Winfree, J. A. (2015). Determinants of secondary market sales prices for National Football League personal seat licenses and season ticket rights. *Journal of Sports Economics*, 16(3), 227-253.

Schreyer, D., Schmidt, S. L., & Torgler, B. (2017). Game outcome uncertainty and the demand for international football games: Evidence from the German TV market. *Journal of Media Economics*, 30(1), 31-45.

Tainsky, S., Xu, J., Mills, B. M., & Salaga, S. (2016). How success and uncertainty compel interest in related goods: Playoff probability and out-of-market television viewership in the National Football League. *Review of Industrial Organization*, 48(1), 29-43.

Welki, A. M., & Zlatoper, T. J. (1994). U.S. professional football: The demand for game-day attendance in 1991. *Managerial and Decision Economics*, 15(5), 489-495.

Appendix A: Tables

Table 1

Variable Descriptions

Variable	**Type**	**Description**
Year	Numeric	Year NFL season took place in.
X2019Franchise	Factor, 32 Levels	NFL Franchise as of 2019, base level is “Arizona Cardinals.”
Vkey	Text, Identifier	Identifier Key, not used in analysis. Represents every season across every year.
PercentageOccupancy	Numeric	The percentage occupancy of a game in a given season-week. Raw measurement of the total raw attendance number divided by the stadium capacity.
CensoredOccupancy	Numeric	The percentage occupancy of a game in a given season-week. Censored measurement of the total raw attendance number divided by the stadium capacity where values above 100% are replaced with 100%.
WeeklyAttendance	Numeric	The weekly attendance of a given NFL game in a given season-week.
WeeklyCapacity	Numeric	The weekly stadium capacity of a given NFL game in a given season-week.
Sellout?	Factor, 2 Levels	0-1 dummy variable where 1 represents a Percentage Occupancy value greater than or equal to 100%.
Home?	Factor, 2 Levels	0-1 dummy variable where 1 represents a "Home" game. All observations have a value of 1 after data cleaning.
SeasonWeek	Numeric	Integer value representing the Week number of the game.

Table 1 (continued)

Variable	**Type**	**Description**
Sum of HomeStadiumCapacity	Numeric	Home Stadium capacity of a given team. Observations with different values for "Sum of HomeStadiumCapacity" and "WeeklyCapacity" mean that the game took place outside of the team's traditional home venue.
StadiumName	Text, Identifier	The name of the stadium where the game was played.
StadiumLocation	Text, Identifier	The location of the stadium where the game was played.
StadiumType	Factor, 3 Levels	The type of the stadium where the game was played.
TeamClimate	Factor, 2 Levels	The climate of the location where the game was played.
Home	Numeric	Season total of attendance for all games played at home by a given team in a given NFL season.
Sum of SeasonEndWins	Numeric	The Season-Ending Win total for a given team for a given season.
StadiumAge	Numeric	The age of a given stadium in a given year, at the start of the NFL season.

Table 2

Season Average Percentage of Total NFL Stadium Seats Occupied

Summary Statistics: League Wide Annual Percentage Occupancy:	
Year	**Average Percentage Occupancy Across all 256 Games**
2002	94.8581%
2003	95.4314%
2004	96.5633%
2005	96.6366%
2006	98.8565%
2007	98.6818%
2008	97.6678%
2009	95.1293%
2010	94.1276%
2011	94.9668%
2012	95.0559%
2013	96.0779%
2014	96.9212%
2015	96.3409%
2016	96.0263%
2017	94.8234%
2018	94.1997%
2019	93.5159%
Total Average	**95.8822%**

Table 3

Dependent "Experiential Factors" variables unique to our model:

Unique Dependent Variables in Modeling	
Excitement Variables	**Comfort Variables**
Sum of SeasonEndWins	StadiumType
Sum of SeasonEndLosses	TeamClimate
Sum of SBTPS	StadiumAge
Sum of SuperBowlOdds	
Sum of WLOU	

Table 3 (continued)

