Spring 5-5-2024

An Analysis of NFL Betting Strategies: A Look Into High Value Bets

Brady Trenchard
btrencha@skidmore.edu

An Analysis of NFL Betting Strategies: A Look Into High Value Bets

Brady Trenchard

May 5, 2024

This thesis is submitted in partial fulfillment of the requirements for the course Senior Seminar (EC 375), during the Spring Semester of 2024

While writing this thesis, I have not witnessed any wrongdoing, nor have I personally violated any conditions of the Skidmore College Honor Code.

Name: Brady Trenchard

Signature: ____________________

Abstract

This paper uses data from NFL games throughout the 2013-2019 seasons and analyzes the profitability of several correlated parlays and straight bets. Independent variables such as temperature, wind, prime time, and divisional game status were incorporated. Using OLS regression analysis, several strategies are presented that would have been profitable throughout the sample period. Results suggested that wagers on the under in divisional games with high wind were profitable. Furthermore, it was profitable to wager on a home favorite and over parlay in prime time and non-divisional games. Results also indicated it was profitable to wager on a home underdog and under parlay in divisional games when the spread line was between 10 and 13.5 points. However, these results were only profitable during the sample time period and are unlikely to persist due to the extremely efficient nature of bookmakers.

I. Introduction

Sports betting became legalized in 2018 after the Supreme Court struck down the Professional and Amateur Sports Protection Act (PASPA), which allowed all states to determine their own sports betting laws (CBS Sports, 2023). Prior to 2018, sports betting was illegal everywhere in the United States except Nevada, where it has been legal since 1951. However, bettors in other states could wager illegally on off-shore sites and with bookmakers (Rolfe and Robinson, 2023). The recent legalization of sports betting in the United States has dramatically increased the popularity of sports betting apps and websites such as FanDuel, DraftKings, and BetMGM. Sports betting revenues have immensely increased since legalization in 2018, and revenue for the sports betting industry reached nearly $11 billion in 2023 (Sports Business Journal, 2023). Legalization has led to much more data being available within the industry which has opened opportunities for research into new avenues such as parlay wagers and high-value bets. This paper will analyze the profitability of a variation of correlated parlays and straight bets in the NFL. The main question being is a parlay between the home underdog and the under a profitable strategy in the NFL? Furthermore, is it profitable to wager on other correlated parlays such as a favorite and over parlay? An analysis will also be completed for independent variables to see if they have an impact on the profitability rate for the wagers.

This research presents an opportunity for insight into how regressions and analytical work can fit into the picture within the sports betting industry. Sportsbooks and bookmakers could utilize this information to adjust lines to limit any potential exploitations or inefficiencies within the market. Bookmaking is a finely tuned science requiring many analytics and copious amounts of data. Any edge that sports bettors or bookmakers can acquire on the competition can be tremendously beneficial for either side. This research also applies to the NFL. In terms of broadcasting and scheduling, the NFL can utilize this information to determine a more efficient schedule to maximize television audience and fan attendance.

Davis et al. (2018) introduced the concept of correlated parlays with their research in college football. They analyzed the profitability of specific parlays that had correlative relationships such as a parlay between the favorite and the over as well as an underdog and under parlay. Their results identified positive returns for both types of correlated parlays in college football. Shank (2018, 2019) analyzed the specific effects of prime time and divisional games on the outcome of betting wagers in the NFL. He identified inefficiencies within the point spread and totals markets and determined various profitable betting strategies. Anderson (2019) also examined the effect of prime time games on scoring in the NFL and discovered similar results that games played in prime time result in closer final scores. Borghesi (2007, 2008) examined the impact of weather effects on NFL games and patterns that are present when an underdog plays at home. These scholars' work contributed significantly to this paper and provided an exceptional foundation for future research to be done into the efficiency of the NFL betting market.

The purpose of this paper is to analyze the profitability of high-value wagers in the form of correlated parlays in the NFL. A combination of independent variables was used to determine relationships that could significantly impact the profitability of parlays and how often they hit. A few straight bets were also analyzed for regression and profitability purposes. A regression was run with eight independent variables for all the parlays and straight bets that were studied. Regression results determined characteristics of specific games that indicated a higher probability for each wager to hit. First, an analysis was completed to determine the exact profitability of each parlay or straight bet throughout the 2013-2019 seasons. Then, the regression results were incorporated into the profitability analysis to uncover strategies that would have been profitable during the sample time period.

Overview

In the NFL sports betting arena, bettors have the opportunity to engage in wagering on point spreads and over/unders for every game throughout the season. Point spread bets rank among the most popular wagers. A point spread bet is a prediction that a team will either "cover the spread" by winning by a larger margin or losing by a narrower margin than the set point spread established by bookmakers. For instance, a common point spread might be 6.5 points, implying that the favored team must secure victory by 7 points or more for the bet to be successful. Conversely, if the favored team falls short of this margin, the underdog is said to have "covered the spread." Over/under bets center on the total score of the game. Should the over/under line be set at 50.5, the combined score of both teams must exceed 51 for the over bet to prevail; otherwise, wagers placed on the under are victorious.

Traditionally, sportsbooks strive to attain equal amounts of money wagered on each side of a designated line. For instance, if the over/under line is set at 44, sportsbooks aim for a 50-50 split in the betting funds between over 44 and under 44. This equilibrium across both sides of the line serves to mitigate the sportsbooks' risk and generates profit through the vigorish (commission) on losing bets (Humphreys et al., 2013). Sportsbooks earn commission because the odds are typically -110 for both sides of an even bet, meaning bettors stake $11 to win $10. Thus, with an evenly balanced book (50% of bettors on each side), the sportsbook pays out $10 to the winning bettor from the $11 wagered by the losing bettor, retaining the remaining $1 as commission, called the vigorish. In such a scenario, if a balanced sportsbook were to have only two bettors (one on each side) each betting $11 to win $10, it would generate a mere $1 profit. However, it is evident how the commission accumulates as the number of bettors and the money wagered increases (Pfitzner et al, 2009). Due to the fact that spread and over/under bets give -110 odds, a bettor must wager $11 to win $10. Therefore, in order for a bettor to break even and overcome the vigorish, he/she must win 52.4% of their bets.

II. Literature Review

Home Underdog Strategy

A notable and historically profitable strategy in the NFL betting realm involves wagering on the home underdog to cover the spread. Extensive literature delves into the evidence supporting this strategy. Borghesi (2007) uncovered its effectiveness in late-season scenarios, revealing that home underdogs cover the spread by an average of 3.13 points during weeks 15-18 and average an outright victory margin of about 9 points in the playoffs. The bias toward home underdogs persisted throughout the sample period from the 1981-2000 seasons, challenging the notion that sportsbooks adjust over time.

Similarly, Dare and Holland (2004) and Golec and Tamarkin (1991) uncovered biases favoring home underdogs as well. Golec and Tamarkin (1991) reported a winning percentage of 53.7% from 1973-1987 and Gray and Gray (1997) identified statistically significant returns exceeding 4% for a bettor wagering on the home underdog strategy. Shank (2018) further supported the profitability of betting on the home team when they were considerable underdogs, with a success rate of approximately 63% when the home team was projected to lose by 10 or more points. Overall, there is ample evidence supporting the home underdog strategy. However, with the exception of Shank's (2018) research, all other research supporting the home underdog strategy is quite dated, so it is possible that sportsbooks have adjusted over time to mitigate this inefficiency.

Point Spread Market Efficiency

Extensive research has also focused on the efficiency of the NFL point spread market. Gray and Gray (1997) examined home teams, employing a probit-based regression analysis, which suggested that underdogs and home teams are more likely to cover the spread than favorites and away teams. Their results also indicated that betting markets tend to overreact to a team's recent performances while discounting the team's overall season performance. However, this research is also quite dated and likely lacks relevance in the current NFL point spread market. Shank (2018) identified a slight inefficiency where the home team covered the spread approximately 49% of the time. He also further delved into inefficiencies associated with prime time games, revealing that games played during prime time were more likely to result in the home team covering the spread. Shank (2019) also unveiled that games played against divisional opponents resulted in a lower probability of the home team covering the spread and the game going over.

