2018

NFL Betting Market Efficiency: A Closer Look at the Final Day of Betting

Shankar Mani
Skidmore College, smani@skidmore.edu

NFL Betting Market Efficiency: A Closer Look at the Final Day of Betting

By

Shankar Mani

A Thesis Submitted to

Department of Economics

Skidmore College

In Partial Fulfillment of the Requirement for the B.A Degree

Thesis Advisor: Qi Ge

May 1, 2018

Abstract

This paper utilizes opening and closing betting lines on the day of play in the NFL to investigate if changes in the spread are a result of uninformed bettors. I formulate and empirically test the changes in the spread as they relate to home field advantage, favorites and hot hand betting. The results show that bettors tend to overvalue information and as a result, the actual scores shift less dramatically than the spreads on the final day of betting. A profitable betting strategy can be implemented betting against the shift in the spread. Additionally, it is more profitable to bet on home underdogs than away underdogs and more profitable to bet on away favorites than home favorites when betting against the shift in the spread.

Acknowledgement

I would like to express my deep gratitude to Jiebei Luo for her patient guidance in helping me develop and organize my data, as well as create my hot hand variables. Her willingness to help me under tight time constraints is very much appreciated.

Introduction

The sports betting market has expanded exponentially over the last ten years with an estimated market cap of close to 1 trillion dollars (Statista.com). Sports economists start to pay particular attention to this market in light of its gaining popularity, especially over the last ten years. The NFL betting market is the largest subsection of the gambling market overall, so this market receives the most attention from economists with regards to market efficiency. Economists analyze the stock market endlessly for market efficiency, as the general consensus is that the stock market is efficient over an extended time period (McGroarty and Urquhart 2016). However, stock prices are always a representation of the "collective judgement" of all stock traders. The NFL betting market, contrarily, provides a concrete value in that there is a final score of every game that reflects the ability of each team on the given day. As a result, the NFL betting market is a much more accessible market to analyze for efficiency and given its increase in popularity, is a necessary topic to explore in sports economics.

The NFL betting market has several components. The three most popular bets are to bet on the money line, the spread and the over/under. The money line is a bet on a team to win the game and categorized by either a positive or negative number. If the number is positive, the team is predicted to lose, and the number represents the money received on a 100 dollar bet (additional to the 100 dollars placed on the bet). So, if a team is +100, the bettor receives 100 dollars of profit. If the money line is negative, the team is expected to win. Therefore, the payout will be much less on a win. If a team is -200, the bettor will win 50 dollars on a 100 dollar bet (oddsshark.com). The over/under line is a bet on whether the total score (between both teams) will be above or below the given line. The payout for these bets are always one-to-one (the same as +100). The spread line is a predicted score difference (away-home) on the game. So if a bettor places a bet on the home team and the spread is -9, the home team must win by more than 9 to win the bet. In both over/under betting and spread betting, if the prediction is exactly the same as the result, there is a "push," which means that all bettors get their money back. To combat this result, bookmakers will often us 0.5 spread lines. Bookmakers also strategically make their lines to attract equal betting on each side. Bookmakers receive commission on each bet of about 5 percent (oddsshark.com). If the public bets heavily on one side and wins, the bookmaker is at risk. As a result, bookmakers are constantly changing the spreads in order to hedge their bets, and prevent losses.

The spread lines are both the most popular and most representative of the NFL market as a whole, which is why I choose to research this particular aspect of the NFL betting market in length (Ge 2018). Specifically, I analyze the changes in the spreads, by bookmakers, in order to hedge against possible loss. When bookmakers make their initial spread, the only thing that they analyze is the quality of the two teams and conditions the field will be played on. However, as the spread starts to change, it becomes more of a combination of the information about the game, and how the public view the game (oddsshark.com). Zuber et al (1985) serves as the baseline of the idea that the closing spread is the best predictor of the game outcome because information is constantly changing right up until game time, and earlier spreads may not incorporate these changes. Far fewer economists cite the bookmakers "hedging" as reason to believe that the closing spread is more of an indication of the public opinion, rather than quality of the teams (Humphreys 2011).

In this paper, I build off Paul and Weinbach (2011), who find that bettors in the last hour are uninformed and find a profitable betting strategy betting against the majority of last hour bettors. I analyze the changes in the spreads in order to draw conclusions on the best indicators of the score, as well as look for a profitable betting strategy in this market. I collect data on the opening and closing money lines on the day of NFL games from the 2006-2014 seasons, and use an OLS third order regression model to predict the point spreads on each game. I also collect information regarding temperatures and conditions to look for profitable betting strategies in that light. I run OLS regressions of the actual score difference on these different variables to test for efficiency. I compare the results from the open and close spreads.

I also include variables for home and away favorites to analyze how successfully home favorites can cover the spread because many analysts believe in "home field advantage." Dare and Dennis (2011) find that home teams are often underestimated; however, Aadland and Wever (2012) find that all favorites (home and away) are overestimated. The inclusion of favorite information will help me analyze the discrepancies in the literature regarding home field advantage and favorites.

My results show that bettors tend to overvalue information, as shifts in the spread by one result in a less than one point shift in the actual score. My results also show that the opening spread lines provide a more accurate prediction of game outcomes compared to the closing line spreads. These coefficients are statistically significant at the 1% level. As a result, it may be profitable to bet against shifts in the spread on the final day of betting. Furthermore, if the spread shifts in favor of the favorite, it is more profitable to bet on a home underdog than an away underdog. If the spread shifts in favor of the underdog, it is more profitable to bet on the away favorite than the home favorite.

My findings show that early bettors are more informed than the bettors that participate on the last day of betting. Despite a lack of information, bookmakers more accurately predict the game outcomes than the bettors do, even after all information is released leading up to game time. My model exposes inefficiencies in the NFL betting market, provides proof that bettors perform worse than the 50 percent that bookmakers aim to control, and provide a potential profitable betting strategy. My results suggest that in order to make this market more efficient, bookmakers must release the information they use to predict spreads because their predictions are better than average bettors.

Literature Review

Efficient Market Hypothesis

Fama (1970) develops an Efficient Market Hypothesis (EMH), which states that all stock prices accurately and fully represent all available information to the public. The EMH often references the random walk of prices, which results in individual investors not able to "beat the market" because no investor has access to information not available to everyone. Fama (1970) assumes that current prices are only impacted by current information, and not impacted by historical prices. Fama (1970) also assumes that all information is available for free. Fama (1970) considers three diverse types of efficiency: weak form, semi strong form, and strong form. Weak form refers to using historical prices or return sequences, semi strong form refers to newly released information (earnings, news reports, etc.), and strong form refers to "monopolistic information" (insider information not available to everyone). Fama (1970) note that though there is a severe lack of information regarding insider information (and rules to regulate that kind of trading), the first two forms give us strong insights to test the efficiency of a market. Fama (1970) concludes, as stated earlier, that stock price history and previous return sequences do not have any impact on the performance of the stock in the future. Fama (1970) also finds that stock prices only reflect semi-strong form, and therefore reflect the information currently released regarding the performance of the business.

There are several drawbacks to the EMH, though this literature should serve as the basis for any efficient market analysis. Howden (2009) critiques the EMH on a few accounts. Firstly, the literature assumes that all market information is interpreted in the same way; in reality, there have been several investors to beat the market, and people value different stocks in different ways. Secondly, he cites the transactions costs associated with trading to take away from market efficiency. As a result, a profitable investment strategy that "beats the market" must account for the commission costs investors' pay. Thirdly, the time it takes to process a transaction often times results in a discrepancy between the bid price and buy prices. The conclusion is that the EMH is too unrealistic to apply to the financial markets. Despite these limitations, the EMH can still serve as a basis for analyzing markets.