Excitement Variables	**Comfort Variables**
SeasonEndBookFinish	
HighestPlayoffLevel	

Table 4

Base Ordinary Least Squares Estimation Results

Coefficients	**Estimate**	**Std. Error**	**t values**	**Pr(>|t|)**	**Significant notation**
(Intercept)	0.715	0.432	1.655	0.098	*
Year	0.000	0.000	0.385	0.700	
X2019FranchiseAtlanta Falcons	-0.008	0.001	-6.269	0.000	***
X2019FranchiseBaltimore Ravens	0.006	0.003	1.917	0.055	*
X2019FranchiseBuffalo Bills	0.002	0.003	0.720	0.472	**
X2019FranchiseCarolina Panthers	-0.019	0.002	-11.450	0.000	***
X2019FranchiseChicago Bears	0.011	0.004	2.937	0.003	***
X2019FranchiseCincinnati Bengals	0.006	0.003	1.936	0.053	*
X2019FranchiseCleveland Browns	0.012	0.003	3.587	0.000	***
X2019FranchiseDallas Cowboys	-0.006	0.002	-2.848	0.004	***
X2019FranchiseDenver Broncos	-0.001	0.003	-0.155	0.877	
X2019FranchiseDetroit Lions	0.018	0.002	7.288	0.000	***
X2019FranchiseGreen Bay Packers	0.004	0.004	0.957	0.339	***
X2019FranchiseHouston Texans	-0.011	0.001	-10.418	0.000	***
X2019FranchiseIndianapolis Colts	0.021	0.003	8.350	0.000	***
X2019FranchiseJacksonville Jaguars	-0.015	0.002	-9.271	0.000	***
X2019FranchiseKansas City Chiefs	-0.001	0.004	-0.377	0.706	
X2019FranchiseLos Angeles Chargers	-0.022	0.002	-14.350	0.000	***
X2019FranchiseLos Angeles Rams	0.015	0.002	6.758	0.000	***
X2019FranchiseMiami Dolphins	-0.003	0.002	-1.458	0.145	**

Table 4 (continued)

Coefficients	**Estimate**	**Std. Error**	**t values**	**Pr(>|t|)**	**Significant notation**
X2019FranchiseMinnesota Vikings	0.025	0.003	9.719	0.000	***
X2019FranchiseNew England Patriots	0.015	0.003	4.256	0.000	***
X2019FranchiseNew Orleans Saints	-0.009	0.001	-6.293	0.000	***
X2019FranchiseNew York Giants	-0.003	0.004	-0.837	0.403	
X2019FranchiseNew York Jets	-0.004	0.003	-1.082	0.279	
X2019FranchiseOakland Raiders	-0.019	0.002	-10.427	0.000	***
X2019FranchisePhiladelphia Eagles	0.010	0.003	2.863	0.004	***
X2019FranchisePittsburgh Steelers	0.010	0.004	2.786	0.005	***
X2019FranchiseSan Francisco 49ers	-0.001	0.002	-0.293	0.770	
X2019FranchiseSeattle Seahawks	0.010	0.003	3.026	0.002	***
X2019FranchiseTampa Bay Buccaneers	-0.012	0.002	-7.797	0.000	***
X2019FranchiseTennessee Titans	-0.008	0.002	-4.831	0.000	***
X2019FranchiseWashington Redskins	-0.007	0.004	-1.967	0.049	**
WeeklyAttendance	0.000	0.000	345.269	0.000	***
WeeklyCapacity	0.000	0.000	-182.940	0.000	***
SeasonWeek	0.000	0.000	0.273	0.785	
Sum.of.HomeStadiumCapacity	0.000	0.000	-0.383	0.702	
StadiumTypeHybrid	0.003	0.001	2.529	0.011	**
StadiumTypeOpen Air	0.008	0.002	5.409	0.000	***
TeamClimateWarm	0.021	0.003	8.075	0.000	***
Home	0.000	0.000	2.676	0.007	***
Sum.of.SeasonEndWins	0.000	0.001	-0.307	0.759	
Sum.of.SeasonEndLosses	0.000	0.001	-0.366	0.714	
Sum.of.SBTPS	-0.001	0.000	-2.387	0.017	**
Sum.of.SuperBowlOdds	0.000	0.000	2.611	0.009	***
Sum.of.WLOU	0.001	0.000	4.223	0.000	***