Anderson (2019) found that teams whose home field is grass tend to have a better performance against the spread. He found the grass coefficient (2.52) to be statistically significant. This indicated that away teams tend to perform better when the game is played on grass as opposed to turf. Marino (2017) also found that the field variable held statistical significance at the 10% level. This also supported that games on turf decreased the likelihood of the away team covering the spread by 4.14%. Interestingly, Anderson (2019) found dome and temperature variables to be insignificant when regressing for performance against the spread. Anderson (2019) also analyzed prime time games and found that scores in prime time games were closer than in other games. Furthermore, he found that home teams tend to perform better in prime time regardless of favorite or underdog status. Overall, previous research determined various inefficiencies in the NFL point spread betting market, presenting evidence for profitable strategies.

Over/Under Market Efficiency

A considerable body of literature also exists in examining efficiency within the totals market. Paul and Weinbach (2002) identified market inefficiencies, noting that unders prevailed over 50% of the time. Their analysis, spanning from 1979 to 2000, indicated that when the total was set at 47.5 points or higher, the under hit 58.74% of the time. Borghesi (2008) also demonstrated that the NFL totals betting market is not efficient as he found the percentage of over bets that win was 48.92% with a p-value of 0.0635, demonstrating a statistical inefficiency. Therefore, simply betting the under could potentially win at a rate of over 50%. Pfitzner et al. (2009) adopted a systematic approach in predicting team scores and comparing them against actual contest outcomes during the 2005-2006 NFL season, also revealing inefficiencies in the over/under market during that period of time. Shank's (2018) research revealed that games played against divisional opponents were more likely to hit the under. He also unearthed inefficiencies which indicated a higher likelihood of hitting the over with very high or low lines, no precipitation, and when 60% of bettors favored the over. Furthermore, analysis by Marino (2017) revealed that games played under a dome raised the likelihood of the over by 5.5%. There are clear inefficiencies within the NFL totals and point spread betting markets, which provides the foundation of this paper. In examining parlays with various point spread and total wagers, this paper synthesizes the inefficiencies in both markets to analyze the potential profitability of high-value wagers.

Correlated Parlays

Davis et al. (2018) looked into correlated parlays in college football and analyzed the linkage between underdog/under and favorite/over parlays. Their results showed that games, where underdogs covered the spread, were also more likely to go under the set total. Comparatively, contests, where the favorite covered the spread, were more likely to go over the set total. Davis et al. (2018) analyzed the point spread and found that in games where the point spread was 21.5+ points, the favorite/over parlay produced returns of a surprising 19.23%. Furthermore, Davis et. al (2018) analyzed “multiples” which refer to a value equal to the total divided by the spread. For example, in a game where Team X is favored by 17 points over Team Y and the total is set at 42, the multiple would be 2.47. He analyzed multiples of less than 1.5, (1.5 - 2), (2 - 2.5), (2.5 - 3), and 3 or higher. For games where the multiple was 2 - 2.5, the favorite/over parlay generated returns of a remarkable 50.52%. Correlated parlays generated significant returns in the college football betting market. With hundreds of games each week it is easy to see how inefficiencies could become present in the college football scene. However, with approximately only 16 games played each week in the NFL, bookmakers can likely develop a more systematic approach to setting lines, which could reduce the success and profitability of correlated parlays in the NFL betting market.

Previously Profitable Strategies

Previous literature also delved into profitable betting strategies in the NFL. Paul and Weinbach (2011) identified that during the final hour before games, many uninformed bettors tend to heavily sway toward the favorite. Therefore, betting on the underdog when the percentage wagered on the favorite surges in the final hour resulted in a success rate exceeding 60% (Paul & Weinbach, 2011). Analysis by Wever and Aadland (2012) also demonstrated that betting on underdogs with substantial closing line spreads yielded winning percentages nearing 60%.

Dare and Holland (2004) identified profitability in betting on teams with strong overall performances relative to recent ones, achieving out-of-sample returns of 16.67%. Paul and Weinbach (2002) highlighted a profitable strategy of betting on the under when the total line for the game was 5, 6, or 7 points above the mean. Similarly, Pfitzner et al. (2009) explored a profitable strategy during the 2005-2006 season, generating a win rate of 62.5% when the predicted total differed from the actual total line by 10 or more points. There are certainly several strategies that have historically produced profits in the NFL betting market, however as technology becomes more adept, it is possible for bookmakers to develop methods to correct for inefficiencies. Regardless, this literature provides evidence of previously profitable strategies that established the potential for profitability within the NFL betting market.

Weather Impact on Point Spread and Totals Market Efficiency

Borghesi (2008) found that win rates for the over decreased as values for temperature, wind, and rain variables increased. Furthermore, he found that when games are played in the hottest, windiest, or rainiest quartiles, bets on the under can produce win rates of 56.13%, 53.32%, and 59.37% respectively. He used an OLS regression model to analyze the impact of independent weather variables (temperature, wind, rain, snow, and humidity) on the outcome of the game. Results suggested that weather was significant, and he found that rushing yards were decreased when the temperature was high and passing yards were reduced with high wind and rain. Turnovers also increased with rain. Overall, scoring was significantly reduced by all variables, temperature, wind, and rain.

Borghesi (2008) used a probit model to determine the relationship between weather variables and the outcome of over/under bets. The dependent variable was a value of 1 if a bet on the over won and 0 if a bet on the under won. He added an independent variable for the value of the closing over/under line. Results indicated that bets on the under were more likely to win when the total lines were highest. Furthermore, similar to the OLS model, high amounts of wind, heat, and rain significantly reduced the percentage of bets on the over that won. Borghesi (2007) also delved into the impact of temperature conditions on human performance, noting significant effects on both gross and fine motor skills due to exposure to low temperatures. He found that cold weather had a larger impact on the away team's performance than compared to heat. Therefore, home teams tended to perform better in cold weather games, presenting a profitable betting opportunity for wagers on home teams playing in cold weather.

Overview of Model

There are two primary categories of betting wagers: a "straight up" wager and a parlay wager. A "straight-up" wager is one wherein only a single outcome needs to occur for the wager to win. A typical spread or over/under wager exemplifies a straight-up wager. Typically, the odds on spread and over/under wagers hover around -110, meaning that a bettor staking $11 wins $10 if the bet hits. Parlay wagers, on the other hand, require the occurrence of two or more outcomes to secure victory. For example, a bet on the spread and total score. For instance, if a bettor elects to wager on the favored team with a spread of -2 and anticipates a low-scoring affair by opting for the under, given a total score of 48, the favored team must win by more than 2 points, and the cumulative score of both teams must fall below 48 points for the bet to succeed. Notably, both outcomes are required for victory. Parlays can also consist of multiple wagers and can even range into double-digit legs if the bet is on many games and outcomes. Given the reduced likelihood for wagers to simultaneously occur, the odds for parlays are significantly elevated compared to standard straight bets. For instance, the odds for a parlay involving both the spread and the total score are around +260, translating to a potential win of $26 on a $10 stake. The appeal of parlays and their high value is intriguing, as a bettor can still secure profits despite hitting fewer than 50% of his/her wagers.

The model in this paper is based upon a two-leg parlay, the home underdog to cover the spread, and the under to occur in the game. As the odds for spreads and totals are both typically -110, every time the parlay hits, a $10 bet returns $36 (a win of $26 as the $10 stake is returned).

Therefore, the home underdog/under parlay needs to win at a minimum rate of 28% to yield profits. The parlay includes the home underdog, so the wager would only be placed on games where the home team was the underdog.