Unprofitable Betting Strategies

Applying the EMH to the NFL betting market, if the NFL betting market was efficient, then there would be no profitable gambling strategies. Furthermore, since gambling commission is around 5 percent, all betting strategies, in the long run, should result in a net loss of about 5 percent. This is different from the short run, where profitable betting strategies may appear successful. Zuber et al. (1985) analyze the efficiency of the NFL betting market in two ways. First, using traditional finance literature, they use an equation that relates the actual point spread of the NFL game to the Las Vegas Odds for the spread of the game. It is important to note that the Las Vegas Odds do not reflect the bookkeeper's initial prediction of the game's outcome, but rather uses all of the bets placed to reflect the optimal spread on the game. Just as stock prices reflect the continuous trading by investors and the prices they are willing to pay, the Las Vegas point spread represents the "collective judgement" of all bettors involved. Therefore, the Las Vegas odds for the spread are closing spreads (the spread when betting closes for a given game). If the market is efficient from the EMH, the coefficient under OLS regressions should be 1; in other words, the relation between the Vegas odds and the actual point spread should be a one to one ratio. Though Zuber et al. (1985) find a strong correlation, the study must also test to see if the Las Vegas odds have no impact on the actual results of the game (which intuitively makes sense). There is a stronger correlation of the latter, implying that additional tests of efficiency must be conducted.

Zuber et al. (1985) perform a second test that aims to find a profitable betting strategy, which if found, would mean there are some inefficiencies in the market. The model is an OLS regression of the realized point difference of a certain game on a vector of variables that includes yards rushing, yards passing, number of wins prior to the game, fumbles, interceptions, number of penalties, proportion of passing places to total plays and number of rookie players. The vector, similar to the realized point difference, is a difference of the two team's statistics. The results show less than a five percent loss, which shows inefficiencies in the market as stated above. The data analyzed is only over one season, so there are definitely significant restrictions on the results, but the model serves as the basis to test NFL betting market efficiency.

Amundson et al. (2006) uses the same ideas as the literature before; however, the variable of interest is the surface of the field the game is played on. The OLS regression is of the realized point difference on the predicted point spread, using a dummy variable to analyze if the field is grass or not. The other main variables are the overall record and overall point spread for the season. The overall record and overall point spread play a major role in the Las Vegas Odds spread decision, so the model can effectively analyze whether the field effects the teams playing. Teams that are used to turf may have trouble performing on grass. However, the results did not find a favorable betting strategy, which means that NFL players do not have trouble making a transition to the different types of fields. Amundson et al. (2006) considers a variety of different stadium characteristics, including the size of the stadium, but none of these result in a profitable betting strategy. As a result, it may help in a different context, but is not as relevant to my study.

Borghesi (2007) considers field conditions as well, as the study analyzes if teams can't perform in colder conditions than they are used to. The model uses the same regression model as Amundson et al. (2006), but implements an additional variable that analyzes the difference between the average temperatures in the team's hometown verse the average temperatures where the game is played. This is a very impactful study because many bettors believe that teams from the south will struggle in the north. The case is less believed when cold weather teams travel down south. The results show that historically teams from warm weather areas struggle in cold temperatures, as there seems to be a profitable betting strategy betting on teams from colder weather to do better. This is another variable that I am very interested in, and given the results, should include in my analysis, as cold weather is both something that affects players and gamblers.

The literature that does not find profitable betting strategies use variables that have already been included in the point spread set by Las Vegas, as offensive and defensive statistics are considered before setting point spreads. However, this literature does serve as the basis for all point spread analysis. Using my dataset, I will first run a regression of realized point difference on point spread to test for market efficiency (a coefficient of 1 indicates a perfectly efficient market). Then, I can include other variables, such as type of field and temperature, to test for inefficiencies.

Favorites and Home Field Advantage

The previous section discusses variables that should already be factored in to the point spreads, which is why the betting strategies prove not to be profitable. In this section, I will look at variables that are popular amongst bettors, but may not be as much of a part of the point spread, in order to find profitable betting strategies against the rest of the market. The two most common variables that bettors tend to over-consider are home field advantage (teams play better than their statistics may show at home) and favorites (the better team has a much larger spread to cover).

A common betting strategy, especially present in the NFL, relates to the home-underdog bias. The home-underdog bias is based on the idea that road favorites have trouble covering the spread. Bettors tend to overvalue the success of the road favorites. Dare and Dennis (2011) analyze this idea to see if there is a profitable betting strategy in betting for home underdogs. To analyze this phenomena, the study looks at the NFL seasons from 2005-2011. Dare and Dennis (2011) look at a mean forecast error. In other words, the study looks at the difference between the spread and the final result. Once the forecasting error is determined, the study runs a simple t test against 0, to see where the biases lie with respect to favorites and home teams. Their results find that bettors underestimate the scoring potential of home teams, both favorites and underdogs, and the results are statistically significant. As a result, there could be a profitable betting strategy from betting on home teams, as the spreads are generally lower than the results indicate.

Aadland and Wever (2012) look at a different betting strategy, which is the idea that favorites have trouble covering large spreads. As a result, Aadland and Wever (2012) use a probit model to analyze how the favorites do at covering the spread. The model regresses whether the home team wins the bet on home favorites, home underdogs and creates an interaction variable between the two variables and the closing line. To analyze the results, Aadland and Wever (2012) use the coefficients to predict the closing line that will have a winning probability of 53 percent (which accounts for the implied commission). The model predicts that underdogs are significantly underpriced, which means that the spread on a game involving a significant underdog is usually higher than it should be. Specifically, if the home team is the underdog, and the spread is above 6.5, a profitable betting strategy can be implemented betting on the underdog. If the visiting team is the underdog and the spread is above 10.5, there can be a profitable betting strategy betting on the underdog. Bettors tend to bet on favorites, and as a result, the spreads are a little higher than they should be. These conclusions prove that there may be inefficiencies in the market because bettors tend to both favor the home team and bet on the underdog. These two variables seem important to gamblers, and therefore should definitely be considered when I do my analysis.

Humphreys et al. (2013) critique the previous literature by using data from more seasons. First, the study replicates the model from before in order to see how their hypothesis holds up under a longer time interval. The results are similar to Dare and Dennis (2011); however, Humphreys et al. (2013) find an interesting difference. Humphreys et al. (2013) could reject the null hypothesis that the mean forecast error was zero for away favorites, but could not reject the null hypothesis for home underdogs. The study infers that the home-underdog favorite is only a small part of a much bigger story. Similar to Aadland and Wever (2012), Humphreys et al. (2013) believe that bettors just tend to favor teams that have been successful and therefor the favorite teams to cover the spreads. Paul and Weinbach (2002) find that, in general, due to psychological reasons, bettors tend to bet on the better teams. Using this theory, Humphreys et al. (2013) propose a different model to analyze if this hypothesis was true. Using data from sportsnights.com, they first look at betting percentages on favorites over the years in question to see if their hypothesis is true. As expected, the percentages prove that bettors tend to bet on favorites to cover the spread. It is important to note that these percentages look at total bets places and are not effected by total money placed on bets (Las Vegas sets their odds based on total money on each side of the bet, as mentioned before in order to hedge their losses). However, Humphreys et al. (2013) claim the high percentages are enough to hypothesize that there could be a bias. Instead of analyzing the mean forecast error, the study looks at determining a profitable betting strategy by looking at the percentage of bets that would win. Using the same data as Dare and Dennis (2011), Humphreys et al. (2013) find no profitable betting strategy in betting against favorites with this new approach. In conclusion, although bettors definitely prefer betting on favorites, the betting market correctly anticipates this outcome, and no profitable betting strategies can be found through this type of strategy.