Table 4 (continued)

Coefficients	**Estimate**	**Std. Error**	**t values**	**Pr(>|t|)**	**Significant notation**
SeasonEndBookFinishpush	0.000	0.001	0.814	0.416	
SeasonEndBookFinishunder	-0.001	0.000	-1.749	0.080	*
HighestPlayoffLevelDR	0.000	0.001	-0.242	0.808	
HighestPlayoffLevelNP	0.001	0.001	1.373	0.170	
HighestPlayoffLevelSB	0.001	0.001	0.937	0.349	
HighestPlayoffLevelSBW	0.001	0.001	0.921	0.357	
HighestPlayoffLevelWC	0.002	0.001	3.008	0.003	***
PriceDeflator	-0.009	0.007	-1.356	0.175	
StadiumAge	0.000	0.000	4.426	0.000	***
Residual standard error: 0.007788 on 4553 degrees of freedom					
Multiple R-squared: 0.9917, Adjusted R-squared: 0.9916					
F-statistic: 1.01e+04 on 54 and 4553 DF, p-value: < 2.2e-16					

Table 5

Base Random Effects Estimation Summary

Coefficients	**Estimate**	**Std. Error**	**t values**	**Pr(>|t|)**	**Significant notation**
(Intercept)	0.719	0.955	0.753	0.451	
Year	0.000	0.000	0.170	0.865	
X2019FranchiseAtlanta Falcons	-0.008	0.003	-2.833	0.005	***
X2019FranchiseBaltimore Ravens	0.006	0.007	0.860	0.390	
X2019FranchiseBuffalo Bills	0.002	0.008	0.327	0.744	
X2019FranchiseCarolina Panthers	-0.019	0.004	-5.167	0.000	***
X2019FranchiseChicago Bears	0.011	0.008	1.320	0.187	***
X2019FranchiseCincinnati Bengals	0.006	0.007	0.872	0.383	
X2019FranchiseCleveland Browns	0.012	0.007	1.622	0.105	**
X2019FranchiseDallas Cowboys	-0.006	0.005	-1.334	0.182	
X2019FranchiseDenver Broncos	-0.001	0.008	-0.083	0.934	
X2019FranchiseDetroit Lions	0.018	0.005	3.288	0.001	***

Table 5 (continued)

Coefficients	**Estimate**	**Std. Error**	**t values**	**Pr(>|t|)**	**Significant notation**
X2019FranchiseGreen Bay Packers	0.003	0.008	0.411	0.681	
X2019FranchiseHouston Texans	-0.011	0.002	-4.729	0.000	***
X2019FranchiseIndianapolis Colts	0.021	0.006	3.759	0.000	***
X2019FranchiseJacksonville Jaguars	-0.015	0.003	-4.197	0.000	***
X2019FranchiseKansas City Chiefs	-0.001	0.008	-0.186	0.852	
X2019FranchiseLos Angeles Chargers	-0.022	0.003	-6.495	0.000	***
X2019FranchiseLos Angeles Rams	0.015	0.005	3.038	0.002	***
X2019FranchiseMiami Dolphins	-0.003	0.004	-0.655	0.513	
X2019FranchiseMinnesota Vikings	0.025	0.006	4.389	0.000	***
X2019FranchiseNew England Patriots	0.015	0.008	1.917	0.055	*
X2019FranchiseNew Orleans Saints	-0.009	0.003	-2.868	0.004	***
X2019FranchiseNew York Giants	-0.003	0.008	-0.397	0.691	
X2019FranchiseNew York Jets	-0.004	0.008	-0.503	0.615	
X2019FranchiseOakland Raiders	-0.019	0.004	-4.729	0.000	***
X2019FranchisePhiladelphia Eagles	0.010	0.007	1.293	0.196	*
X2019FranchisePittsburgh Steelers	0.010	0.008	1.246	0.213	*
X2019FranchiseSan Francisco 49ers	-0.001	0.005	-0.134	0.893	
X2019FranchiseSeattle Seahawks	0.010	0.007	1.366	0.172	*
X2019FranchiseTampa Bay Buccaneers	-0.012	0.004	-3.512	0.000	***
X2019FranchiseTennessee Titans	-0.008	0.004	-2.162	0.031	**
X2019FranchiseWashington Redskins	-0.007	0.008	-0.911	0.363	