Due to Davis et al.'s (2018) findings regarding the correlation of the under hitting and the underdog covering, the home underdog/under parlay hitting is the dependent variable of the model which equals a value of 1 if the parlay hit in the game and 0 if it did not hit. The dependent variable was based upon the foundation of previous research including Shank's (2018), Dare and Holland's (2004), Golec and Tamarkin's (1991), and Gray and Gray's (1997) who confirmed the biases and potential profitability of wagering on the home underdog. Paul and Weinbach (2002), Borghesi (2008), Shank (2018, 2019), and Marino (2017) also confirmed various biases favoring the under. While the parlay's success rate only needs to reach 28%, the hypothesis is that simply betting on the parlay every time the home team is the underdog will not result in profitability. Therefore, this paper aims to analyze variables that could potentially raise the likelihood of the parlay succeeding in specific matchups. These variables encompass factors such as temperature, wind, playing surface (natural grass or artificial turf), divisional matchups, venue type (indoor dome or outdoor stadium), prime time (games aired during prime television slots), point spread lines, and over/under lines.

III. Data And Variables

My data set is a time series data set that consists of every NFL game from the 2013-2019 seasons. The data set provides variables such as home and away teams, home and away scores, the total score of the game, difference in scores, if the game went to overtime, spread, odds for the game, home and away money lines, total line, divisional game status, surface, stadium (dome or outdoors), temperature, wind, and more. All games, including Super Bowls, that were played at a neutral location were excluded as there is no true home team. Games played at neutral locations were excluded because certain important characteristics that come with being the home team are not present at neutral locations. Examples include fan occupancy, crowd noise, familiarity with the stadium and surface, as well as other variables that come with home field advantage. The dependent variable also includes the home underdog, which is impossible to measure if the game is played at a neutral location.

Analysis was completed for all games played during the 2013-2019 seasons, except games played at neutral locations. Using Stata, dummy variables were created for if the home team was the underdog, if the home team covered the spread, if the under hit in the game, and a final variable for if all three variables (home team as an underdog, home team covering the spread, and under hitting) were true. If the value of the final variable was equal to 1, the parlay hit in that specific game. This variable was named "parlay_hits" and represented the dependent variable (the home underdog/under parlay). Using Stata, a regression was run on parlay_hits with all independent variables: magnitude of temperature, wind, prime time, divisional, surface, roof type, and magnitude of over/under lines and spreads. The roof variable (roof_numeric) was incorporated due to Marino's (2017) findings of correlation with the over and games played in a dome/closed roof stadium. Anderson (2019) also used roof type to examine performance against the spread. The variable was assigned a value of one if the game was played in an open, outdoor setting, and zero if it was played in a dome or stadium with a closed roof. The surface variable (surface_numeric) was assigned a value of one if the game was played on natural grass and zero if the game was played on artificial turf. Boulier et al. (2006) identified linkages between playing surfaces and teams' performance against the spread. Furthermore, Anderson (2019) found playing surfaces to be statistically significant when regressing for teams' performance against the spread as well. The wind variable (wind) was defined as the wind speed in miles per hour. Borghesi's (2008) results correlated high levels of wind with increased win rates for bets on the under. Therefore, wind was incorporated based on its impact on the under leg of the parlay.

To create the magnitude of the spread variable (magnitude_of_spread), values were assigned for individual ranges of the point spread lines. For example, a value of one was assigned for games when the spread was 0 to 2.5 points, two for a 3 to 6.5 point spread, three for games where the point spread was 7 to 10 points, and so on. The highest spread line for an away favorite was 18 points in favor of the New England Patriots in their matchup against the Miami Dolphins. The game was played on September 15, 2019, in Miami and the Patriots triumphed over the Dolphins, winning 43-0. The highest spread line for a home favorite throughout the sample time period was 27 points on October 13, 2013, when the Denver Broncos hosted the Jacksonville Jaguars in Denver. The final score ended 35-19 in favor of the Broncos, however, Denver did not cover the spread. Magnitude_of_spread was incorporated based on Shank's (2018) findings that home underdogs with large point spreads performed well against the spread.

A similar setup was used to measure the magnitude of the total line (magnitude_of_totalline). Values were assigned to specific ranges of the total lines. A value of one was assigned for games with totals of 38 and below, two for totals of 38.5 to 41, three for totals of and 41.5 to 45, and so on. The lowest total line was set at 35 points in a matchup between the Baltimore Ravens and the Pittsburgh Steelers. The game was played on December 29, 2019, in Baltimore and ended with the Ravens securing victory by a score of 28-10. The largest total line, set at 63.5, occurred in a matchup between the Kansas City Chiefs and the Los Angeles Rams. The Rams played at home and won by a slight margin with a final score of 54-51. The magnitude_of_totalline variable was included based on evidence from Paul and Weinbach (2002) that games with total lines significantly above the mean hit the under at a 59% rate.

The prime time variable (primetime) was defined as any game that was played after 6:00 pm EST. Primetime was included on the foundation of Anderson’s (2019) findings which concluded that prime time games resulted in closer scores than regularly scheduled games. Moreover, prime time was incorporated due to Shank’s (2018) results that home teams covered the spread more often in prime time games. Games played in "prime time" time slots were assigned a value of 1. The divisional game variable (div_game) was characterized as any game played against a divisional opponent. The NFL consists of eight separate divisions, each including four teams. Games played against divisional opponents were assigned a value of one. The divisional aspect was included by virtue of Shank's (2019) research which uncovered a correlation between divisional games and the home team covering the spread.

Temperature was measured in degrees Fahrenheit. Similar to the magnitude of the spread and the magnitude of total line variables, a variable was added for specific ranges of temperatures. A value of one was assigned for temperatures of 10 degrees and below, two for 11 to 20 degrees, three for games played in 21 to 30 degree temperatures, and so on. This variable was named "temp_numeric." The magnitude of the temperature variable was also included on the basis of research from Borghesi (2008) who discovered a correlation between temperature levels and win rates on over/under bets. The population regression function includes all of the magnitude variables as opposed to the original variables (spread_line, total_line, and temp). “Spread_line” consisted of all the point spread values for all games played in the NFL from 2013 to 2019. “Total_line" contained all of the total line values for games played from 2013-2019. "Temp" was composed of all of the temperatures in degrees Fahrenheit for all NFL games played from 2013 to 2019. Magnitude variables were used instead of original variables to prevent the effects of severe outliers in the data, such as an extremely high or low game-time temperature.

The regression identified statistically significant variables. Further analysis was done on statistically significant variables to identify which variables could lead to hitting the parlay at a higher rate. Games that possessed specific characteristics of the significant variables were analyzed. For example, if temperature was a significant variable and the relationship with the parlay hitting was negative, this would indicate that the parlay hit at a higher rate when temperatures were colder. Therefore, a model would be created that only included games played in cold temperatures to determine profitability. The population regression function is stated below:

Population Regression Function

$$Parlay\_hits_{gt} = \beta_0 + \beta_1 Primetime_{1gt} + \beta_2 Magnitude\_of\_spread_{2gt} +$$

$$\beta_3 Magnitude\_of\_totalline_{3gt} + \beta_4 Div\_game_{4gt} + \beta_4 Roof\_numeric_{4gt} +$$

$$\beta_5 Surface\_numeric_{5gt} + \beta_6 Temp\_numeric_{6gt} + \beta_7 Wind_{7gt} + \mathcal{E}_{gt}$$

IV. Analytical Framework

Subscripts g and t signal that the data set is composed of individual games (g) played from seasons 2013-2019 (t). Individual hypotheses were developed for each variable according to prior research and knowledge. Surface_numeric was hypothesized to have a positive correlation with the parlay hitting. Prior research has shown that teams with grass as their home field tend to perform better against the spread (Anderson, 2019). Grass has also been shown to boost away teams' performance as opposed to turf, which could result in a closer game, and increase the odds for the parlay to hit (Anderson, 2019). Temp_numeric was hypothesized to have a negative correlation with the parlay hitting. As temperatures decrease, the likelihood for the parlay to hit will increase. This is because colder temperatures often resulted in fewer points (Shank, 2019). In cold temperatures, it is more difficult to complete precise body movements that are required in football, making offensive movements more difficult and resulting in fewer points and closer games (Borghesi, 2007). Wind was expected to have a strong and positive relationship with the parlay hitting. As wind speeds increase, it is more difficult for players to throw and catch, resulting in fewer points and closer games (Borghesi, 2008). Primetime was also expected to have a strong and positive relationship with the parlay hitting as games in prime time are often played between relatively evenly matched opponents, creating an atmosphere for the game to be decided by a few points (Anderson, 2019), (Shank, 2018). Furthermore, games played against divisional opponents were also expected to increase the likelihood for the parlay to hit. This is because in the NFL, every team plays the other three teams in their division twice every season, once at home and once away. Analysis from Shank (2019) revealed that divisional games resulted in a higher probability for the game to hit the under. Due to this familiarity, it is hypothesized that divisional games will more often result in fewer points and tighter scores.

