Betting Markets and Market Efficiency: Evidence from College Football

Michael Sinkey* and Trevon D. Logan†

December 2009

Abstract

Existing studies of market efficiency in sports betting markets concentrate on professional sports. We test for betting market efficiency in an amateur sport, American college football, using data from over 11,000 games from 1985 to 2003. Sizable betting markets exist for amateur sports, but little is known about their efficiency. We find that the market is inefficient. In particular, we find that favorites are systematically overpriced. We show that the magnitude of this bias is large enough to generate both economic and statistical inefficiency in this betting market. Furthermore, we provide suggestive evidence for the cause of this inefficiency— betting houses deliberately inflate the betting lines for favorites in order to counteract bettor's "hot hand" beliefs. Tempering the "hot hand" bias results in consistently profitable betting strategies.

---

*Corresponding author. Department of Economics, Ohio State University, 410 Arps Hall, 1945 N. High Street, Columbus, OH 43210. e-mail: sinkey.4@osu.edu

† Ohio State University and NBER, e-mail: logan.155@osu.edu

Andrew Weinbach graciously shared data sources. Nathan Fritz-Joseph and Justin Sawyer provided excellent research assistance. We thank participants in the OSU Micro Workshop, Ben Anderson, Rodney Andrews, Vicki Bogan, Lisa D. Cook, John Gray, P.J. Healy, Pok-Sang Lam, Pete McGee, Jim Peck, Brian Roe, and Bruce Weinberg for helpful comments and conversations. The usual disclaimer applies.

1. Introduction

In prediction markets individuals price the likelihood of future events. If the prediction market price accurately represents all of the information available in the market, there is no room for arbitrage. This "no arbitrage" condition has a natural analog to the efficient markets hypothesis (EMH), where traders predict the present value of an asset, which features prominently in the economics and finance literatures. Sports betting markets are particularly unique prediction markets since the predicted event will be realized with certainty. For this reason, researchers have focused on sports betting markets as fertile ground for tests of market efficiency (Sauer et al. 1988, Zuber et al. 1985, Camerer 1989, Woodland and Woodland 1994, Gil and Levitt 2007).

The existing literature has focused on professional sports in testing for market efficiency (Gray and Gray 1997, Sauer et al. 1988, Gandar et. al 1988, Zuber et al. 1985 , Woodland and Woodland 1994, Brown and Sauer 1993, Scheibehenne and Bröder, 2007). Few studies have focused on amateur sports. Despite their amateur status, there are a number with large betting markets. For example, it is estimated that as much as $12 billion is bet on the NCAA Final Four Basketball tournament each spring (Matuszewki 2009). Although the basic structure of the markets is similar to that in professional sports, the information environment can vary considerably. For example, college teams have different characteristics from professional teams. When moving from professional sports to their college counterparts, the number of teams increases by four fold, there is greater variation in team quality, player turnover is inherently greater, and, unlike professional sports, injuries and suspensions are not always released publicly. There may also be fewer sources of public information in college sports; while every professional sports team receives substantial newspaper and television coverage, relatively few college programs are subject to such intense scrutiny. Indeed, many major college sports powers are outside of major media markets, which can act to limit the availability of information about teams.[1] More important, strategies, such as betting on teams that are playing at home, may be found to be efficient in betting on a professional sport, yet inefficient in an amateur one.

Despite the size and economically interesting features of amateur betting markets, little is known about their economic properties, such as efficiency. In this paper we focus on the market efficiency of the betting market for American college football. The few existing studies of the market have suggested that college football betting is efficient (Dare and McDonald 1996, Golec and Tamarkin 1991, and Paul et. al 2003). One drawback of the existing studies is that they used limited data and employ methodological approaches which are indirect tests of market efficiency, as we describe below. We use new data on over 11,000 college football games from 1985 to 2003 and their characteristics to test for market efficiency in the college football betting market. Contrary to existing literature, we find robust evidence that the market is inefficient when using recently developed, theoretically grounded empirical methods. Specifically, our direct test for market efficiency estimates whether or not bettors can consistently make profit by betting clearly identifiable simple strategies. In particular, we find that favorites are consistently overpriced. We estimate that, for example, if one bet $1000 on underdogs in prominent games, one would receive $1,116.99 in expectation, before accounting for the transaction cost to the betting house.