Table 5 (continued)

Coefficients	Estimate	Std. Error	t values	Pr(>|t|)	Significant notation
WeeklyAttendance	0.000	0.000	483.553	0.000	***
WeeklyCapacity	0.000	0.000	-244.869	0.000	***
SeasonWeek	0.000	0.000	0.306	0.760	
Sum.of.HomeStadiumCapacity	0.000	0.000	-1.265	0.206	
StadiumTypeHybrid	0.003	0.002	1.167	0.243	
StadiumTypeOpen Air	0.008	0.003	2.433	0.015	**
TeamClimateWarm	0.021	0.006	3.637	0.000	***
Home	0.000	0.000	2.042	0.041	**
Sum.of.SeasonEndWins	0.000	0.001	-0.120	0.904	
Sum.of.SeasonEndLosses	0.000	0.001	-0.150	0.881	
Sum.of.SBTPS	-0.001	0.001	-1.059	0.289	
Sum.of.SuperBowlOdds	0.000	0.000	1.187	0.235	
Sum.of.WLOU	0.001	0.000	1.908	0.056	*
SeasonEndBookFinishpush	0.000	0.001	0.371	0.711	
SeasonEndBookFinishunder	-0.001	0.001	-0.791	0.429	
HighestPlayoffLevelDR	0.000	0.001	-0.110	0.913	
HighestPlayoffLevelNP	0.001	0.001	0.632	0.528	
HighestPlayoffLevelSB	0.001	0.002	0.424	0.672	
HighestPlayoffLevelSBW	0.001	0.002	0.423	0.672	
HighestPlayoffLevelWC	0.002	0.001	1.364	0.173	
PriceDeflator	-0.010	0.015	-0.616	0.538	
StadiumAge	0.000	0.000	2.028	0.043	**
## Total Sum of Squares: 9.7858					
## Residual Sum of Squares: 0.13608					
## R-Squared: 0.98609					
## Adj. R-Squared: 0.98593					
## Chisq: 322857 on 54 DF, p-value: < 2.22e-16					

Table 6

Base Tobit Estimation Summary

Coefficients	**Estimate**	**Std. Error**	**t values**	**Pr(>|t|)**	**Significant notation**
(Intercept):1	1.3010	0.5038	NA	NA	
(Intercept):2	-4.8400	0.0122	-397.548	0.0000	***
Year	-0.0002	0.0003	-0.8940	0.3712	
X2019FranchiseAtlanta Falcons	-0.0134	0.0016	-8.5530	0.0000	***
X2019FranchiseBaltimore Ravens	0.0185	0.0036	5.0960	0.0000	***
X2019FranchiseBuffalo Bills	0.0120	0.0038	3.1850	0.0014	***
X2019FranchiseCarolina Panthers	-0.0154	0.0019	-8.2990	0.0000	***
X2019FranchiseChicago Bears	0.0216	0.0042	5.1570	0.0000	***
X2019FranchiseCincinnati Bengals	0.0164	0.0035	4.7080	0.0000	***
X2019FranchiseCleveland Browns	0.0203	0.0036	5.5550	0.0000	***
X2019FranchiseDallas Cowboys	-0.0061	0.0027	-2.2450	0.0248	**
X2019FranchiseDenver Broncos	0.0114	0.0037	3.0570	0.0022	***
X2019FranchiseDetroit Lions	0.0184	0.0026	7.0830	0.0000	***
X2019FranchiseGreen Bay Packers	0.0130	0.0042	3.1160	0.0018	***
X2019FranchiseHouston Texans	-0.0090	0.0012	-7.3810	0.0000	***
X2019FranchiseIndianapolis Colts	0.0261	0.0028	9.2650	0.0000	***
X2019FranchiseJacksonville Jaguars	-0.0114	0.0017	-6.5840	0.0000	***
X2019FranchiseKansas City Chiefs	0.0072	0.0038	1.8900	0.0587	*
X2019FranchiseLos Angeles Chargers	-0.0176	0.0016	-10.9180	0.0000	***
X2019FranchiseLos Angeles Rams	0.0169	0.0024	6.9330	0.0000	***
X2019FranchiseMiami Dolphins	-0.0029	0.0021	-1.3560	0.1751	
X2019FranchiseMinnesota Vikings	0.0251	0.0028	8.8350	0.0000	***
X2019FranchiseNew England Patriots	0.0291	0.0043	6.7390	0.0000	***