The magnitude of the spread variable was expected to have a negative relationship with the parlay hitting. As the spread gets lower, to less than two and a half points, opponents are more evenly matched which will more likely result in closer final scores of games. The hypothesis for the magnitude of the total line variable was a slightly positive relationship with the parlay hitting. Shank (2018) found that games with very high or very low totals were more likely to cover the over, therefore games with totals closer to the mean, are more likely to cover the under. Lastly, the roof variable is estimated to have a positive relationship with the parlay hitting. Marino (2017) found a relationship between the over and games played in indoor settings. Therefore, games played outside are expected to raise the likelihood for the under and the parlay to hit due to outside weather effects.

In addition to the main research question, a regression was run for a straight bet on the home underdog covering the spread and a straight bet for the under hitting. Regressions were also run for alternative variations of the main dependent variable. Eight total regressions were run including the main regression with the home underdog and under parlay (parlay_hits). All regressions were checked for robustness and multicollinearity tests were completed. The seven other regressions all included the eight original independent variables and consisted of the following dependent variables, the regression names for the dependent variables are in parentheses:

1. The home underdog covering the spread (HUnderdog_covers)
2. The under hitting in the game (cover_total)
3. The home team covering the spread, regardless of favorite or underdog status, (cover_spread)
4. The home favorite covering the spread (HFavoritecovers)

5. A parlay with the home favorite covering the spread and the over hitting (HomeFavoriteoverlay)
6. A parlay with the favorite team covering the spread (regardless of home or away) and the over hitting (FAV_OVER_LAYhits)
7. A parlay with the underdog covering the spread (regardless of home or away) and the under hitting (DOG_UNDER_LAYhits).

V. Results

Home Underdog/Under Parlay Profitability and Regression

Using Stata, the number of games in which the parlay would have hit was totaled and divided by the games in which the parlay would have been placed (all games where the home team was the underdog). This resulted in a value of 24.26%, confirming the hypothesis, and revealing that simply betting the parlay for all games where the home team was the underdog during the 2013-2019 seasons did not break the threshold for profitability of 28%. Consequently, during this time period, the parlay would have been unprofitable. Therefore, a regression was completed with the eight independent variables to determine their individual impact on the parlay hitting and to examine a variation of game characteristics that could yield a higher percentage of profitability.

The regression was completed and checked for robustness on the home underdog and under parlay (parlay_hits) with the original independent variables: primetime, magnitude_of_spread, magnitude_of_totalline, temp_numeric, surface_numeric, div_game, wind, and roof_numeric. Magnitude_of_spread and div_game were statistically significant at the one percent and ten percent levels, respectively. Roof_numeric was omitted due to collinearity. The statistical significance of magnitude_of_spread was 0.003 with a coefficient of -0.019. This signified that as the level of the spread decreased, the parlay was more likely to hit. For example, if the spread was a low number such as 1.5 points, the game had a higher probability of hitting the home underdog/under parlay. Div_game was significant with a p-value of 0.075 and a coefficient of 0.029. Therefore, div_game had a positive correlation with the parlay hitting, indicating that games played against divisional opponents raised the likelihood for the home underdog/under parlay to hit. This aligned with Shank’s (2019) findings that games played against divisional opponents increased the probability for the under to hit. A test for multicollinearity was completed with the VIFs for all variables falling between 1 and 1.1, with a mean VIF of 1.04. Table 1 provides a summary table of descriptive statistics for all variables. Table 2 details the linear estimates of the home underdog and under (parlay_hits) regression model. Table 3 provides the VIF estimates for the multicollinearity check for the parlay_hits regression.

Profitable Parlay Strategy

Using the statistically significant variables from the parlay_hits regression, a profitable strategy was found during the 2013-2019 time period. When the home underdog was playing a divisional opponent and the spread was between 10 and 13.5 points, the home underdog/under parlay hit 45% of the time, well exceeding the 28% threshold for parlay profitability. However, these characteristics occurred in only 20 games throughout the time period, with the parlay hitting in nine of them, which is a very small sample size.

Over/Under Profitability and Regression

Profitability analysis was also completed for overs and unders throughout the sample time period. During the 2013-2019 seasons, the under hit 49.6% of the time, the over hit 47.6% of the time, and 51 games pushed. The under hitting (cover_total) was regressed and checked for robustness with the eight independent variables, div_game, primetime, magnitude_of_spread, magnitude_of_total, surface_numeric, temp_numeric, roof_numeric, and wind. Cover_total was equal to a value of 1 if the under hit and 0 if the over hit. Div_game was statistically significant with a value of 0.024 and a coefficient of 0.063. This indicated that games played against divisional opponents raised the likelihood that the under hit. Therefore, analysis was completed for all divisional games during 2013-2019. The under hit at a rate of 53.3% in games played between divisional opponents during this time period, which exceeded the threshold for profitability of 52.4%. This is also in line with Shank's (2019) research which also revealed that games played against divisional opponents had a higher probability of hitting the under. Furthermore, wind was also significant with a p-value of 0.004 and a coefficient of 0.007, indicating that wind speed slightly increased the probability for the under to hit. When adding a variable for wind speed, the profitability jumped to a 61.11% hit rate for the under in games played against divisional opponents when the wind speed was 10 miles per hour or greater.

These results aligned with Borghesi's (2008) findings that games with high levels of wind more often resulted in the under.

Given that div_game and wind were significant, contrarily, games played in low wind and against non-divisional opponents were more likely to hit the over. Therefore, analysis was completed in non-divisional games played when the wind speed was 10 mph or less. The over hit 53.3% of the time, which also would have been profitable during the 2013-2019 seasons. The regression was tested for multicollinearity resulting in VIFs ranging from 1 to 1.1 for all variables and a mean VIF of 1.04. Table 2 details the linear estimates of the over/under (cover_total) regression. Table 4 provides the VIF estimates for the multicollinearity check for the cover_total regression.

Home Underdog Covering the Spread Profitability and Regression

Throughout the 2013-2019 seasons, the games in which the home underdog covered the spread were totaled and divided by all games where the home team was the underdog. The home underdog covered the spread 48.6% of the time during 2013-2019. A regression was completed for the home team covering the spread (HUnderdog_covers). Magnitude_of_spread was significant with a p-value of 0.000 and a coefficient of -0.049, indicating that as the spread decreased, the home underdog was more likely to cover the spread. This contradicted Shank's (2018) findings that the home underdog covered the spread 63% of the time when the spread was 10 points or more. Multicollinearity was tested and all VIF values ranged from 1 to 1.1, resulting in a mean VIF of 1.04. However, despite the significance of the magnitude_of_spread, there was nothing found to be profitable in wagering on the home underdog to cover the spread during the 2013-2019 seasons. Table 2 details the linear estimates of the home underdog covering the spread (HUnderdog_covers) regression. Table 5 provides the VIF estimates for the multicollinearity check for the HUnderdog_covers regression.