Clearly, bettors over-bet the favorites in the NFL. Davis et al. (2013) looks at betting against the favorites in a different light. During week one of the NFL betting season, most of the information regarding the spreads rely on the prior season. As a result, the study looks at how teams fare during week one of the season that make the playoffs the season before. Specifically, the study looks at teams that made the playoffs competing against teams that did not make the playoffs to see if there is a bias towards these teams. In the simulation, which looks at season 2004-2012, teams cover the spread only 35.7 percent of the time. In other words, playoff teams do not cover the week one spread against non-playoff teams. Though this paper makes interesting points, the EMH is a long run theory, so looking at only one week of the season does not point out inefficiencies in the market.

The majority of the literature shows that bettors prefer to bet on favorites and that there could be a profitable betting strategy betting against the favorites. Specifically, the literature shows that away teams have trouble covering and/or beating the spread. As a result, I will include a dummy variable for the home team, favorite and home favorite and create interaction variables to see if I can find statistically significant differences. With this separation, I can analyze the "home underdog bias" and draw conclusions about how successfully home and away favorites cover the spread.

Hot Hand

The "hot hand" belief refers to the idea that teams that have recent successes will continue to succeed. The two competing theories are that bettors tend to overvalue recent successes and teams on winning streaks are more likely to succeed. The theory is first associated with the NBA because many believe that shooters go on streaks. Camerer (1989) and Brown and Sauer (1993) develop this model, claiming that bettors will bet heavily for teams on winning streaks and against teams on losing streaks. As a result, both studies find a profitable betting strategy, finding that the "mythical hot hand" skews the point spreads for teams on streaks. Paul and Weinbach (2005) further this research by finding a profitable betting strategy betting against teams on winning streaks.

The "hot hand" ideology can also be applied to the NFL betting market, as spreads could be skewed towards teams on winning streaks and against teams on winning streaks. Woodland and Woodland (2000) analyze this idea looking for the same profitable betting strategy found in the NBA. In their study, there are no profitable betting strategies betting against the streaks. The study concludes that there are no inefficiencies due to hot hand bettors. However, this idea assumes that bookkeepers change the spreads in order to attract equal bets on either side, and as a result the spreads on teams on winning streaks are falsely high. Humphreys (2011) shatters that idea with a study that proves bookkeepers actually take a stance on the games in order to exploit misinformed bettors and produce higher returns. He did this by comparing the initial and final spreads of games to the actual scores, and analyzed the returns the bookkeepers received. If bookkeepers did keep "balanced books," then in the long run their returns should be solely their commission. However, this study shows that bookkeepers in the NFL could calculate spreads independent of bettors. If the initial spread results in the public betting heavily on one side, then bookmakers miscalculate the initial spread, and must adjust the spread to fit public opinion. Hence, bookmakers must use public opinion to calculate spreads (Bookmakers use the spread to even the betting field). In the same light, bookmakers make excess profits from their initial line being misinterpreted by the public, rather than the way they adjust the spreads.

Paul et al (2014) take this idea a step further by critiquing the previous literature. Though there may not be a profitable betting strategy betting against teams on streaks, there still could be evidence of bettors betting on streaks. Instead of looking at data on the spreads, this study looks at betting percentages, so that even if the spread does not account for hot hand betting, evidence can be shown that it exists. The study defines winning streaks by how many wins consecutively against the spread a team has, leading up to the game. Losing streaks, similarly, are defined by consecutive losses against the spread leading up to the game. Using data from the 2005-2006 season and 2008-2009 season, the study finds that teams on winning streaks against the spread attract more bets than other teams. The same was concluded for losing streaks. Therefore, initial spreads often move in favor of "hot" teams and away from "cold" teams. The literature aforementioned consistently refer to winning streaks against the spread; however, the average bettor may only consider winning streaks outright when making their decisions. As a result, analysis should be conducted in order to determine if traditional winning streaks can either lead to a profitable betting strategy or expose "imbalanced books."

Though it seems there can be no profitable betting strategy betting against the hot hand principle, there is evidence that bettors over bet teams with recent success. Applying the ideas of the hot hand coupled with the "imbalanced bookkeepers" could bring to light meaningful conclusions about the discrepancies between the "collective judgment" mentioned earlier and the actual spread of these games.

Last Hour of Betting

The last hour of betting is a relatively new and under-researched topic in betting. The idea was first analyzed by Ottaviani and Sorenson (2008) in the horse betting market. The idea was simple: the most informed bettors place their bets in the last hour of betting because they wanted access to all available information before making their decision (as shown the EMH). Through their simple analysis, bets placed within the last hour of betting did exceptionally well supporting their hypothesis. The horse betting market differs significantly from the NFL betting market because the odds on horses are not released until after all bets are placed, so all bettors receive the same odds. In the NFL, all bettors may have a slightly different spread because the spreads are updated continuously. As a result, it makes sense that horse bettors wait until as close to post time as possible because the odds at that time will be most predictable of the odds they will receive. Despite this difference, NFL betting markets generally receive between 20 and 25 percent of their bets within the last hour of betting.

Miller and Rapach (2013) analyze the changes in the spread over the course of the week in the 1972 season because of their theory that the spread becomes more accurate as the week progresses. They separate their spreads into three categories; beginning of the week, Tuesday's opening line and closing line. Miller and Rapach use OLS regression analysis to compare the differences between the spreads and conclude that closing spreads are the best indicator of outcome, Tuesday spreads are the second best, and opening is the worst. Later, they use betting percentages and lag variables to put together a more in depth analysis, but my data limits me from performing such analysis. This data is also from only one season, and a season 45 years ago, so there are significant limitations to this study, but the simpler models comparing the different lines are relevant and necessary for my analysis.

With such a high percentage of bets occurring in the last hour, Paul and Weinback (2011) wanted to analyze whether bettors in the last hour were also more informed and therefore could produce higher returns than the other bettors. After conducting an analysis of the last hour of betting, looking for trends in the data, the study found that bettors tend to prefer the same betting strategies found in most of the literature (betting on favorites, home teams, etc.), which does mimic the betting of the general public. In other words, not only does the last hour of betting not show evidence of informed trading, but additionally the last hour of betting could be uninformed. As a result, the study looks for a profitable betting strategy by betting against the bettors in the last hour. To do this, the study looks at how the spread changes over the last hour of betting and sets up betting simulations against the move in the spread. For example, if the spread increases over the last hour, bettors in the last hour bet on the favorite, and therefore their simulation would bet on the underdog. Through these simulations, the study did find a profitable betting strategy, as 60 percent of the time, these bets are successful. The study has two notable conclusions relevant to my study. Firstly, in an efficient market, the time the bet is placed should have no impact on betting strategies because the spreads should accurately depict all available information on the game, and therefore still should show no profitable betting strategy. Secondly, if the bettors in the last hour are similar to the overall betting market, there may be a profitable betting strategy simply betting against the public, which would further back Humphreys (2011) claim that bookkeepers take a stance on games to strengthen their earnings potential.

Comparing the last hour of betting to the opening spreads could provide information regarding two things. First, there could be a profitable betting strategy simply betting against the public, which has been the common theme in every section. But secondly, I can analyze the changes in the spread to determine if there exists "imbalanced bookkeeping." Since the literature on last hour of betting is minimal, my paper hopes to reexamine the variables mentioned before in this new light.

My paper will extend traditional literature of field conditions and temperature as they relate to the spread by analyzing opening and closing spreads, and look for discrepancies, as well as determine which line more accurately represent the game outcome. After this, I will create additional variables for both home team and favorite in order to further analyze the relationships between the spread and the realized point score. Finally, I will analyze how changes in the spread effect the home team's chances of covering the spread. Much of the previous literature analyzes these variables with respect to closing spreads only, so I will be able to extend the analysis in to comparing opening and closing spreads, and hope to find a profitable betting strategy using public opinion.