Table 6 (continued)

Coefficients	Estimate	Std. Error	t values	Pr(>|t|)	Significant notation
X2019FranchiseNew Orleans Saints	-0.0150	0.0016	-9.2060	0.0000	***
X2019FranchiseNew York Giants	0.0062	0.0039	1.5890	0.1120	**
X2019FranchiseNew York Jets	0.0044	0.0037	1.1900	0.2340	*
X2019FranchiseOakland Raiders	-0.0130	0.0021	-6.1560	0.0000	***
X2019FranchisePhiladelphia Eagles	0.0207	0.0037	5.6360	0.0000	***
X2019FranchisePittsburgh Steelers	0.0217	0.0042	5.2200	0.0000	***
X2019FranchiseSan Francisco 49ers	0.0007	0.0035	0.2140	0.8306	
X2019FranchiseSeattle Seahawks	0.0203	0.0036	5.6810	0.00000	***
X2019FranchiseTampa Bay Buccaneers	-0.0079	0.0018	-4.5240	0.0000	***
X2019FranchiseTennessee Titans	-0.0073	0.0022	-3.3420	0.0008	***
X2019FranchiseWashington Redskins	0.0129	0.0041	3.1180	0.0018	***
WeeklyAttendance	0.0000	0.0000	315.7810	0.0000	***
WeeklyCapacity	0.0000	0.0000	-170.427	0.0000	***
SeasonWeek	0.0000	0.0000	-0.3910	0.6960	
Sum.of.HomeStadiumCapacity	0.0000	0.0000	0.6390	0.5227	
StadiumTypeHybrid	-0.0020	0.0013	-1.4980	0.1340	**
StadiumTypeOpen Air	-0.0009	0.0017	-0.5340	0.5935	
TeamClimateWarm	0.0266	0.0029	9.2800	0.0000	***
Home	0.0000	0.0000	0.6980	0.4853	
Sum.of.SeasonEndWins	0.0011	0.0008	1.4670	0.1424	
Sum.of.SeasonEndLosses	0.0012	0.0008	1.5310	0.1257	
Sum.of.SBTPS	-0.0013	0.0005	-2.6370	0.0084	***
Sum.of.SuperBowlOdds	0.0000	0.0000	2.6090	0.0091	***
Sum.of.WLOU	0.0006	0.0001	3.9430	0.0001	***
SeasonEndBookFinishpush	-0.0006	0.0007	-0.9290	0.3527	
SeasonEndBookFinishunder	-0.0006	0.0005	-1.0900	0.2755	
HighestPlayoffLevelDR	0.0010	0.0007	1.4570	0.1451	
HighestPlayoffLevelNP	0.0004	0.0007	0.5130	0.6078	

Table 6 (continued)