Home Team Covering the Spread Profitability and Regression

The home team only covered the spread 46.5% of the time during the 2013-2019 seasons. The variable for the home team covering the spread (cover_spread) was regressed, and checked for robustness, with div_game, primetime, magnitude_of_spread, magnitude_of_total, surface_numeric, temp_numeric, roof_numeric, and wind. Cover_spread was equal to a value of 1 if the home team covered the spread and 0 if the away team covered the spread. Primetime, surface_numeric, and wind were statistically significant. Roof_numeric was omitted for collinearity. Primetime was significant with a value of 0.040 and a coefficient of 0.072, indicating that games played during primetime were more likely for the home team to cover the spread. This reinforced Shank's (2018) findings of the same relationship between prime time games and the home team covering the spread. Surface_numeric was significant with a value of 0.007 and a coefficient of -0.084. This revealed that the home team was more likely to cover the spread when the game was played on turf. This contradicted Anderson's (2019) research that teams performed better against the spread if their home field was grass. However, the surface_numeric significance could be questionable as it is possible that teams whose home field was turf, were more dominant and successful in covering the spread during this time period, so this variable may be slightly biased. Furthermore, it is not uncommon for NFL teams to change their home field playing surface. Within recent years several teams have switched their home field from turf to grass, which could also impact results. Wind was also significant with a p-value of 0.098 and a coefficient of 0.004, specifying that games played in higher wind slightly increased the probability that the home team covered the spread.

Using the statistically significant variables, the percentage that the home team covered the spread in all prime time games was calculated, revealing that during 2013-2019, the home team covered the spread 49.6% of the time in all prime time games. When field surface was taken into account, the hit rate for the home team covering the spread jumped to 53.9% in prime time games that were played on turf. Furthermore, in games played during prime time and when the total line was 49.5-53, the home team covered the spread at a rate of 53%. The home team also covered the spread 53% of the time when the game was played in prime time and the spread was between 14-17.5 points. However, the home team only covered the spread 49.2% of the time during prime time games when the wind speed was 10 mph or higher. Table 2 details the linear estimates of the home team covering the spread (cover_spread) regression. Table 6 provides the VIF estimates for the multicollinearity check for the cover_spread regression.

Home Favorite Covering the Spread Profitability and Regression

The home favorite covered the spread 46.6% of the time in games played during the 2013-2019 seasons. A regression was run on the home favorite covering the spread (HFavoritecovers) with the same independent variables. The regression was checked for robustness. Roof_numeric was omitted due to collinearity. A multicollinearity test was completed, resulting in a mean VIF of 1.04. Primetime was significant with a p-value of 0.020

and a coefficient of 0.075, suggesting that prime time games raised the likelihood of the home favorite covering the spread. This aligned with Shank's (2018) and Anderson's (2019) research on the impact of prime time games. Magnitude_of_spread was the most significant with a p-value of 0.000 and a coefficient of 0.044, indicating that as the spread increased, the likelihood that the home favorite covered the spread also increased. This is similar to the relationship Davis et al. discovered between high point spreads and the favorite/over parlay. Div_game was also significant with a p-value of 0.058 and a coefficient of -0.048, illustrating that the home favorite was more likely to cover when playing against non-divisional opponents. This was emphasized by Shank (2018) who initially discovered the relationship between home teams covering the spread against non-divisional opponents. Surface_numeric was also significant with a p-value of 0.005 and a coefficient of -0.082, revealing that games played on turf raised the likelihood that the home favorite covered the spread. Despite the statistical significance of several variables, there was no profitable strategy found regarding wagers on the home favorite to cover the spread from 2013 to 2019. Table 2 details the linear estimates of the home favorite covering the spread (HFavoritecovers) regression. Table 7 provides the VIF estimates for the multicollinearity check for the HFavoritecovers regression.

Home Favorite/Over Parlay Profitability and Regression

Davis et al.'s (2018) research on correlated parlays in college football indicated a relationship between the favorite covering the spread and the over hitting as well as the underdog covering the spread and the under hitting. This is the case because when the favorite covers the spread it occasionally results in a blowout win, resulting in many points and the over, whereas when the underdog covers the spread, it is likely to be a close game and result in fewer points. Therefore, an analysis was completed with the home favorite and the over parlay in the NFL. During the 2013-2019 seasons, the home favorite/over parlay hit at a rate of 23.2%, which would not have been profitable during this period of time. Therefore, regression was run, which was checked for robustness, with the home favorite/over parlay as the dependent variable (HomeFavoriteoverlay) with the eight independent variables: primetime, surface_numeric, magnitude_of_totalline, magnitude_of_spread, roof_numeric, div_game, temp_numeric, and wind. Roof_numeric was omitted for collinearity. A multicollinearity test was completed resulting in VIF values falling between 1 and 1.2 for all variables and a mean VIF of 1.05. Primetime was significant with a value of 0.003 and a coefficient of 0.181, revealing that prime time games resulted in a greater likelihood for the parlay to hit. This finding aligned with Anderson's (2019) and Shank's research (2018) about the correlation between prime time games and the home team covering the spread. Magnitude_of_spread was significant with a value of 0.004 and a coefficient of 0.066, suggesting that as the spread increased, the likelihood that the parlay hit also increased. Davis et al. (2018) also revealed a relationship between a high point spread and an increased hit rate of a favorite/over parlay in general. However, the coefficient was not extremely high, which could indicate that a range of spreads closer to the mean are responsible for the positive coefficient. Therefore, analysis was completed for all ranges of the magnitude_of_spread variable. Div_game was significant with a value of 0.002 and a coefficient of -0.169, signaling that games played against non-divisional opponents had a higher probability that the parlay to hit. This supported Shank's (2019) research which also found that non-divisional games more often resulted in the over. Surface_numeric was significant with a value of 0.057 and a coefficient of -0.112, revealing that games played on turf had a higher probability of hitting the parlay.

Using the statistically significant independent variables, analysis was completed to determine the profitability of the home favorite/over parlay for games played during prime time, against non-divisional opponents, and with the six ranges of the spread that comprise the magnitude_of_spread variable. In the first analysis, with the first range of the spread (2.5 points or less), in prime time games against non-divisional opponents, the home favorite/over parlay hit 35% of the time, which was enough to break the 28% threshold for profitability during the 2013-2019 seasons. With the next range of spreads (3-6.5 points), in prime time and non-divisional games, the home favorite/over parlay hit exactly the breakeven point for profitability at 28%. In prime time and non-divisional games when the spread was 7-9.5 points, the home favorite/over parlay hit 35% of the time, which also would have been profitable. Non-divisional, prime time games when the spread was 10-13.5 points, hit the home favorite/over parlay 33% of the time – also profitable. Similarly, in non-divisional and prime time games, with a spread between 14-17.5 points, the home favorite/over parlay hit 37.5% of the time, which was also profitable. However, during 2013-2019 there were only 8 instances where the home favorite was playing in a prime time, non-divisional game, when the spread was 14-17.5 points, which is an extremely small sample size. For the last range of spreads (18 points and up), there were 0 instances where the home favorite was playing in a prime time, non-divisional game with an 18+ point spread, so the analysis was not possible.

It is quite stunning that essentially all ranges of spreads generated profitable results when combined with prime time and non-divisional factors, with the exception of the 0 instances of the 18+ point range and the 3-6.5 point range, which broke even at 28%. Moreover, in games played only in prime time and against non-divisional opponents, the home favorite/over parlay hit 31.5% of the time. Overall, with a combination of spreads of 2.5 and less, 5.5 to 6.5, 8 to 9.5, and 11 to 13.5 points, the home favorite/over parlay hit 41.5% of the time in prime time games played against non-divisional opponents, which would have generated significant profits. Table 2 details the linear estimates of the home favorite and over parlay (HomeFavoriteoverlay) regression. Table 8 provides the VIF estimates for the multicollinearity check for the HomeFavoriteoverlay regression.