To understand the source of this inefficiency we analyze how this betting market functions. The key piece of information used in this market is the point spread, also known as the betting line, which is the predicted margin of victory for a given game. Bettors place bets that a team will "beat the spread" (exceed the predicted margin of victory) or not. Existing analysis of market function

[1] For example, every single NFL (National Football League) city has all four major TV affiliates and a daily newspaper with their own dedicated sports reporters, and many cities have multiple newspapers.

postulates betting lines as being independent from game to game, previous performance against the line should not be predictive of future performance if the market is efficient (Fama 1991, among others). We find, however, that betting lines are not independent from game to game. In particular, current betting lines have memory; they are functions of previous betting market results. Because of this serial correlation, we test to see whether or not teams that exceed the betting line are likely to do so in the following week. We find that teams who exceed the betting line in one week are no more likely to do so in the following week because betting lines are *systematically greater* for teams who beat the betting line the previous week.

This bias suggests that betting houses are particularly sensitive to potential "hot hand" bias among bettors. As a rule, betting houses are particularly sensitive to *any* bias among bettors— their profit motive is to have an equal amount of money on either side of the bet. If bettors believe in the hot hand, the profit motive of betting houses will cause them to overstate the betting lines of "hot" teams in order to avoid risk. We support our conclusion with both qualitative and statistical evidence. Qualitatively, we use narrative evidence to show that this sensitivity drives the overpricing of favorites that we observe in the market. Using a variety of sources, we document how previous performance against the betting line is commonly used to predict current performance. Statistically, we show how the mispricing of games varies by whether or not they contain "hot" teams. One consequence of this adjustment is that it makes other conditional strategies, such as betting on favorites that beat the betting line in the week before, consistently profitable. Both of these sources of evidence are consistent with our finding that counteracting the "hot hand" bias of bettors creates profitable strategies that cause the college football betting market to be inefficient.

2. Prediction Markets and College Football Betting

Prediction markets have recently gained prominence due to their ability to forecast future events with less variance and more accuracy than other techniques, such as polling experts. Research on prediction markets has increased along with their use. Theoretical papers in this literature include Ottaviani and Sorenson (2007, 2008), and Wolfers and Zitzewitz (2004, 2007), while empirical papers cover a wide variety of issues, including macroeconomic events and political elections (Rhode and Strumpf (2005), Snowberg et. al (2007)). Furthermore, many firms now use prediction markets to estimate the likelihood of future events[2]. Among prediction markets, gambling on sports is well-developed, and the results of a particular event are public and known with certainty.[3]

Sports betting markets are of particular interest for a number of reasons. First, sports betting markets hinge on a relatively simple metric, the betting line, which is the estimate of how large the margin of victory will be in a given sporting event. Bettors simply place wagers on whether or not a given team will win (lose) by more (less) than the predicted margin of victory (the betting line). In order to win, a bettor need only be on the correct side of the betting line. That is, bettors who bet on the favorite win if the team predicted to win wins by a margin greater than the predicted margin of victory, and bettors who bet on the underdog win if the team predicted to lose loses by less than the predicted margin of victory. Second, the sports betting market is quite large and many bettors are repeat participants— if a consistently profitable strategy were available bettors are likely to exploit it. A priori, there should be limited room for arbitrage in sports betting markets. Third,

[2] For example, Google uses prediction markets to predict events relevant to company profitability, including the number of users for Gmail, the quality rating of Google Talk, whether or not Apple will release an Intel-based Mac, etc. (Cowgill, Wolfers, Zitzewitz 2009).

[3] Previous Intrade contracts have been controversial. In 2006, the bet of whether or not North Korea would successfully fire a missile outside of its own airspace was improperly worded, leading to confusion and anger from bettors in the betting market.

outcomes are observed, and we can directly assess whether or not a given bet or betting strategy was profitable.

Among sports betting markets, amateur betting markets share characteristics with professional betting markets, but also contain economically interesting differences. College football games have different properties from professional football games. College coaches are not required to provide weekly injury reports; whether or not key players are playing may be obscured from both the opponent and the betting market. College football teams also have new players enter and exit more frequently than the NFL. It may be more difficult to discern the quality of a team with younger (freshman or sophomore) players because of few observations. Even if the status of key players is known, their substitutes are more likely to be unknown quantities. Given these informational differences, the question of whether or not the college football betting market is efficient is not clear. To a certain extent, efficiency of the betting market in college football can be seen as general test of information environment differences on market efficiency, with college football being a coarser information environment.