Data

In order to analyze the efficiency of the NFL Betting market, I will look at games from the 2006-2014 NFL seasons and look for inefficiencies with respect to the type of field, as well as the temperature. My data is divided up by season. My first data set comes from armchairanalys.com. This data set includes dummy variables for grass fields, as well as dome fields. The data set also includes the temperature at the start of every game. The final scores for each game result are included as well (away-home). To distinguish my paper from previous literature, I compile an additional data set from Pinnaclesports.com. This data set includes both opening and closing money lines on the day of the game. In order to combine the data sets, I match up the games by season, away team and home team. As a result, I omit all observations that do not include opening money line, closing money line or the result of the game. I also omit all preseason games because teams play preseason games differently than regular season games to avoid injury. Since I have different sources of data, I omit all duplicate games, and only use the first entry for each game used. There are a total of 9,142 observations omitted. Despite the missing data points, I have 1,641 observations. My data is a time series, as the total aggregate games are separated by season. I will analyze the differences between opening and closing lines and spot any differences. Finally, I will look at the popular vote to see if how the spread changes can provide a profitable betting strategy. All of these models hope to spot inefficiencies in the market.

Zuber et al (1985) serve as the basis for NFL betting analysis, and much of the literature to follow agree that the closing lines are the most accurate representation of the game result. This can be attributed to last minute injuries and other information that may not become available until minutes before kickoff. Miller and Rabach (2013) perform an analysis of the 1972 NFL season that proves that as more information is analyzed and more bettors participate in the market, the line shifts toward a better representation of the game outcomes. However, Paul and Weinbach (2011) develop a new idea that as more and more bettors place their bets, the line becomes a less accurate representation of the game. Bookmakers study the NFL with much more scrutiny than the majority of bettors, so Paul and Weinback (2011) conclude that bettors actually disrupt the betting market. As a result, I hope to analyze the conflicting theories using opening and closing line data.

Ge (2018), Cortis (2015), and Card and Dahl (2011) prove in their literature the importance of the spread in the NFL, as a better indicator of the betting market than the money line. Cortis (2015) explains that the spread attracts equal bets on either side, while the money line does not. The spread is both easier to interpret, and more relevant for my research. As a result, I must convert my money line data to spread data. Ge (2018) outlines a model to convert money line data to the probability that the home team wins by using a model from Cortis (2015). The model is relatively simple, but does provide an easy way to determine the probability. Equation (1) shows the model below.

$$(1) \quad Money\ Line = \begin{Bmatrix} +100\left(\frac{1-p_i}{p_i}\right), p_i \leq 0.5 \\ -100\left(\frac{p_i}{1-p_i}\right), p_i > 0.5 \end{Bmatrix}$$

Using this equation, I convert all of my money lines to a probability that the home team will win. The next step, similar to Card and Dahl (2011) is to create a relationship between the realized score difference in the game (away team points minus home team points) and the probability that the home team will win. Since the relationship between the point spread and the realized score difference is one-to-one for most data sets of high volume, I will use the same relationship as point spread to probability in order to predict the pregame point spreads. Card and Dahl (2011) claim that there is a third order polynomial relation between the two variables; when the spread is within three points, a change in the spread will have a linear relationship to probability. As the spread climbs up above 10 points, changes in the point spread will have a smaller impact on the probability of the game because on team is already so heavily favored. As a result, I come up with a simple third order regression model that relates realized score difference with probability that the home team will win, as listed below in equation (2).

$$(2)\ PointDifference_i = \beta_0 + \beta_1 WinProb_i + \beta_2 WinProb_i^2 + \beta_3 WinProb_i^3 + E_i$$

After running this regression, I predict the spreads at the beginning of the day and at the close of betting for each game in my dataset. I have now calculated a predicted point spread that I will use for the remainder of my methodology and results. Using the predicted opening and closing spreads, I also create a dummy variable to describe if the spread moves in favor of the home team, as well as a variable that measures the difference between the opening and closing spread, which will be helpful in my analysis.

Table 1

Variable	**Observation**	**Mean**	**St Deviation**	**Mine**	**Max**
Opening Spread	1,641	-1.981718	5.83235	-17.48295	11.10715
Closing Spread	1,641	-1.981718	5.928049	-17.76627	12.2923
Difference	1,641	-1.984775	15.31817	-59	45
Temperature	1,641	49.39038	26.79876	0	99

Table 1 shows some summary statistics of my data. The mean opening and closing spread is -1.981718, which means that the home teams are favored more often than the away teams. The values are also the same, which shows that there may not be a large deviation between the opening and closing spreads. Additionally, the actual difference in scores is -1.984775, which is slightly different, but the close relation shows that the spread is a good indicator of the final score. Furthermore, the standard deviation is much higher for the actual score than the spreads because bookmakers try not to favor one teams heavily. A high standard deviation can also be an indicator of inefficiency;

however, further regression analysis needs to be done in order to draw more meaningful conclusions. The average temperature is 49.39038.

My data also includes information regarding the home team and favorites in the game, as much of the literature consider these variables. As seen in Figure 1 below, the betting market heavily favors the home team, as both bettors and bookmaker's alike value the home field advantage. Furthermore, when the home team is favored initially, the spread tends to increase in magnitude much more frequently than for the away favorites. As a result, it is necessary to consider the differences between home and away favorites. After taking a closer look at the data, I can get in to the methodology.

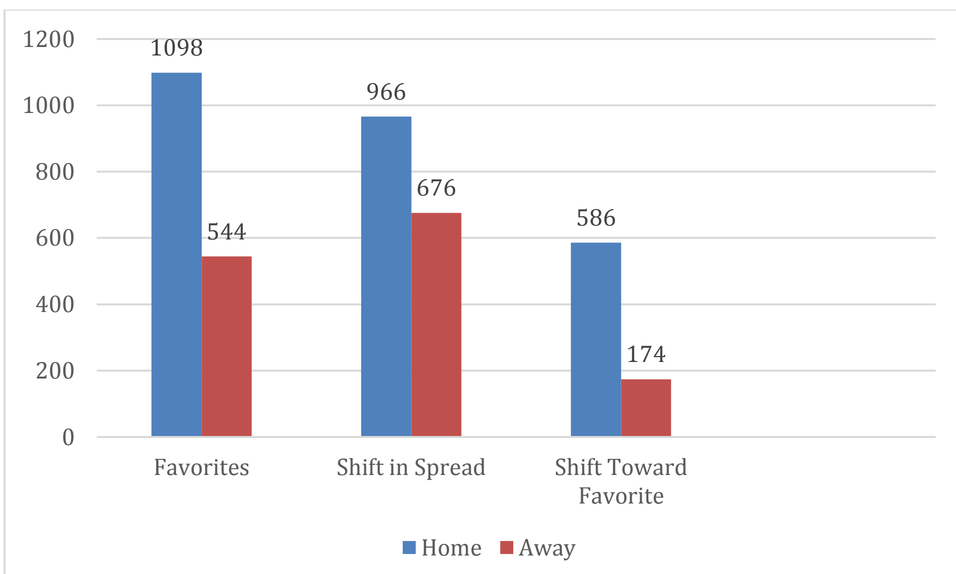

Figure 1

Lastly, my data includes hot hand information on both teams of each game in the dataset. In other words, each team (away and home) of each game has a number for the consecutive wins and losses over the last three games. I will use this data to analyze the effects of streaks on the outcomes of the game. Streaks do not go over multiple seasons, so during week one, none of the teams are on streaks.