Favorite/Over Parlay Profitability and Regression

Regardless of home or away status, the favorite and over parlay hit 22.2% of the time throughout all games played during the 2013-2019 seasons, which was not profitable. Therefore, a regression was run with the favorite/over parlay hitting as the dependent variable (FAV_OVER_LAYhits) along with the eight original independent variables. Roof_numeric was omitted for collinearity. The regression was checked for robustness and div_game and wind were statistically significant. Div_game was significant with a p-value of 0.012 and a coefficient of -0.058, revealing that non-divisional games had a higher likelihood for the favorite/over parlay to hit. This was supported by Shank's (2018) findings that non-divisional games more likely resulted in the home team covering the spread and the over hitting. Wind was statistically significant with a p-value of 0.046 and a coefficient of -0.004. This suggested that as wind speeds decreased, the probability for the favorite/over parlay to hit slightly increased. This relates to Borghesi's (2008) results that low levels of wind more often result in the over. Multicollinearity was tested and all VIF values ranged from 1 to 1.1 with a mean VIF of 1.04. Nothing was found to be profitable with either of the significant variables for the favorite/over parlay. Table 2 details the linear estimates of the favorite and over parlay (FAV_OVER_LAYhits) regression. Table 9 provides the VIF estimates for the multicollinearity check for the FAV_OVER_LAYhits regression.

Underdog/Under Parlay Profitability and Regression

In all games played during the 2013-2019 seasons, the underdog/under parlay hit 455 times for a 24.3% hit rate – not enough to turn a profit. With the underdog/under parlay hitting as the dependent variable (DOG_UNDER_LAYhits), a regression was run with the original independent variables. The regression was checked for robustness. Roof_numeric was omitted for collinearity. Multicollinearity was tested, resulting in the values for each variable ranging from 1 to 1.1 and a mean VIF of 1.04. Only surface_numeric was statistically significant with a p-value of 0.082 and a coefficient of 0.045, revealing that games played on grass had a higher probability of hitting the parlay. This aligned with Anderson's (2019) findings regarding games played on grass. However, despite the significance of surface_numeric, there was nothing found to be profitable in wagering on underdog/under parlays. Table 2 details the linear estimates of the underdog and under parlay (DOG_UNDER_LAYhits) regression. Table 10 provides the VIF estimates for the multicollinearity check for the DOG_UNDER_LAYhits regression.

Profitability Comparisons with Previous Research

Contrary to Shank's (2018) findings that home underdogs covered the spread 63% of the time when the spread was 10 points or more, during 2013-2019, teams that were home underdogs and projected to lose by 10 or more points, only covered the spread 44.74% of the time. However, during the time span within this, there were only 38 games where this was the case, which is a very small sample size.

Furthermore, contrary to Borghesi's (2007) findings that home underdogs covered the spread at a higher rate within the last three weeks of the season, during the final three weeks of the season, from 2013 to 2019, the home team only covered the spread at a rate of 43.75%. This could be evidence that since Borghesi's (2007) research, sportsbooks have adjusted lines to account for the home underdog strategy, especially in the concluding weeks of the regular season. In alignment with Shank's (2018) findings that home teams cover the spread around 49% of the time, during prime time games during the 2013-2019 time period, the home team covered the spread 49.6% of the time. These results attest to the overlapping time periods between this study and Shank's (2018), as well as the persistent efficiency of bookmakers.

Regressions With Magnitude and Original Variables

Regressions were also completed with different combinations of each of the original variables (spread_line, total_line, and temp) and magnitude variables (magnitude_of_spread, magnitude_of_totalline, and temp_numeric). All regressions were run on parlay_hits, the main dependent variable of the original model. To avoid collinearity, magnitude variables were never combined with the original variables in the regressions. For example, if a regression was run with magnitude_of_spread, spread_line was not included. With the combinations of variables, without overlapping a magnitude and an original variable, seven other combinations of regressions could be run, excluding the original population regression function which included the three magnitude variables: temp_numeric, magnitude_of_spread, and magnitude_of_totalline. All of the regressions included the other five independent variables, primetime, div_game, wind, surface_numeric, and roof_numeric.

When the regression was run and checked for robustness, on parlay_hits with temp_numeric, magnitude_of_totalline, and the original spread_line variable (the exact spread values for every game), div_game became no longer significant and spread_line became significant with a p-value of 0.000 and a coefficient of -0.016. This could indicate slight collinearity between div_game and magnitude_of_spread. However, a multicollinearity test was completed and all VIFs still ranged from 1 and 1.1 with a mean of 1.04.

Furthermore, when spread_line was included with total_line (the exact total line values for every game) and temp_numeric, spread_line was still the only significant variable with the same value of 0.000 with the same coefficient of -0.016. Moreover, when temp_numeric was replaced with temp and regressed with total_line and spread_line, only spread_line was significant again at 0.000 and a coefficient of -0.016. Both regressions were checked for robustness and multicollinearity, resulting in the same previous VIF results.

When magnitude_of_spread was regressed with magnitude_of_totalline and temp, magnitude_of_spread and div_game were statistically significant. Magnitude of spread was again significant at 0.003 with a coefficient of -0.019, and div_game was significant at 0.075 with a coefficient of 0.029 once more. The regression was checked for robustness and multicollinearity and the mean VIF was 1.04 again.

When parlay_hits was regressed with magnitude_of_spread, total_line, and temp, magnitude_of_spread remained significant at 0.003 with the same coefficient of -0.019. Div_game also remained significant at 0.077 with the same coefficient of 0.029. Similarly, when magnitude_of_spread was regressed with total_line and temp_numeric, magnitude_of_spread and div_game were significant again with the same values and coefficients. Lastly, when parlay_hits was regressed with spread_line, magnitude_of_totalline, and temp, spread_line was significant at 0.000 with a coefficient of -0.016. All regressions were checked for robustness and multicollinearity, resulting in a mean VIF of 1.04 for all of them. The only relationship seemed to be between magnitude_of_spread and div_game, which could indicate multicollinearity. However, all of the VIF tests indicated low levels of multicollinearity. No other variables seemed to influence any of the remaining independent variables.

VI. Discussion

Three strategies were found to be most profitable during the sample period (2013-2019). The first was the home underdog and under parlay when the game was played between divisional opponents and the spread was between 10 and 13.5 points. This strategy was profitable at a rate of 45%, well exceeding the threshold for parlay profitability at 28%. The second strategy was a parlay with the home favorite and the over in prime time, non-divisional games when the point spread was 2.5 points or less, 5.5 to 6.5 points, 8 to 9.5 points, or 11 to 13.5 points. This parlay hit at a rate of 42%, also exceeding the 28% threshold for parlay profitability. The last strategy that was most profitable was a straight bet on the under when the game was played between divisional teams and the wind speed was 10 mph or greater. This strategy hit 61% of the time, exceeding the 52.4% straight bet threshold for profitability. However, it is important to note that these results were only profitable during the sample time period and will not likely persist due to the extremely efficient nature of bookmakers.

Roof_numeric was omitted for collinearity in all of the regressions which made it impossible to analyze the impact of a closed vs. open stadium setting. In general, for all regressions, magnitude_of_spread, div_game, primetime, wind, and surface_numeric were the only statistically significant variables. Each regression had different statistically significant variables and there was no obvious pattern that stood out between strategies and statistically significant variables.

Magnitude_of_spread was significant in the parlay_hits, HomeFavoriteoverlay, HFavoritecovers, and HUnderdog_covers regressions. To specify, the level of the point spread was statistically significant when considering the home underdog/under parlay, home favorite/over parlay, home favorite covering the spread, and home underdog covering the spread. The home team aspect is a commonality between the four dependent variables.

Magnitude_of_spread had a slightly negative relationship with the home underdog/under parlay hitting, a slightly positive relationship with the home favorite covering the spread, a slightly positive relationship with the home favorite/over parlay hitting, and a slightly negative relationship with the home underdog covering the spread. Overall, magnitude_of_spread had positive relationships when the home favorite was included and negative relationships when the home team was the underdog. Therefore, in general, this indicated that the home favorite tended to perform better when the point spread was higher. Furthermore, the home underdog tended to perform better against the spread when it was a smaller number. However, it was likely that there was a parabolic relationship between the point spread and the home underdog/under parlay. The parlay was most successful in divisional games when the point spread was 10-13.5 points, which is evidence of a parabolic relationship with the point spread and the parlay hitting, as 10-13.5 points is within the middle range of all point spread values.