Betting houses facilitate the betting process in both professional and amateur sports in the same way, by setting the betting line that bettors wager on. They derive their profit by taking a fixed percentage of all bets placed, known as the vigorish. Intuitively, betting houses do not risk losing money if exactly half of the total amount of money bet is on one side of the betting line and the other half on the other side, regardless of the outcome. The money that is bet on the losing side of the outcome is used by to pay off the winning wagers. If this does not occur, the betting house incurs some risk.

We provide more intuition with a simple example. Suppose that a betting house sets a betting line on the Michigan versus Ohio State game where Ohio State is favored to win by ten points. Suppose that the game occurs and Ohio State wins by 31 points, thus “beating” the betting line. For the betting house, they encounter no risk if half of the total money wagered was bet on Ohio State to win by more than ten points and half of the money bet on Ohio State to win by fewer than ten points. If, however, sixty percent of money wagered is bet on Ohio State beating the line, then the betting house will only be able to cover 2/3 of the winning bets with money from the losing bets. In this case, the betting house will lose money, paying out not only the money bet on Ohio State to lose to the line, but the fees charged all wagers (the profit for the betting house) and additional monies from outside of this particular bet.[4]

Although betting houses could profit handsomely if there were more losing bets than winning bets, such a strategy would involve substantial risk to the betting house. For this reason, betting houses are primarily interested in setting betting lines that will guarantee equal betting on either side of the betting line[5]. The betting line can be better understood as the wager-weighted median of the distribution of bettor beliefs about the outcome of the game in question as opposed to the expected value of the outcome. This profit motive on the part of the betting house, and its implication for what the betting line should represent, forms the basis of the tests of market efficiency. Intuitively, bettors will be indifferent about a particular betting line when it contains all of the available information that bettors would use to place their wager. If information is available that would induce bettors to be on a particular side of a bet, then the betting line has not been properly set.

[4] Levitt (2004) uses professional football data to suggest that betting houses may sometimes take very strong positions on the outcome of the game, in order to exploit biases that bettors may have. By exploiting these biases, betting houses may be able to make more profit (in the example above, the betting house would profit substantially if 60% of bettors had bet on Michigan beating the spread). We cannot test this claim directly with our data; however, the odds we use in the following empirical analysis are aggregates over many different betting houses, so even if one particular betting house is exploiting this strategy, other betting houses might quickly move to eliminate the viability of this strategy by offering sophisticated bettors different odds.

[5] Kilby (2002) and Roxborough (1991) point out in guides for sports book management that a bookmaker’s primary objective is, in fact, to minimize risk.

3. Conceptual Framework

We start by establishing a minimum acceptable threshold for a bettor to make profit in the betting market. With some probability, *p*, the bet will be successful, and if the event has only two outcomes, the remaining probability, *1-p*, captures the instance when the bet is not successful. The revenue of a bettor is multiplied by the bet size, *B*.[6] Also, every bet has a fixed percent that is given to the betting house, *c,* the vigorish. The threshold for any particular bet is thus a probability of success that exceeds the sum of the probability of failure and the transaction cost. For a market to be efficient, it must be the case that, *for all strategies,*

$$pB - cB \ \leq (1-p)B \tag{1}$$

and

$$(1-p)B - cB \leq \ pB \tag{2}$$

The left hand side of (1) represents the profit made by betting on a team to beat the betting line and the bet being successful. The left hand side of (2) represents the profit made by betting against a team to beat the betting line and the bet being successful. The right hand sides of both equations represent the loss realized from an unsuccessful bet. Factoring out *B* and rearranging terms yields the key relationship: $p \ \in \left\{\frac{1}{2} - \frac{c}{2}, \frac{1}{2} + \frac{c}{2}\right\}$. A bettor will place a bet on a team to beat the spread if she believes that the team has a greater than $\frac{1+c}{2}$ percent chance of beating the betting line. Conversely,

[6] Betting houses may cap the size of the bet; for example, Kilby and Fox (2002) suggest that gambling houses should cap the size of the bet at $2,000 for college football, although betting houses may choose to pursue higher limits if they feel that bettors are particularly uninformed. However, this would not prevent bettors from having others bet for them. Thus, we model bet size to be arbitrarily large, although in principle the bet size is capped. We return to this idea later, when we discuss potential limits to arbitrage in this market.

a bettor will bet against a team beating the spread if she believes that the team has less than a $\frac{1-c}{2}$ percent chance of beating the betting line.[7]

We test to see whether a particular strategy can exceed this threshold. If $c = 0$, there were no vigorish, then a risk-neutral bettor would be indifferent between betting that a team would beat (or, conversely, lose to) the line if the probability that a team beat the spread was equivalent to the probability that a team lost to the line. (This would represent a coin flip). Otherwise, the bettor would bet on the outcome that was more likely. Note that since all participants in the betting market pay the vigorish, they are inherently risk-loving.