Methodology

I choose not to use a probit model, similar to Aadland and Wever (2012). Though many bettors only care whether their bet wins or loses, the exact differences between the spread and the realized point difference are important for large datasets. My methodology comes from both Amundson et al. (2006) and Bhorghesi (2007), who take the simple OLS regression model of realized score difference on point spread and add a dummy variable for field condition, as well as a variable for temperature. The dummy variable will be whether the stadium is a dome or not. I hypothesize that home teams will have an advantage under dome conditions because they are used to it. Considering the difficulty that comes with playing in cold weather, I hypothesize that road teams will struggle in colder conditions. I will conduct the analysis on both initial lines and closing lines, as well. Sudden changes in the temperature may result in bettors making bets against the favorite to cover the spread. I also include a variable for fixed effects by season in my regression, so that the data is analyzed on a per season basis. Since there may be home team fixed effects as well, I will also include a variable for home team fixed effects. I have included the two new models below.

$$(3)\quad PointDifference_{it} = \beta_0 + \beta_1 OpenSpread_{it} + \beta_2 Dome_{it} + \beta_3 Temperature_{it} + \tau_t + E_{it}$$

$$(4)\quad PointDifference_{it} = \beta_0 + \beta_1 CloseSpread_{it} + \beta_2 Dome_{it} + \beta_3 Temperature_{it} + \tau_t + E_{it}$$

For each game i in season t, if β_1 is greater than 1, the favorites tend to cover the spread, as a one point increase/decrease in the spread results in more than a one point increase/decrease in the realized score difference. If β_1 is less than one, the favorites tend to not cover the spread as a one point increase/decrease in the spread result in a less than one point increase/decrease in the realized score difference. Additionally, I will run two more regressions including a variable for temperature squared, as Borghesi (2007) outlines is an additional test for how temperature effects the score. The two models are listed below.

$$(5)\quad PointDifference_{it} = \beta_0 + \beta_1 OpenSpread_{it} + \beta_2 Dome_{it} + \beta_3 Temperature_{it} + \beta_4 Temperature_{it}^2 + \tau_t + E_{it}$$

$$(6)\quad PointDifference_{it} = \beta_0 + \beta_1 CloseSpread_{it} + \beta_2 Dome_{it} + \beta_3 Temperature_{it} + \beta_4 Temperature_{it}^2 + \tau_t + E_{it}$$

Next, I will conduct my analysis on the "home field advantage," as many bettors and bookmaker's alike feel that teams play better at home. I have a dummy variable for when the home team is the favorite. I will include this new variable, and then create interaction variables between the home favorite variable and the variable of interest (spread). In this case, I am adding two variables to equations (3) and (4). Before running this regression, I must test for collinearity, as there is a high chance that there could be collinearity between the spread variable and the two new variables. In the case of the home favorite model, there does exist collinearity, so I drop the interaction variable. My new model regresses the realized point difference on temperature, dome nand the dummy variable for home favorite.

$$(7)\quad PointDifference_{it} = \beta_0 + \beta_1 OpenSpread_{it} + \beta_2 HomeFavoriteOpen_{it} + \beta_3 Dome_{it} + \beta_4 Temperature_{it} + \tau_t + E_{it}$$

$$(8)\quad PointDifference_{it} = \beta_0 + \beta_1 CloseSpread_{it} + \beta_2 HomeFavoriteClose_{it} + \beta_3 Dome_{it} + \beta_4 Temperature_{it} + \tau_t + E_{it}$$

I hypothesize, similar to Dare and Dennis (2011) and Aadland and Wever (2012), bettors tend to underestimate the ability of the home team to score. Therefore, I expect the coefficient b_1 in both equation (5) and (6) to be greater than 1, indicating that the home favorites tend to cover the spread more often than otherwise. I also expect the coefficient to be greater than b_1 from equations (3) and (4), indicating that home favorites cover the spread more often than overall favorites do.

Next, I will explore models that incorporate a dummy variable for when the home team is an underdog and use a similar model to before. With this dummy variable and interaction variable addition, there is no collinearity between the additional variables and spread, so I can create these models as shown below.

$$(9)\quad PointDifference_{it} = \beta_0 + \beta_1 AwayFavorite_{it} * OpenSpread_{it} + \beta_2 AwayFavorite_{it} + \beta_3 Dome_{it} + \beta_4 Temperature_{it} + \beta_5 OpenSpread_{it} + \tau_t + E_{it}$$

$$(10)\quad PointDifference_{it} = \beta_0 + \beta_1 AwayFavorite_{it} * CloseSpread_{it} + \beta_2 AwayFavorite_{it} + \beta_3 Dome_{it} + \beta_4 Temperature_{it} + \beta_5 CloseSpread_{it} + \tau_t + E_{it}$$

For equations (7) and (8), staying in line with the "home underdog bias," I hypothesize away favorites teams to have trouble covering the spread, so b_1 should be less than 1. I also expect b_1 to be lower than b_1 from equations (3) and (4).

I will also analyze the effects of recent success on the scores of the games by including a variable for streak. In order to do so, I will include a fixed effects variable for the streak. The streak variable will be 0 if the team is not on a winning streak, 1 if the team is on a one game winning streak, 2 if the team is on a two game winning streak, and 3 if the team is on a three game winning streak. I will then do the same for both the home and away team, for winning and losing streaks, and for opening and closing spread. The models will be the same as equations (3) and (4); however, there will be hot hand fixed effects. I hypothesize that bettors will overvalue teams on winning streaks and undervalue teams on losing streaks. As a result, I predict a profitable betting strategy betting against the "hot hand" and with the "cold hand."

Finally, I will analyze how the change in the point spread (from open to close) effects the realized point difference. In order to do this, I will run a standard two tail t test between the open and closed coefficients in each of the six equations to see if their difference is significant. The equations for the t statistic is listed below. The number of observations is the same for open and close, so $n_{OPEN} = n_{CLOSE} = n$

Null Hypothesis: $b_{1CLOSE} = b_{1OPEN}$

Alternative Hypothesis: $b_{1CLOSE} \neq b_{1OPEN}$

$$(11) \qquad t = \frac{b_{1CLOSE} - b_{1OPEN}}{\sqrt{\frac{(S^2_{b_{1CLOSE}} + S^2_{b_{1OPEN}})}{n}}}$$

Results

Table 2

VARIABLES	(3) Difference	(4) Difference
OpenSpread (b_{1OPEN})	0.881***	
	(0.0647)	
ClosingSpread (b_{1CLOSE})		0.886***
		(0.0631)
Dome	-1.576	-1.796
	(2.612)	(2.605)
Temperature	-0.00431	-0.00624
	(0.0202)	(0.0200)

Constant	0.00493	0.353
	(2.949)	(2.946)
Season FE	Yes	Yes
Team FE	Yes	Yes
Observations	1,641	1,641
R-squared	0.174	0.178

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2 shows the results from equation (3) and (4). The opening and closing spread coefficients are significant at the 1% level. A change in the opening spread by one point results in a 0.881 change in the actual score and a change in the closing spread by one point results in a 0.886 difference in the actual score. The closing spread is a slightly better indicator of actual score. The coefficients of the dome and temperature variables are both positive, but insignificant. The t statistic for the difference in opening and closing spread is approximately 0.056, so we fail to reject the null hypothesis that b_{1OPEN} from equation (3) equals b_{1CLOSE} from equation (4) at the 5% level.