Div_game was significant in the parlay_hits, cover_total, HFavoritecovers, HomeFavoriteoverlay, and the FAV_OVER_LAYhits regressions. To clarify, the independent variable for games played against divisional opponents was statistically significant when regressed with the dependent variables for the home underdog/under parlay hitting, over and under hitting, home favorite covering the spread, home favorite/over parlay hitting, and favorite/over parlay hitting. Div_game had a positive relationship with the home underdog and under parlay hitting, a negative relationship with the home favorite and over parlay hitting, a negative relationship with the home favorite covering the spread, a positive relationship with the under hitting, a negative relationship with the over hitting, and negative relationship with the favorite and over parlay hitting. Div_game clearly had a negative relationship with the over hitting, as all regressions with dependent variables that include the over presented a negative coefficient for div_game. This further supported Shank's (2018) findings that games played against divisional opponents favor the under. Div_game also consistently had a negative relationship with the favorite covering the spread, as all regressions with dependent variables which included the favorite or home favorite resulted in negative coefficients for div_game. Therefore, home favorites and straight-up favorites performed better against the spread in non-divisional games.

Surface_numeric was significant in the HomeFavoriteoverlay, HFavoritecovers, DOG_UNDER_LAYhits, and cover_spread regressions. Therefore, it was statistically significant for the home favorite and over parlay hitting, the home favorite covering the spread, the underdog and under parlay hitting, and the home team covering the spread. Surface_numeric had a negative relationship with the home favorite and over parlay, a negative relationship with the home favorite covering the spread, a positive relationship with the underdog and under parlay hitting, and a negative relationship with the home team covering the spread. Surface_numeric consistently had a negative relationship when the dependent variable included the home team. Therefore, these findings revealed that the home team performed better against the spread when the field was turf. This contradicted Anderson's (2019) findings that teams whose home fields were grass performed better against the spread. These differences likely reflect the differing sample time periods and could also be influenced by teams switching their playing surfaces. However, it is possible that these findings could also reveal that teams whose home fields were turf were more dominant during the sample period, so the surface variable likely does not hold much weight.

Primetime was significant in the HomeFavoriteoverlay, HFavoritecovers, and cover_spread regressions. It was significant in the home favorite and over parlay hitting, the home favorite covering the spread, and the home team covering the spread. In the three regressions where primetime was statistically significant, it had a positive relationship with all three of the dependent variables. The three dependent variables all included a home team factor, indicating that home teams and home favorites performed better against the spread in prime time games, supporting Shank's (2018) and Anderson's (2019) findings.

Wind was significant in the cover_total, FAV_OVER_LAYhits, and DOG_UNDER_LAYhits. To clarify, wind was statistically significant in the over/under, favorite and over parlay, and underdog and under parlay regressions. Wind was slightly positive for the over/under regression, revealing that games with high levels of wind hit the parlay more often. Wind was slightly negative with the favorite and over parlay hitting, and slightly positive for the underdog and under parlay hitting. None of the wind coefficients were strong, however, it is clear that the under hit more often in windier games, supporting Borghesi's (2008) findings.

VII. Conclusions

This paper contributes to the existing literature on market efficiency in the NFL betting market as it sheds light on the value and profitability of various parlay wagers in the NFL. These findings can be useful for sportsbooks and bookmakers as they may want to adjust lines to limit the profitability of specific strategies that have proven to be profitable in this paper. Furthermore, the NFL could take this information into account to help identify the types of games that produce high/low scores. This could help the league decide games to broadcast at specific times to maximize fan engagement and audience.

For future research within the realm of betting efficiency and parlay betting in the NFL, it would be interesting and useful to incorporate offensive and defensive statistics. For example, a regression could be run including passing yards per game, rushing yards per game, and first downs to identify exactly which statistics are the most indicative of the outcome of the game. This would especially be interesting to incorporate when analyzing parlay and over/under wagers, as it is likely that specific statistics will have a higher likelihood of hitting parlays and totals. For example, if two teams are playing who both average high values for passing yards per game, it is presumably more likely that the game will be high scoring and the over will hit. Moreover, if a team with a high level of average passing yards is playing a defense that gives up a lot of passing yards, the game would more likely favor the over and the team with the high-scoring offense. There are many different angles to approach an analysis with game statistics and for further research, it would be interesting to consider the impact of average statistics on outcomes of games and potentially profitable strategies.

It would also be intriguing to implement power rankings into the regression. The teams with the higher ranking would more likely win games, but it would be interesting to analyze how power rankings relate to teams covering the spread. Most power rankings, which rank teams in the NFL weekly, are primarily based on the previous week's performance, so there is an aspect of momentum that is included as well. Essentially, it would be a measurement of how efficiently bookmakers account for momentum and a team's previous performance.

Injuries would also be compelling to take into account. All positions and players have a different effect on the game, so a point system would need to be developed based on player and player value in terms of points. For example, if a star quarterback is injured, his point value would be relatively high, presumably around 5-6 points. However, if a second-string linebacker was out, the point value would be much lower. It would be difficult to incorporate into a regression, but it would be interesting to analyze the point impact of players and how injuries could subsequently impact the outcome of games.

It would have been useful to explore a larger data set with more games to increase the sample size, however, games played prior to 2013 lacked relevance, as playing styles and coaching philosophies change, and games played after the 2019-20 season were impacted by Covid-19 factors which would have skewed the data and results. A probit model also would have been interesting to incorporate in addition to the linear estimations to possibly achieve more accurate results. In the future, it would also be useful to implement a quadratic variable for point spread and total line variables. It is likely that especially the spread had a parabolic relationship with the home underdog and under parlay hitting. In addition, it would be useful to implement a solution for the collinearity of the roof variable, so it would be possible to analyze the impact of playing in a dome vs. an outdoor stadium.

With the legalization and increasing popularity of sports betting around the world, there is ample opportunity for research and data analysis. The sports betting market can provide insight into alternative financial markets with regard to overall market efficiency. Live betting, which is a term used for betting on a sporting event while it is occurring, could also be an avenue for future research. Live betting is probably the most similar to a financial market due to the constantly fluctuating prices and lines. For live betting, as the game is going on, bookmakers are constantly changing lines and prices on player statistics, including passing yards, rushing yards, receptions in football and points, rebounds, and assists in basketball, and so on, depending on the nature of the sport. This type of betting most similarly reflects the nature of trading stocks in financial markets. Analysis of live betting could give insight into the intricacies and functions of a market that is constantly shifting, which could be applied to similar financial markets. The analysis would also be more feasible within a live sports betting market due to the smaller size of the market and lack of outside influences.

Sports betting markets mirror financial markets and provide a unique medium for examination of the functionality and efficiency of other similar markets. This study contributes to the ongoing analysis of efficiency within the NFL betting market and can be used as a foundation for further research. The sports betting industry is primed for rapid growth following legalization and presents an avenue for insight into the intricacies of market functions.