Betting houses aggregate this betting behavior to make profit. From the perspective of a betting house, a betting line *l* is efficient if it guarantees profit in expectation. More formally, the betting house chooses *l* to satisfy both

$$\mathbb{E}\left(p_t^b \middle| l_t, \Omega_t(l_t, I_t)\right) B - \mathbb{E}\left(1 - p_t^b \middle| l_t, \Omega_t(l_t, I_t)\right) B + cB \geq 0 \tag{3}$$

and

$$\mathbb{E}\left(1 - p_t^b \middle| l_t, \Omega_t(l_t, I_t)\right) B - \mathbb{E}\left(p_t^b \middle| l_t, \Omega_t(l_t, I_t)\right) B + cB \geq 0 \tag{4}$$

where l_t is the line at time *t*, Ω_t is the distribution of bettor beliefs about the likelihood of beating a betting line, and I_t is the information set of the bettors at time *t,* which includes all relevant information a bettor chooses to use when betting on a particular team to beat the betting line. This information set includes any information that bettors may use, such as injuries, whether a team has a strong tradition, opponent strength, and past results against the line. From a betting house's

[7] For most betting markets, this number is 52.4 percent.}

perspective, any $p_t^b \in \left\{\frac{1}{2} - \frac{c}{2}, \frac{1}{2} + \frac{c}{2}\right\}$ guarantees that (3) and (4) hold. Note further that for common distributions of beliefs, such as uniform and normally-distributed beliefs, setting a line close to the median is efficient—it guarantees that bettors will be indifferent between betting on either side of the betting line, which in expectation would give equal amounts of money on either side of the bet.

Under certain assumptions, risk-minimizing behavior is profit-maximizing. Suppose that, for a given line l_t, the cumulative probability distribution associated with individuals who believe the line is too low is $f(l_t)$ and the cdf associated with individuals who believe the line is too high is $g(l_t)$, such that $f(l_t) + g(l_t) = 1$. Furthermore, suppose that the betting house has the same beliefs as the bettors, such that $q(l_t) = f(l_t)$. For a given line, the betting house's profit function becomes:

$$f(l_t)B + g(l_t)B + c(f(l_t)B + g(l_t)B) - 2q(l_t)f(l_t)B - 2(1 - q(l_t))g(l_t)B \quad (5)$$

Here, setting l_t such that $f(l_t) = .5$ is both risk-minimizing and profit-maximizing.

We define a market as efficient only when there are no profitable betting *strategies* that can be employed by bettors to make profit in the long run. That is, while betting a particular strategy in any given game may result in a win or loss, betting a simple strategy (such as always betting on home teams) will not result in consistent profit. This is analogous to the semi-strong form of the efficient markets hypothesis (Fama 1970), which states that fundamental analysis cannot be used to make long-run profits. We define fundamental analysis in our environment as always using a particular characteristic of the game (location of game, favorite status, etc.) to predict the likelihood of beating the line.[8] Using this measure as our test does not require the actual margin of victory to be any

[8] This would be similar to always using a particular set of characteristics to choose the stocks for a portfolio that would be held for a long period of time, such as only choosing utility companies or stocks with low price/earnings ratios or low betas.

particular distance from the betting line. The only criterion for profit is whether or not a particular game or team characteristic is statistically related to beating the betting line, which is a binary outcome. Thus, we can test to see whether or not particular characteristic beat the line more often than not (or particular characteristics lose to the line more often than not), which allows us to directly test this variant of the EMH.

4. Data

We exploit two sources of data to test for market efficiency in this betting market. One source is betting lines taken from sports handicapper Jim Feist's workbook, as used in Paul et. al, (2003). This data source consists of betting lines from all Division 1-A games from 1985-2005, for which betting lines exist. From this data we obtain information on game location, game results, and the betting lines themselves. As this data is comprehensive and is the total universe of college football betting lines, we refer to this in our analysis as the ``total sample."