Table 3

VARIABLES	(5) Difference	(6) Difference
OpenSpread (b_{1OPEN})	0.880***	
	(0.0647)	
ClosingSpread (b_{1CLOSE})		0.885***
		(0.0631)
Dome	-0.941	-1.341
	(2.690)	(2.674)
Temperature	0.0376	0.0237
	(0.0708)	(0.0700)
tempsq	-0.000455	-0.000325
	(0.000756)	(0.000749)
Constant	-0.678	-0.136

	(3.066)	(3.052)
Season FE	Yes	Yes
Team FE	Yes	Yes
Observations	1,641	1,641
R-squared	0.174	0.178

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3 shows the results from equations (5) and (6), which includes a temperature squared term. Once again, the coefficients of opening and closing spreads are statistically significant at the 1% level, while the dome and temperature coefficients are not statistically significant. A change in the opening spread by one point results in a change in the actual score of 0.88 and the change in the closing spread by one point results in a 0.885 change in the actual score. The t statistic for the difference between opening and closing spread is still 0.056, so we fail to reject the null hypothesis that b_{1OPEN} from equation (3) equals b_{1CLOSE} from equation (4).

Table 4

VARIABLES	(7) Difference	(8) Difference
OpenSpread (b_{1OPEN})	0.788***	
	(0.104)	
HomeFavoritesOpen	-1.417	
	(1.236)	
ClosingSpread (b_{1CLOSE})		0.731***
		(0.101)
HomeFavoritesClose		-2.436*
		(1.246)
Dome	-1.403	-1.674
	(2.607)	(2.600)
Temperature	-0.00377	-0.00600
	(0.0201)	(0.0199)
Constant	0.608	1.535

	(3.023)	(3.052)
Season FE	Yes	Yes
Team FE	Yes	Yes
Observations	1,641	1,641
R-squared	0.174	0.180

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 4 shows the results from equations (7) and (8). As with the previous models, the spread coefficients are statistically significant and the dome and temperature coefficients are not. In equation (7), the coefficient for the dummy variable is insignificant and negative in magnitude and in equation (8), the coefficient is significant at the 10% level and negative as well. When the home team is the favorite at the start of game day, the actual score difference is 1.5 less than otherwise and 2.43 less when the home team is the favorite when betting closes. The t statistic for the difference in opening and closing spread lines is well above 3.2, so we can reject the null hypothesis that b_{1OPEN} equals b_{1CLOSE} from equations (7) and (8) at the 1% level.

Table 5

VARIABLES	(9) Difference	(10) Difference
OpenSpread (b_{1OPEN})	0.811***	
	(0.114)	
AwayFavoritesOpen	1.758	
	(1.432)	
AFOpenSpread	-0.125	
	(0.264)	
ClosingSpread (b_{1CLOSE})		0.770***
		(0.110)
AwayfavoritesClose		3.032**
		(1.485)
AFCloseSpread		-0.210

		(0.264)
Dome	-1.440	-1.784
	(2.610)	(2.609)
Temperature	-0.00371	-0.00626
	(0.0201)	(0.0200)
Constant	-0.668	-0.621
	(2.989)	(2.966)
Season FE	Yes	Yes
Team FE	Yes	Yes
Observations	1,641	1,641
R-squared	0.174	0.180

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5 shows the results from equations (9) and (10). The coefficients of the spreads are statistically significant at the 1% level. The dome, temperature, and the interaction variables are not significant. The away favorite dummy variable is statistically significant at the 5% level in the closing spread model, but insignificant in the opening spread model. When the away team is the favorite at the start of game day, the spread is 1.758 higher than otherwise and when the away team is the favorite at the end of betting, the spread is 3.032 higher than otherwise. The t statistic for the difference in opening and closing spread lines is well above 3.2, so we can reject the null hypothesis that b_{1OPEN} equals b_{1CLOSE} from equations (7) and (8) at the 1% level.

Table 6

VARIABLES	(12) Difference	(13) Difference	(14) Difference	(15) Difference
OpenSpread	0.882***	0.880***	0.875***	0.888***
	(0.0670)	(0.0669)	(0.0676)	(0.0672)
Dome	0.212	0.210	0.261	0.252
	(1.118)	(1.114)	(1.117)	(1.117)
Temperature	0.0193	0.0190	0.0195	0.0195
	(0.0172)	(0.0172)	(0.0172)	(0.0172)
1.HomeLoseStreak	0.0226			

	(0.860)			
2.HomeLoseStreak	-0.737			
	(1.132)			
3.HomeLoseStreak	0.638			
	(1.095)			
1.HomeWinStreak		-0.102		
		(0.858)		
2.HomeWinStreak		0.591		
		(1.153)		
3.HomeWinStreak		0.394		
		(1.154)		
1.AwayLoseStreak			0.872	
			(0.822)	
2.AwayLoseStreak			-0.0506	
			(1.148)	
3.AwayLoseStreak			0.727	
			(1.255)	
1.AwayWinStreak				-0.696
				(0.883)
2.AwayWinStreak				-0.537
				(1.118)
3.AwayWinStreak				0.169
				(1.120)
Constant	0.227	0.198	-0.0206	0.394
	(1.493)	(1.483)	(1.484)	(1.487)
Season FE	Yes	Yes	Yes	Yes
Team FE	Yes	Yes	Yes	Yes
Observations	1,641	1,641	1,641	1,641
R-squared	0.149	0.149	0.149	0.149

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 6 shows the results from the hot hand model for the opening spread. In this model, the coefficient is much closer to 1 than the previous models, but the streak data does not provide any statistically significant coefficients. I expect teams on winning streaks to do better and teams on losing streaks to do worse, so the signs for home winning streak and away losing streak to be negative, and the signs for away winning streak and home losing streak to be positive. However, the results show home teams on two or three game winning streaks to favor the away team and away teams on one or three game winning streaks to favor the away team as well. When the home team is on a two game losing streak or the away team is on a one or two game winning streak, the results show the home team to have an advantage. Given the insignificance of the results coupled with the randomness of the signs of the coefficients, I conclude that the hot hand does not provide an opportunity for a profitable betting strategy.

Table 7

VARIABLES	(16) Difference	(17) Difference	(18) Difference	(19) Difference
ClosingSpread	0.888***	0.886***	0.880***	0.892***
	(0.0658)	(0.0657)	(0.0663)	(0.0659)
Dome	0.139	0.140	0.196	0.187
	(1.119)	(1.116)	(1.119)	(1.118)
Temperature	0.0165	0.0161	0.0168	0.0168
	(0.0172)	(0.0172)	(0.0172)	(0.0172)
1.HomeLoseStreak	0.0368			
	(0.862)			
2.HomeLoseStreak	-0.740			
	(1.125)			
3.HomeLoseStreak	0.603			
	(1.092)			
1.HomeWinStreak		-0.126		
		(0.855)		
2.HomeWinStreak		0.634		
		(1.150)		
3.HomeWinStreak		0.505		
		(1.148)		
1.AwayLoseStreak			0.957	
			(0.817)	
2.AwayLoseStreak			-0.0240	
			(1.142)	
3.AwayLoseStreak			0.722	
			(1.250)	
1.AwayWinStreak				-0.795
				(0.878)
2.AwayWinStreak				-0.436

				(1.120)
3.AwayWinStreak				0.174
				(1.123)
Constant	0.408	0.368	0.132	0.571
	(1.498)	(1.485)	(1.488)	(1.489)
Season FE	Yes	Yes	Yes	Yes
Team FE	Yes	Yes	Yes	Yes
Observations	1,641	1,641	1,641	1,641
R-squared	0.154	0.153	0.154	0.154

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7 shows the results from the hot hand closing spreads model. The results are very similar to the opening spread model, so I draw the same conclusion that the "hot hand" does not provide opportunity for profitable betting strategy. The t statistic between the opening and closing spreads is less than 1, so we fail to reject the null hypothesis that b_{1OPEN} equals b_{1CLOSE}.