Coefficients	Estimate	Std. Error	t values	Pr(>|t|)	Significant notation
HighestPlayoffLevelSB	0.0014	0.0010	1.3000	0.1936	
HighestPlayoffLevelSBW	0.0017	0.0011	1.6530	0.0982	*
HighestPlayoffLevelWC	0.0020	0.0007	2.9290	0.0034	***
PriceDeflator	0.0003	0.0082	0.0320	0.9748	
StadiumAge	0.0001	0.0000	3.2510	0.0012	***
Names of linear predictors: mu, loglink(sd)					
Log-likelihood: 10984.52 on 9160 degrees of freedom					
R-squared = 0.9905546					

Table 7

Adjusted Tobit Estimation Results

Coefficients	Estimate	Std. Error	t values	Pr(>|t|)	Significant notation
(Intercept):1	1.6090	2.801	NA	NA	
(Intercept):2	-3.0790	0.013	-238.366	0.000	***
Sum.of.SuperBowlOdds	0.0000	0.000	0.714	0.475	
Sum.of.WLOU	0.0007	0.001	0.961	0.337	
Sum.of.SeasonEndWins	0.0040	0.004	0.956	0.339	
Sum.of.SeasonEndLosses	0.0027	0.004	0.661	0.509	
SeasonEndBookFinishpush	-0.0001	0.004	-0.016	0.987	
SeasonEndBookFinishunder	0.0029	0.003	0.981	0.326	
HighestPlayoffLevelDR	0.0005	0.004	0.123	0.902	
HighestPlayoffLevelNP	-0.0007	0.004	-0.173	0.862	
HighestPlayoffLevelSB	0.0087	0.006	1.556	0.120	
HighestPlayoffLevelSBW	0.0072	0.005	1.321	0.187	
HighestPlayoffLevelWC	-0.0003	0.004	-0.067	0.946	
Sum.of.SBTPS	-0.0019	0.000	-3.940	0.000	***
Sum.of.HomeStadiumCapacity	0.0000	0.000	-83.707	0.000	***
StadiumTypeHybrid	0.0113	0.003	3.526	0.000	***
StadiumTypeOpen Air	0.0029	0.002	1.339	0.181	
TeamClimateWarm	-0.0049	0.002	-2.945	0.003	***
Home	0.0000	0.000	92.834	0.000	***
Year	-0.0004	0.001	-0.269	0.788	

Table 7 (continued)

Coefficients	Estimate	Std. Error	t values	Pr(>|t|)	Significant notation
SeasonWeek	-0.0007	0.000	-4.767	0.000	***
PriceDeflator	0.0103	0.045	0.227	0.820	
StadiumAge	-0.0001	0.000	-1.505	0.132	
Names of linear predictors: mu, loglink(sd)					
Log-likelihood: 4845.331 on 9193 degrees of freedom					
R-squared = 0.7625348					

Table 8

Adjusted Predictor Set Model Coefficients and Standard Errors

	OLS		Random Effects		Upper Bound Censored Tobit	
Coefficients	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**	**Estimate**	**Std. Error**
(Intercept):1	0.9045	2.2820	0.9044	2.2813	1.6090	2.8010
(Intercept):2	.	.	.	.	-3.0790	0.0129
Sum.of.SuperBowlOdds	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Sum.of.WLOU	0.0000	0.0006	0.0000	0.0006	0.0007	0.0008
Sum.of.SeasonEndWins	0.0026	0.0034	0.0026	0.0034	0.0040	0.0041
Sum.of.SeasonEndLosses	0.0016	0.0034	0.0016	0.0034	0.0027	0.0041
SeasonEndBookFinishpush	0.0041	0.0030	0.0041	0.0030	-0.0001	0.0038
SeasonEndBookFinishunder	0.0013	0.0024	0.0013	0.0024	0.0029	0.0029
HighestPlayoffLevelDR	-0.0023	0.0030	-0.0023	0.0030	0.0005	0.0038
HighestPlayoffLevelNP	0.0015	0.0031	0.0015	0.0031	-0.0007	0.0039
HighestPlayoffLevelSB	0.0047	0.0042	0.0047	0.0042	0.0087	0.0056
HighestPlayoffLevelSBW	0.0009	0.0043	0.0009	0.0043	0.0072	0.0054
HighestPlayoffLevelWC	0.0008	0.0030	0.0008	0.0030	-0.0003	0.0038
Sum.of.SBTPS	-0.0006	0.0004	-0.0006	0.0004	-0.0019	0.0005
Sum.of.HomeStadiumCapacity	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
StadiumTypeHybrid	0.0048	0.0026	0.0048	0.0026	0.0113	0.0032
StadiumTypeOpen Air	-0.0024	0.0018	-0.0024	0.0018	0.0029	0.0021
TeamClimateWarm	-0.0024	0.0013	-0.0024	0.0013	-0.0049	0.0016
Home	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Year	0.0000	0.0012	0.0000	0.0012	-0.0004	0.0014
SeasonWeek	-0.0007	0.0001	-0.0007	0.0001	-0.0007	0.0001
PriceDeflator	-0.0127	0.0370	-0.0127	0.0370	0.0103	0.0454
StadiumAge	0.0001	0.0000	0.0001	0.0000	-0.0001	0.0000