We also combine the betting line data with the rich data on game characteristics first presented in Logan (2009). The Logan data contains the location of the game, the date of the game, the opponents of the team before and after the game, the conference of the team, the score of the game, the team’s record before and after the game, and detailed information on the teams' opponents including opponent’s contemporaneous and season winning percentages, their poll ranking in both the AP and Coaches Polls. There are limits to the Logan data, however, as it covers only 25 of the most popular teams in college football. These teams are listed in the appendix. The strength of the Logan data is that it allows us to construct measures such as opponent strength, and to consider a richer set of strategies than is available in the raw betting lines themselves. We refer to this matched data as the “Logan sample.”

Table 1 presents the summary statistics for both the total data and the Logan sample. In the total data, the average margin of victory is less than half a point, on average. The average betting line is close to zero, and teams are more likely to be favored when playing at home than away. Because the Logan sample consists of 25 of the most successful programs, these teams are favored more than 70% of the time, and yet beat the betting line only slightly more often (52%). The average betting line in the Logan sample is 7.81 points, which means that teams in the Logan sample are favored to win by slightly more than a touchdown. By game location, teams in the Logan sample are much more likely to be favored. Consistent with the data being for the most successful college football teams, the average ranking of a team in the Logan sample is 10th in both the AP and Coaches polls. On average, the teams in the Logan sample play opponents who win one more game than they lose in a given season, which we would expect given the conference structure under which the most prominent teams play.

5. Empirical Results

5.1 Testing for Betting Strategies

Various tests of sports market betting efficiency exist. One potential method is to regress the actual outcome of a game on the associated betting line and a list of meaningful covariates, and then to test whether or not the betting line is statistically different from the actual outcome. Variations of this method have been used by Zuber et. al (1985), Gandar et. al (1988), Sauer et. al (1988), Golec and Tanarkin (1991), Dare and McDonald (1996), Dare and Holland (2004), and others on NFL data. It has also been used by Fair and Oster (2007) on college football data for a limited number of years. Most of these studies have found that the betting line is not statistically different from the margin of victory on average, and use this to infer that betting markets are efficient.[9]

While this technique is a test of whether or not the betting line is a predictor of the actual margin of victory, it is not a direct test of whether or not the betting market is efficient. The technique described above characterizes efficiency as how close the actual margin of victory is to the betting line, but this is not important to market participants (either the betting houses or bettors, as shown earlier). Betting houses are concerned with getting even money on either side of the bet, and bettors are concerned with whether or not the actual margin of victory will be greater or less than predicted. These standard techniques do not address these questions. At best, the results of standard tests of market efficiency are indirect tests of market efficiency.

To see how this is an indirect test, consider the following: suppose that all of the teams that win by more points than predicted by the betting line possess a certain characteristic, (i.e., they play games at their home field). If this scenario is true, then it is clearly a dominant strategy for a bettor to bet on the home team to beat the line, since the bettor will win more often than lose and make profit. This type of scenario cannot be captured by noting how close a game's predicted margin of victory is to the actual margin of victory, because the two could be arbitrarily close over all games, and yet small enough in others (e.g., home games) to make particular betting strategies profitable. In general, this test is not powerful against reasonable alternatives.

Our approach to testing for efficiency in the market begins by considering the basic claims of EMH, which hinge on information and strategic actions. We denote whether a team beats the line as Y_i, which takes the value 1 if team *i* beats the line and 0 otherwise. Strategies, we argue, would

[9] We show the results of the traditional test for our data in the appendix.

be based on game characteristics, which are independent sources of variation between games. For example, whether or not a team is playing at home, or whether or not a team is favored by the betting line itself are both pieces of information that individuals could use as a rule of thumb with which to bet on the outcome of a game. If favored teams beat the line sufficiently often, it would be profitable to bet on favored teams, and we denote this as a possible *betting strategy*. If the betting market is efficient, then no *strategy* based on readily observable game characteristics should yield positive profit. These sorts of patterns, we argue, would readily be discerned by bettors and betting houses, and the possibility for arbitrage would disappear.

To capture whether or not these strategies meaningfully increase the possibility of winning a bet, we estimate a probit model that takes the following form:

$$\Pr(Y_i = 1) = \Phi\left(X_i'\beta\right) \tag{6}$$

where $\Phi(\cdot)$ is the cdf of the standard normal distribution, and X_i is a vector of characteristics that pertain to specific strategies (see Gray/Gray 1997 for an application of this method to professional football). These characteristics include a team's rank in the AP poll, the strength of the opponent, and the week of the season that the game is played. This approach allows us to uncover whether betting houses, and therefore bettors, misprice the likelihood of a team beating the line based on public information about a given game.