Discussion

Equations (3) and (4) are standard regressions looking at the relationship between the actual score difference on open spread/close spread, dome and temperature, and are displayed in Table 2. The coefficients of interest are the values of b_{1CLOSE} andb_{1OPEN}. Both b_{1CLOSE} andb_{1OPEN}, at the 1% level, do not equal the value of 1 that the EMH predicts. As a result, there are clearly inefficiencies in the market. Additionally, the t statistic between b_{1CLOSE} andb_{1OPEN} shows that I fail to reject the null hypothesis that $b_{1CLOSE} = b_{1OPEN}$. Both values are less than one, which means that changes in the spread are overvalued, and the actual score difference changes less than the spread does. One possible explanation from general finance theory is that markets often overreact to information, as Howden (2009) outlines. When the spread changes, it is a result of new information, which causes bettors to overreact, and shift the spread much more than it should shift. As a result, the actual final score is less sensitive to additional information than the spreads.

The $Dome_i$ coefficient,b_2, is positive, which would indicate that a dome actually favors the away team (recall that spreads are calculated by away score-home score), while holding all other variables constant. However, the values are insignificant. The $Temperature_{it}$ coefficients,b_3, are positive as well, which means that increases in temperature favor the away team and decreases in temperature favor the home team, while holding all other variables constant. These coefficients are statistically significant, and agree with the results from Borghesi (2007), as teams from warm weather struggle in colder conditions. The small magnitude of the coefficient signifies that only dramatic changes in temperature can impact the score of the game. As Borghesi (2007) describes, a change from 60 degrees to 10 degrees gives the home team an unbelievable advantage, and my model predicts a decrease in the $PointDifference_{it}$ by 1 (away-home).

Equations (5) and (6) include a temperature squared variable in order to further analyze temperature as an indicator of the final score. The results are insignificant and small in magnitude as with the first two models, so no significant conclusions can be drawn for these results. As a result, temperature is not a great indicator of final scores.

Equations (7) and (8) incorporate the $HomeFavorite_{it}$ variable to analyze how having a home favorite's effect the spread. The results are shown in Table 4. The $Dome_i$ and $Temperature_{it}$ coefficients have similar results to before, so the same conclusions can be drawn from this model. The b_{1CLOSE} and b_{1OPEN} values are significant and much less than one, which implies, similar to previous models, that bettors tend to overreact to newly released information, and as a result changes in the spread are more dramatic than the change in the actual score. In this model, I can reject the null hypothesis that the b_{1CLOSE} and b_{1OPEN} are equal. Furthermore, the opening spread lines are better indicators of the actual score than the closing spread line, which contradicts Zuber et al. (1985). One possible explanation for this is the idea that earlier in the day, the bettors are bookmakers and more informed bettors. As the day progresses, more casual bettors that do not spend as much time tracking spreads enter the market, which causes the spread to shift away from the actual scores. The dummy variable has a negative coefficient, which is in line with my prediction because when the home team is the favorite the score should decrease and when the away team is the favorite, the spread should increase. The closing spread model has a higher coefficient, which is significant at the 10% level, which means that the favorite has a better chance to cover the spread if the team is favorited at the close of betting than at the start of betting. The fact that the coefficient is statistically significant and less than one does give merit to the idea that favorites are often overestimated, as Aadland and Wever (2012) discuss. Since $b_{1CLOSE}$is less than b_{1OPEN}, bettors on the day of play significantly overestimate the home team's ability to score. However, comparing these results to the results in Table 2, it is clear that bettors overestimate a team's ability to score regardless of who the favorite/home team is.

Equations (9) and (10) incorporate a variable for when the home team is the underdog to analyze how home underdogs effect the spread. The results are shown in Table 5. The $Dome_{it}$ and $Temperature_{it}$ coefficients have similar results to before, so the same conclusions can be drawn from this model. However, a number of interesting conclusions can be drawn from the new additions to the model. The model predicts that increases in the spread make it tougher to cover the spread and decreases in the spread make it easier to cover the spread, which intuitively makes sense. The open spread model suggests that a one point increase in the $OpenSpread_{it}$ results in an additional 0.12 decrease in the $PointDifference_{it}$when the away team is favored, as opposed to when the home team is favored. In other words, away favorites are less likely to cover the spread when the spread increases and more likely to cover the spread when the spread decreases (since $b_{1OPEN} = 0.81$ and $b_{5OPEN} = -0.12$). This coefficient is insignificant. The close spread model predicts similar outcomes. In the closing spread model, an increase in $CloseSpread_{it}$ by one results in an additional decrease of 0.21 in $PointDifference_{it}$ as compared to home favorites. This coefficient is statistically insignificant as well.

The test statistic, when comparing b_{5OPEN} to b_{5CLOSE}, proves at the 1% level that we can reject the null hypothesis, as with the other previous model. Since the two values under analysis are statistically significant as well, there is a discrepancy between the $OpenSpread_{it}$ and $CloseSpread_{it}$ that is statistically significant. A shift in the spread from open to close, as per EMH, implies that additional information is released on the day of the game, causing one team to have an advantage over the previous spread.[1] Generally, on the last day of betting, uninformed bettors must overreact to this newly released information, which shifts the spread much more than it should. Using Table 4 to analyze away favorites and Table 5 to analyze home favorites, I can use statistically significant information, as the dummy variable for $HomeFavorite_{it}$ and

[1] New advantages/disadvantages can range anywhere from injury updates and weather changes to bookmaker's trying to hedge their bets to do the increases frequency of betting on one side.

$AwayFavorite_{it}$ is equal to 0. Specifically, when the away team is the favorite, the spread shifts much more than when the home team is the favorite. In other words, if the spread shifts toward the favorite, away favorites have a tougher time covering a larger spread than home favorites. If the spread shifts toward the underdog, then home favorites have a tougher time covering the spread than the away favorites do. One possible theory for this phenomenon comes from a combination of Dare and Dennis (2011) and Aadland and Wever (2012). When the bettors shift the spread towards the favorite, Dare and Dennis (2011), Aadland and Wever (2012) and my results show that away teams struggle to cover the spread on the road and home favorites have a much better chance to cover high spreads than away favorites. When the bettors shift the spread towards the underdog due to newly released information, my results show the spread will shift more toward the home underdog than the away underdog, so away favorites will cover the spread more often than home favorites. In other words, home underdogs are overestimated compared to away underdogs, whereas, away favorites are overestimated compared to home favorites.

It is important to note that, overall, bettors still overreact to information, as the coefficients of opening and closing spreads are less than one. The results match the results from Paul and Weinbach (2011), who also find a profitable betting strategy betting against the public. Their results look at the last hour, instead of the last day. As a result, a profitable betting strategy can be implemented by betting against the general public on the last day of betting because bettors overreact to newly released information and bettors that participate on the last day are less informed than previous bettors. The results from Table 4 and 5 show that if the spread shifts toward the favorite, betting on the home underdog is more profitable than betting on the away underdog; if the spread shifts toward the underdog, then it is more profitable to bet on the away favorite than the home favorite.

Another theory that could prove this anomaly, is the theory of imbalanced books. Humphreys (2011) develops a model that proves that bookmakers make additional income to the commission charges, and infers that bookkeepers may not fully adjust the spreads to convey the "collective judgement," that it should. In the case of the spread shifting, bookkeepers could over adjust the spread based on their own opinions more than the opinion of the public, with the hope of capitalizing on public misconceptions. No matter which theory is accurate, the model does point to inefficiencies in the NFL betting market.

Table 6 and Table 7 show the results from the hot hand models, which analyze how teams on streaks do in comparison to teams not on a streak. The coefficients of opening and closing spread are the same and much closer to one, which shows that the betting market is inefficient as it relates to teams on streaks. Additionally, the coefficients of the streak variables are not significant, small in magnitude, and vary drastically in sign, which indicates that a team's past performance does not impact future performance, so there are no profitable betting strategies related to the hot hand.