Appendix B: Figures

Figure 1

Annual Season End Win Totals Per Franchise

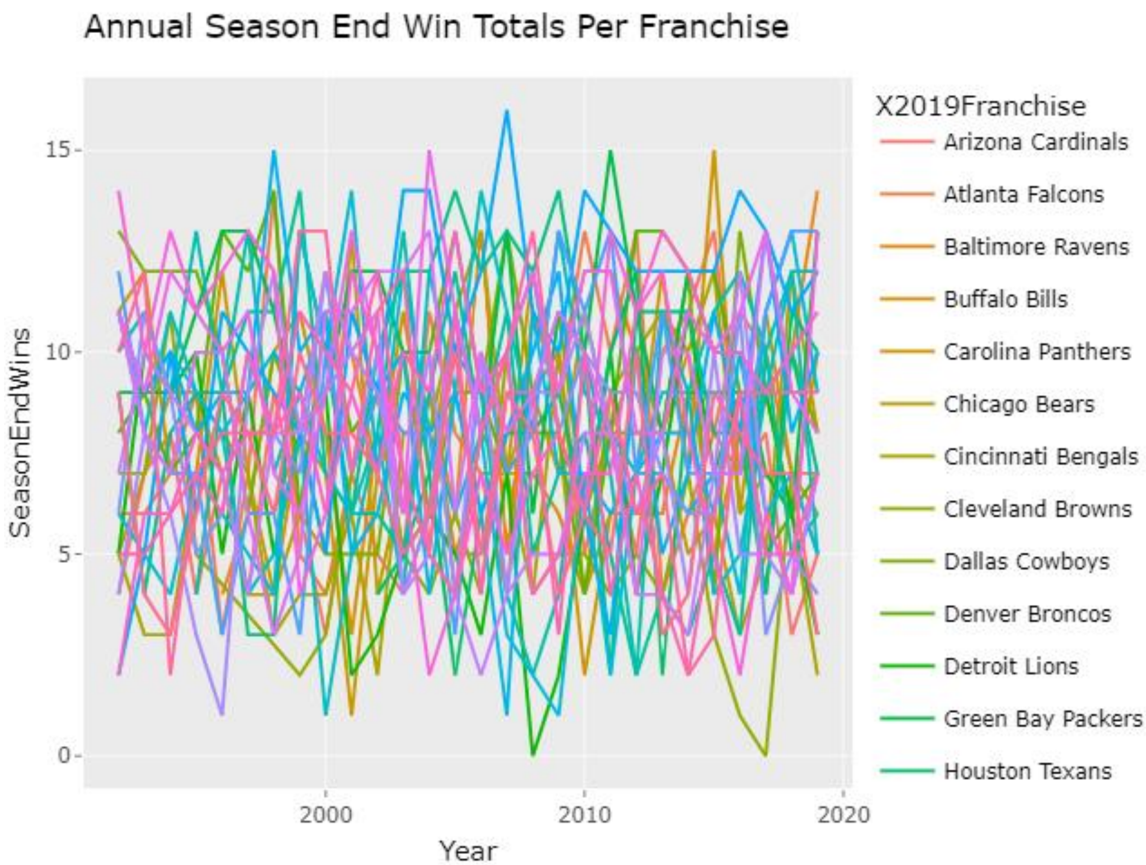


Figure 2