We test a number of different, simple betting strategies in order to fully characterize and test for efficiency. There may be many different rules of thumb that bettors would employ in order to try and make profit consistently in betting markets. Different strategies have different expected outcomes. Suppose an individual chooses to bet on home teams. Home teams may have an advantage because their fans may be able to cheer very loud, affecting the other team's ability to use timeouts and call plays, or the visiting team may be unfamiliar with the playing surface, wind conditions, temperature, or other variables that may impact the result of a game. Additionally, athletes may be nervous when playing in relatively unfamiliar, hostile environments.

Table 2 reports the results of the probit specification, where we test whether or not particular strategies are consistently profitable. We find that there is evidence, both in the total data and in Logan sample, for favorites being overpriced relative to the line. In the total data, favorites are 1.86 percent less likely to beat the line. Home teams are 2.05 percent more likely to beat the line, which suggests that home teams are underpriced and that favorites are overpriced relative to the line. In the Logan sample, favorites are 6.38 percent less likely to beat the line, which suggest that favorites who happen to be teams with strong traditions are significantly overpriced. In the Logan sample, betting against favored teams with strong traditions would allow for large profit if the strategy was consistently applied. Whether or not a team is favored significantly impacts that team's associated probability of beating the betting line in every model. Favorites are consistently less likely to beat the betting line in every specification, from 1.86 percent to 2.54 percent in the total data to 1.82 to 6.84 percent in the Logan sample. Bettors systematically overprice favorites.

We find that betting $1,000 against favorites in the Logan data would, in expectation, yield $1,016.99 after accounting for the betting house's cut of a bet, which is $100 for a winning bet. Table 2 also shows that bettors potentially misprice the game based on realized opponent strength, which we define as wins - losses at the end of the season.[10] Opponent strength is significant in all specifications: the marginal effect of increasing opponent strength by one win reduces the probability of beating the spread by one to two percent (specification values range from 1.19 to 1.96 percent). This is of particular interest, because the line presumably accounts for opponent strength,

---

[10] In specifications where opponent strength is defined at the time the game is played, the coefficient is of similar magnitude but is only marginally significant due to measurement error, particularly early in the season.

which would be key in determining the predicted margin of victory. We find evidence in the opposite direction: as the opponent's strength improves, teams are less likely to beat the betting line (and, for that matter, win the game).[11]

We consider the profitability of a richer set of specific strategies in Table 3, which tests the marginal effects of interaction strategies using both the total data and Logan sample. In particular, we test to see whether or not betting on a team that possesses a particular joint characteristic (e.g., a favorite playing at home) is more profitable than betting randomly. We find that in both the total data and the Logan sample betting on home underdogs is statistically inefficient. In the total data, home underdogs are 1.92 percent more likely to beat the spread, and this is significant at the 95 percent level. Furthermore, in the Logan sample, home underdogs are 18.5 percent more likely to beat the spread. While the Logan results may be a function of a small number of home underdogs, the results suggest the presence of an arbitrage opportunity for this set of teams.

As further evidence, we also check that the data is consistent with the inverse of the strategies. Since betting on home underdogs implies betting against away favorites (excluding games played at neutral sites), we test to see if betting against away favorites is a profitable strategy. We find that betting against away favorites is a profitable strategy in all specifications. In the total data, away favorites are 4.35 percent less likely to beat the spread. We find this to be strong evidence for the existence of profitable strategies in this betting market.[12]

---

[11] Since opponent strength is determined by season it could be sensitive to week of season effects, opponent strength could be difficult to determine early in the season. We test this finding for robustness to week of season by using a fixed effect model, controlling for week of season (reported in the appendix). We find no significant effect for week of season, and the effect of opponent strength is robust.

[12] Dare and Holland (2004) suggest that these effects should be exactly opposite to each other, and that this should therefore be a model restriction. We disagree for two reasons; one is that there may be unobserved heterogeneity that is associated with away favorites but not with home underdogs, particularly because point spreads are not set in all college football games. Another is that, unlike professional football data, our data is not symmetric since a number of games are played at neutral sites. Thus, we do not have exactly the same number of home underdogs and away favorites in the data (see Table 1).