Table 1

Descriptive Statistics

Variable	Obs	Mean	Std. Dev.	Min	Max
primetime	1835	.209	.407	0	1
away score	1835	21.689	9.708	0	59
home score	1835	23.906	10.304	0	57
result	1835	2.217	14.274	-49	52
total	1835	45.596	14.039	6	105
overtime	1835	.056	.23	0	1
away money line	1834	89.214	267.807	-1300	2173
home money line	1834	-134.477	326.179	-5000	880
spread line	1835	2.337	5.804	-18	27
magnitude of spread	1835	2.221	.996	1	6
HomeFavorite2	1830	.66	.474	0	1
away spread odds	1834	-58.462	89.989	-134	120
home spread odds	1834	-57.645	90.666	-133	121
total line	1835	45.453	4.154	35	63.5
magnitude of totalline	1835	3.517	1.114	1	7
under odds	1834	-70.704	78.533	-125	113
over odds	1834	-77.874	72.218	-125	113
div game	1835	.366	.482	0	1
roof numeric	1835	.757	.429	0	1
surface numeric	1835	.579	.494	0	1
temp	1359	58.857	17.51	-6	97
temp numeric	1359	6.339	1.762	1	10
wind	1359	8.107	5.128	0	71
HomeUnderdog	1864	.334	.472	0	1
cover spread	1780	.488	.5	0	1
cover total	1818	.51	.5	0	1
parlay hits	1801	.086	.28	0	1
HomeFavoriteoverlay	437	.643	.48	0	1
HFavoritecovers	1869	.302	.459	0	1
HUnderdog covers	1869	.162	.368	0	1
FAV OVER LAYhits	1869	.222	.415	0	1
DOG UNDER LAYhits	1869	.243	.429	0	1

Table 2
Linear Regressions

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	parlay_hits	cover_total	HUnderdog_covers	HFavoritecovers	cover_spread	HomeFavoriteoverlay	FAV_OVER_LAYhits	DOG_UNDER_LAYhits
primetime	-.009	-.044	-.006	.075**	.072**	.181***	.028	-.023
	(.019)	(.035)	(.025)	(.032)	(.035)	(.061)	(.03)	(.029)
magnitude_of_spread	-.02***	.004	-.049***	.045***	-.003	.067***	.006	.009
	(.007)	(.014)	(.009)	(.013)	(.014)	(.023)	(.011)	(.012)
magnitude_of_totalline	-.006	.012	-.014	.001	-.014	-.024	.001	.002
	(.007)	(.013)	(.009)	(.012)	(.013)	(.025)	(.011)	(.011)
div_game	.03*	.064**	.023	-.048*	-.029	-.17***	-.059**	.023
	(.017)	(.028)	(.021)	(.025)	(.029)	(.056)	(.023)	(.025)
roof_numeric								
surface_numeric	.01	.041	-.004	-.082***	-.084***	-.113*	-.018	.046*
	(.017)	(.031)	(.023)	(.029)	(.031)	(.059)	(.026)	(.026)
temp_numeric	0	-.008	.005	-.01	-.004	-.015	-.01	-.006
	(.005)	(.008)	(.006)	(.007)	(.008)	(.016)	(.007)	(.007)
wind	.002	.007***	.001	.003	.004*	-.007	-.004**	.002
	(.002)	(.003)	(.002)	(.002)	(.003)	(.006)	(.002)	(.002)

_cons	.12**	.412***	.276***	.292***	.583***	.793***	.331***	.197***
	(.048)	(.081)	(.061)	(.073)	(.082)	(.157)	(.067)	(.071)
Observations	1330	1344	1359	1359	1312	322	1359	1359
R-squared	.01	.013	.022	.027	.013	.096	.01	.005

Robust standard errors are in parentheses
**** p<.01, ** p<.05, * p<.1*

Table 3
Parlay_hits

VIF	1/VIF
1.080	0.929
1.050	0.950
1.050	0.954
1.030	0.966
1.020	0.976
1.020	0.982
1.020	0.984
1.040	

Table 4
Cover_total

VIF	1/VIF
1.080	0.926
1.050	0.948
1.050	0.952
1.030	0.967
1.030	0.974
1.020	0.982
1.020	0.984
1.040	

Table 5

HUnderdog_covers

VIF	1/VIF
1.080	0.928
1.050	0.949
1.050	0.954
1.030	0.968
1.030	0.975
1.020	0.981
1.020	0.982
1.040	

Table 6

Cover_spread

VIF	1/VIF
1.080	0.929
1.050	0.951
1.050	0.956
1.030	0.967
1.020	0.977
1.020	0.982
1.020	0.984
1.040	

Table 7

HFavoritecovers

VIF	1/VIF
1.080	0.928
1.050	0.949
1.050	0.954
1.030	0.968
1.030	0.975
1.020	0.981
1.020	0.982
1.040	

Table 8

HomeFavoriteoverlay

VIF	1/VIF
1.120	0.892
1.090	0.919
1.050	0.952
1.050	0.954
1.040	0.964
1.010	0.988
1.010	0.989
1.050	

Table 9

FAV_OVER_LAYhits

VIF	1/VIF
1.080	0.928
1.050	0.949
1.050	0.954
1.030	0.968
1.030	0.975
1.020	0.981
1.020	0.982
1.040	

Table 10

DOG_UNDER_LAYhits

VIF	1/VIF
1.080	0.928
1.050	0.949
1.050	0.954
1.030	0.968
1.030	0.975
1.020	0.981
1.020	0.982
1.040	

References

Anderson, S. (2019). NFL Betting Market Efficiency: Finding a Profitable Betting Strategy.

Borghesi, R. (2007). The home team weather advantage and biases in the NFL betting market. *Journal of Economics and Business*, *59*(4), 340-354.

Borghesi, R. (2008). Weather biases in the NFL totals market. *Applied Financial Economics*, *18*(12), 947-953.

Boulier, B. L., Stekler, H. O., & Amundson, S. (2006). Testing the efficiency of the National Football League betting market. *Applied Economics*, *38*(3), 279-284.

Dare, W. H., & Holland, A. S. (2004). Efficiency in the NFL betting market: modifying and consolidating research methods. *Applied Economics*, *36*(1), 9-15.

Davis, J., Dawson, J., & Krieger, K. (2018). Correlated parlay betting: An analysis of betting market profitability scenarios in college football. *The Journal of Prediction Markets*, *12*(2), 68-84.

Golec, J., & Tamarkin, M. (1991). The degree of inefficiency in the football betting market: Statistical tests. *Journal of Financial Economics*, *30*(2), 311-323.

Gray, P. K., & Gray, S. F. (1997). Testing market efficiency: Evidence from the NFL sports betting market. *The Journal of Finance*, *52*(4), 1725-1737.

Humphreys, B. R., Paul, R. J., & Weinbach, A. P. (2013). Bettor Biases and the Home-Underdog Bias in the NFL. *International Journal of Sport Finance*, *8*(4).

Marino, M. (2017). NFL Betting: Is the Market Efficient?.

Paul, R. J., & Weinbach, A. P. (2002). Market efficiency and a profitable betting rule: Evidence from totals on professional football. *Journal of Sports Economics*, *3*(3), 256-263.

Paul, R. J., Wachsman, Y., & Weinbach, A. P. (2011). The role of uncertainty of outcome and scoring in the determination of fan satisfaction in the NFL. *Journal of Sports Economics*, *12*(2), 213-221.

Pfitzner, C. B., Lang, S. D., & Rishel, T. D. (2009). The determinants of scoring in NFL games and beating the over/under line. *New York Economic Review*, *40*(1), 28-39.

Rolfe, B., & Robinson, D. R. (2023, August 24). *A brief history of betting on the NFL*. Pro Football Network. https://www.profootballnetwork.com/a-brief-history-of-betting-on-the-nfl/

Shank, C. A. (2019). NFL betting market efficiency, divisional rivals, and profitable strategies. *Studies in Economics and Finance*, *36*(3), 567-580.

Shank, C. A. (2018). Is the NFL betting market still inefficient?. *Journal of Economics and Finance*, *42*(4), 818-827.

Sports Business Journal. (2024, February 21). *U.S. sports betting industry sets record with nearly $11B in 2023 revenue*. Sports betting industry hits record $11B in 2023 revenue. https://www.sportsbusinessjournal.com/Articles/2024/02/21/us-sports-betting-industry-record-revenue-2023

Staff, C. S. (2023, October 20). *U.S. sports betting: Here's where all 50 states currently stand on Legalizing Online Sports Betting Sites*. CBSSports.com. https://www.cbssports.com/general/news/u-s-sports-betting-heres-where-all-50-states-currently-stand-on-legalizing-online-sports-betting-sites/#:~:text=Legal%20sports%20betting%20started%20expanding,its%20own%20sports%20betting%20laws.

Wever, S., & Aadland, D. (2012). Herd behaviour and underdogs in the NFL. *Applied Economics Letters*, *19*(1), 93-97.