Annual Super Bowl Odds Per Franchise

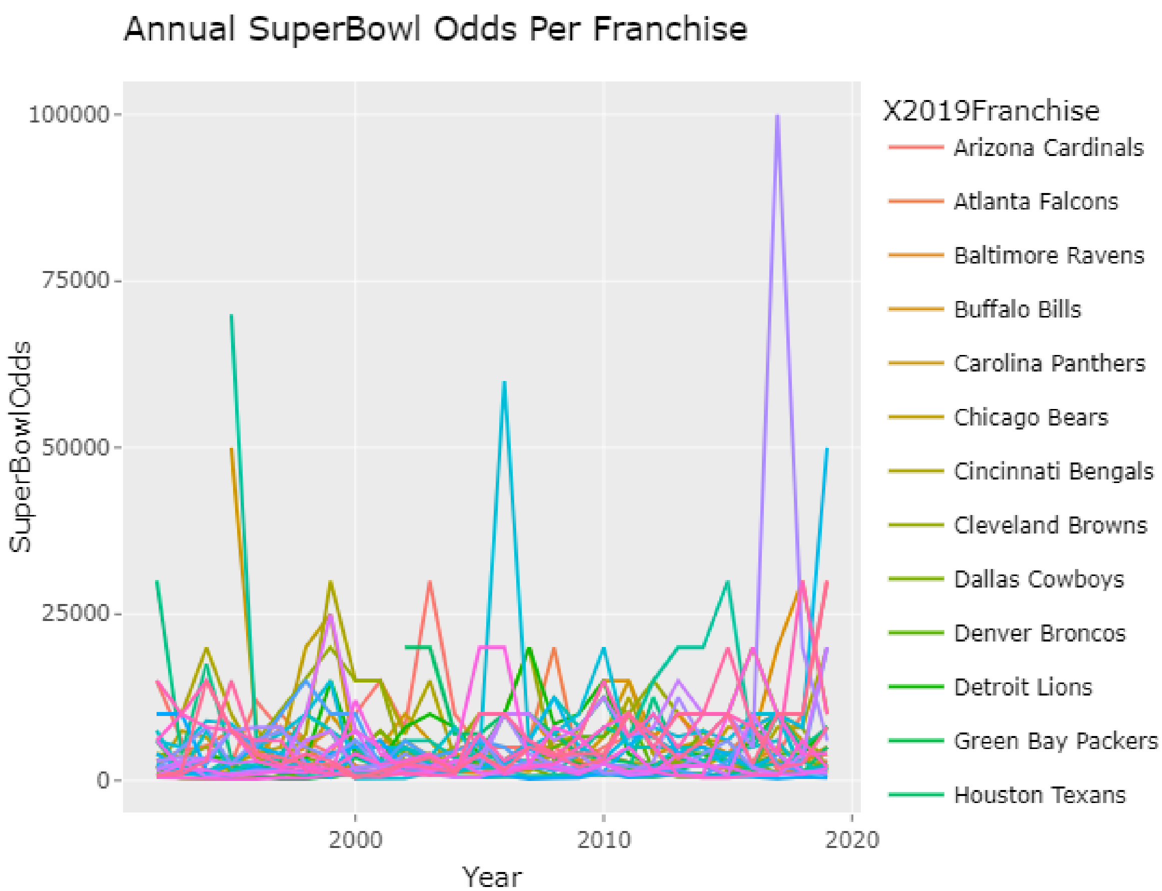


Figure 3

Residual Plot Matrix from Adjusted Tobit Estimation