6. The Mechanism of Inefficiency.

The evidence presented in the previous section establishes that the college football betting market is inefficient when pricing both home teams and favorite teams. While finding inefficiency is interesting, the mechanism for inefficiency is more economically interesting, since the mechanism may explain inefficiencies in other markets. Here, we analyze and identify a potential source of those inefficiencies: memory in the betting lines. We define memory in asset pricing as using past data to predict and price future returns. In betting markets, the betting line is the price, and lines may exhibit memory if previous information and previous performances by teams affect the line in a particular week.

Memory in asset pricing has been documented in other economic settings (see Lo 1991 and Hirshleifer 2001). Past documented increases in an asset's price may lead to beliefs that future prices of the asset may increase, and may indeed lead to price changes in an asset. These beliefs may be correct and past performance may represent strong asset fundamentals. However, these beliefs may also lead to biased estimates of future prices, as investors or market participants may recall or use only favorable subsets of available information to make predictions of future prices. We focus on this second property of memory, that of biases.

. Studies have documented that market participants may believe that data clustering represents future trends rather than statistical anomalies (see Hirshleifer 2001, Barberis/Thaler 2002, and Durham/Heitzel/Martin 2005), which may lead to strategies such as momentum trading. Relying on biases may be a sign of coarse thinking, as described in Ellison and Holden (2008), Mullainathan, Schwartzstein, Shleifer (2008) and others. Looking for memory in betting lines allows us to identify whether or not bettors' beliefs about the likelihood of winning change due to past events.

We proceed in three parts: first, we demonstrate that betting lines exhibit memory. We then document evidence of bettor belief changes in the college football betting market. Finally, we test to see whether or not these belief changes contribute to the aggregate inefficiency found with pricing of home and favorite teams.

6.1 Does Past Performance Predict the Betting Line in College Football?

We first analyze how betting houses set betting lines. Recall that I_t is the information set that a better uses at time *t* to evaluate a particular team, and it includes all of the information available in previous time periods. If betting houses are profit-maximizers, then they set lines to get roughly the same amount of money on both sides of the line to satisfy (3) and (4). Equations (3) and (4) depend on both the betting line, a choice variable for the betting house, and $\Omega_t(l_t, I_t)$, the distribution of beliefs that include the information set that bettors use to assess the likelihood of beating the spread.

To formalize how memory affects how bidders, we first suppose that I_t contains only time-invariant characteristics, which we define as characteristics that bettors believe should be priced independently from time, such as home-field advantage, whether or not a team is favored, whether or not a team is playing an opponent of high quality (or is of high quality themselves), etc. Time-invariant characteristics may change from week-to week, but are generally known in advance. A team may play at home or away, may be favored or not favored, and may play high or low-quality teams, but bettors do not believe that these characteristics should be priced differently in any given week. Moreover, betting lines would not exhibit memory if I_t only contained time-invariant characteristics. Game location, team quality, and favorite status are not contingent on when the game is played, thus bettor beliefs about particular realizations of I_t are independent of time, and $\Omega_{t+1}(l_{t+1} = l, I_t) = \Omega_t(l_t = l, I_t)$, which means that, for any given realization of I_t, a bettor believes that the likelihood of a team beating a betting line l is the same at any given point of time, implying that $\mathbb{E}(p_{t+1}|l_{t+1} =$

$l, \Omega_{t+1}(l_{t+1} = l, I_t)) = \mathbb{E}(p_t | l_t = l, \Omega_t(l_t = l, I_t))$. Since the profit-maximization condition does not change over time when beliefs are independent over time, betting lines are independent over time.

Now suppose that I_t contains both time-invariant and time-variant characteristics, those that bettors believe should be priced differently depending on time. Consider a partition $\zeta_t \subset I_t$ that contains only the time-variant characteristics of I_t, such as whether or not a team beat (or lost to) the spread in the previous week. Furthermore, fix all of the time-invariant characteristics, such as pricing of home teams, opponent strength, favorites, etc. If a betting line does not exhibit memory, then $\Omega_t(l_t = l, \zeta_t) = \Omega_t(l_t = l, \zeta_t')$ for any given *l* and for all $\{\zeta_t, \zeta_t'\}$. However, if a betting line does exhibit memory, then, for different time-varying partitions, it must be the case that $\Omega_t(l_t = l, \zeta_t) \neq \Omega_t(l_t = l, \zeta_t')$ for some pair $\{\zeta_t, \zeta_t'\}$. Intuitively, if every set of different time-varying partitions induced the betting house to choose the same line, then it must be that the beliefs about the likelihood of the team beating that line remained the same, otherwise the betting house bears more risk than necessary, and violates our assumption of profit maximization. Alternatively, if the betting house changes the line, it must be that bettor beliefs have changed, otherwise the betting house is again bearing more risk than necessary.