Limitations and Extensions

My study uses money line data to predict the spreads of the games, as opposed to the actual spreads of the games. As a result, there could be some bias, as a model to predict spreads will accurately represent an efficient market much more than the actual spreads might. The models use the EMH to help predict spreads, so there could be additional inefficiencies that I could not point out. Additionally, I did omit 9,142 observations from my data set, and though it was random and due to an access issue, this can significantly skew my results as well. I also do not have any information regarding betting volume, so I was only able to analyze bettor popularity by the changes in the spread over the last day. A closer look at popularity could also bring light to whether bookmakers are using imbalanced books to maximize their profits. There are many injury updates and additional factors that result in a change in the spread. My data also spans the last day of betting; however, literature proves that 25 percent of the bets in the NFL occur in the last hour of betting, but I am not able to analyze the last hour of betting. I also use an OLS regression model, which has many benefits; however, many bettors are much more interested in whether the bet wins or loses. As a result, some bettors favor the probit model, which does not account for the excess points a team covers the spread by. Lastly, there is a potential for type I error, which occurs if I reject the null hypothesis when the null hypothesis is true. The likelihood is very rare, considering my t statistics were above 50, but it does need to be mentioned.

Despite my papers limitations, the paper provides important insights that can lead to future research. The new idea that shows opening lines as a better indicator of the score than closing lines can be used in many studies in the future. Studies can run betting simulations betting on the shift in the spread to overestimate the shift in the outcome. Since the difference in the coefficients of opening and closing spreads have a very low magnitude, additional research can provide information on the magnitude of a spread shift that provides a profitable betting strategy, as well as the magnitude of the spread prior to the shift. Both of these factors will be relevant in putting together a profitable betting strategy.

There is a plethora of literature analyzing the relationship between home field advantages and favorites. Some literature proves that bettors underestimate the ability of the home team to score. Some literature proves that bettors bet on favorites more often than not. Some literature proves that severe underdogs receive more praise than they deserve (and hence the spread is lower than it should be). Either way all the literature is inconclusive. Since my model is OLS, applying the probit or mean difference error methods I outline in my literature review coupled with this new idea of comparing opening lines to closing lines could provide additional information as well. My paper should pave the way for additional studies to really analyze the changes in the spread throughout the betting season, and potentially expose inefficiencies in the NFL betting market, as my paper has. Combining the home field advantage and favorite betting strategies with the last day (or last hour) of betting might show statistically significant simulations.

The hot hand biases may be more noticeable earlier in the week, as bettors could tend to value past performance when it is much more recent. Further research using spreads from earlier in the week could provide more meaningful conclusions to the idea of the teams on streaks. Since this paper suggests that many bettors are uninformed, bettors may value how the team has been performing overall, as opposed to against the spread. So additional research could also analyze teams on streaks could provide better results than streaks against the spread.

Conclusion

This paper utilizes money line data from the opening and closing of betting on game day to analyze discrepancies between opening and closing line data. I convert the money lines to spread lines using literature. Most literature and theory provide evidence that the closing lines are a better indicator of the game outcome, due to late releases of injury and other information. This paper finds that bettors tend to overreact to newly released information, as spreads move more than the actual scores will move in the NFL betting market. This paper also finds significant discrepancies between the opening and closing spreads. The results indicate that the opening spreads have a much greater correlation to the actual scores than the closing spreads, which contradicts much of the previous literature. As a result, the market is inefficient. Bookmakers and other gamblers that place their bets early in the week will be more informed because they track the spreads early and often. On the last day of betting, more uninformed casual bettors enter the market, who are less successful in predicting final scores. The results suggest that the markets overreact to changes in information. In the stock market, investors tend to overreact to shocks, which is why dramatic changes in the market correct itself over time. Since the NFL betting market does not have time to correct itself, as eventually the game takes place, maybe the market does not have time to fully correct the overreaction. If bookmakers and the media distribute the information more efficiently, the market itself will run more efficiently. The inefficiency and lack of bettor awareness provide insight in to why sports gambling remains illegal in many places. The unpredictability in sports, often due to exogenous factors, does not mirror other markets and can explain the inability of bettors to succeed. Until then, there can be a profitable betting strategy by betting against the shift in the spread, as the shift in the score is predicted to be less than the shift in the spread. Furthermore, if the shift moves toward a favorite, it is more profitable to bet on a home underdog than an away underdog. If the shift moves toward the underdog, it is more profitable to bet on the away favorite than the home favorite.

Work Cited

1. Aadland, D., & Wever, S. (2012). Herd behaviour and underdogs in the NFL. *Applied Economics Letters, 19*(1-3), 93-97.

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

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

4. Brown, W. O., & Sauer, R. D. (1993). Does the basketball market believe in the hot hand? American Economic Review, 83(5), 1377-1386.

5. Camerer, C. (1989). Does the Basketball Market Believe in the Hot Hand? American Economic Review, 79(5), 1257-1261.

6. Card, D., Dahl, G.B., 2011. Family violence and football: the effect of unexpected emotional cues on violent behavior. Q. J. Econ. 126 (1), 103–143.

7. Cortis, D., 2015. Expected values and variances in bookmaker payouts: a theoretical approach towards setting limits on odds. J. Predict. Mark. 9 (1), 1–14

8. Dare, W. H., & Dennis, S. A. (2011). A test for bias of inherent characteristics in betting markets. *Journal of Sports Economics, 12*(6), 660-665.

9. Davis, J, DiFilippo, M., Fodor, A., & Krieger, K. (2013). Inefficient pricing from holdover bias in NFL point spread markets. *Applied Financial Economics, 23*(16-18), 1407-1418.

10. Davis, J., DiFilippo, M., Fodor, A. & Krieger, K. (2014). Early season NFL over/under bias. *Journal of Sports Economics, 15*(2), 201-211.

11. Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The Journal of Finance, 25*(2), 383-417.

12. Ge, Q., 2018. Sports sentiment and tipping behavior. *Journal of Economic Behavior & Organization*, 13(1), 1-15.

13. Lopez, M.J., Matthews, G.J., 2015. Building an NCAA men's basketball predictive model and quantifying its success. J. Quant. Anal. Sports 11 (1), 5–12

14. Howden, D. (2009). Fama's efficient market hypothesis and mises's evenly rotating economy: Comparative constructs. *Quarterly Journal of Austrian Economics, 12*(2), 3-12.

15. Humphreys, B. R. (2011). The financial consequences of unbalanced betting on NFL games. *International Journal of Sport Finance, 6*(1), 60-71.

16. 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), 294-311.

17. Humphreys, B., Paul, R. J., & Weinbach, A. P. (2014). Bettor belief in the 'hot hand': Evidence from detailed betting data on the NFL. *Journal of Sports Economics, 15*(6), 636-649.

18. Humphreys, B. R., & Perez, L. (2012). Who bets on sports? Characteristics of sports bettors and the consequences of expanding sports betting opportunities. *Estudios De Economia Aplicada, 30*(2), 579-597.

19. McGroarty, F., & Urquhart, A. (2016). Are stock markets really efficient? Evidence of the adaptive market hypothesis. *International Review of Financial Analysis, 47*, 39-49.

20. Miller, T. Rapach, D. (2013). An intra-week efficiency analysis of bookie-quoted NFL betting lines in NYC. *Journal of Empirical Finance*. 24(7), 10-23.

21. Paul, R. J., & Weinbach, A. P. (2011). An analysis of the last hour of betting in the NFL. *International Journal of Sport Finance, 6*(4), 307-316.

22. Sauer, R. D. (1988). Hold your bets: Another look at the efficiency of the gambling market for national football league games: Comment. *Journal of Political Economy, 96*(1), 206-213.