We therefore test to see whether betting lines have memory in Table 4 by estimating the following equation:

$$bettingline_t = \gamma_0 + \gamma_1 beatspread_{t-1} + \gamma_2 beatspread_{t-2} + \gamma_3 beatspread_{t-3} + \upsilon_t \quad (7)$$

If beating the betting line is not important to a betting house at time t, we would expect the coefficients on beat spread to be close to zero and statistically insignificant, while if there is evidence that beating the betting line matters to betting houses, we would expect the coefficients to be positive, indicating that betting houses make it more difficult for a team to beat the betting line in successive weeks. There is evidence for the latter scenario in columns I-III in Table 4: beating the betting line adds over 2 points to the betting line for a team, after accounting for both previous games that a team has beat the spread and whether or not a team is playing at home. Additionally, betting lines have memory from weeks prior to the previous week. Column II of Table 4 indicates that if a team beats the spread in the previous three weeks, then the betting house adds over 7.5 points to the betting line in the current week. This result is robust to the addition of further controls, as seen in Column III.

We also test to see if the magnitude by which a team beats the betting line is important to betting houses. Beating the betting line by a small number of points may appear random to a betting house, while beating the line by many points may represent an increase in team quality and perceived strength. That is, we estimate:

$$bettingline_t = \zeta_0 + \zeta_1 magnitude_{t-1} + \zeta_2 magnitude_{t-2} + \zeta_3 magnitude_{t-3} + e_t \quad (8)$$

We report the results in columns IV-VI of Table 4. We find that teams who beat the betting line by many points are favored more heavily in subsequent games, and that this is more important to betting houses than whether or not a team beats the betting line. In particular, every additional point that a team beats the betting line by adds between .082 and .108 points to a subsequent betting line. Intuitively, beating the spread by a touchdown adds between one half to a full point to a betting house's subsequent betting line.

Columns I-VI reject the idea that betting lines have no memory from week to week, and, from the perspective of the betting house, reject the hypothesis that bettor beliefs do not contain serial correlation. If bettor beliefs did not change from week to week based on previous performances against the spread, then the betting house would have no reason to change the line based on time-varying characteristics, because that would cause it to take on unnecessary risk.

We find that betting lines are functions of previous outcomes, but do those previous outcomes predict performance against the spread? It could be that teams who beat the spread in previous weeks are less likely to beat the betting line, or vice versa. Columns VII-IX of Table 4 test the profitability of betting on a team who beat the betting line in previous weeks. Column VII shows that whether or not a team beats the betting line in the previous week does not improve the probability of a team beating the betting line in the current week. Columns VIII-IX show that this result is robust to consecutive weeks of beating the spread.

These results show that betting houses react to changes in time-varying characteristics. A betting house only has incentive to change the betting line if bettor beliefs about the likelihood of a team beating a given line change based on these characteristics. While bettor beliefs may change, it is not clear whether beliefs change due to biases or changes in fundamentals. To identify whether or not belief changes are due to biases, we re-examine how betting houses price *time-invariant* characteristics after accounting for bettor beliefs about the time-varying characteristics. Betting houses' incentives to price these strategies according to (3) and (4) should not change in the presence of time-variant strategies. If time-varying strategies cause betting houses to price time-invariant strategies differently, then the inefficiencies we find may be due to this different pricing. To test this hypothesis, we first document evidence of bettor beliefs in this market, and then examine whether betting houses price time-invariant strategies differently because of those beliefs.

6.2 Market Function and Market Inefficiency

A well-documented time-varying characteristic is beating the spread in consecutive weeks. This characteristic is not known in advance of the season; rather, it evolves over time and changes from week to week. A strategy derived from this characteristic is to bet on teams that perform well relative to the spread in previous weeks. This strategy shares properties with momentum trading, data clustering, and other behavioral strategies in financial markets (see Jeegadeesh/Titman 2001 and others). This strategy is known as the "hot hand effect," and is well documented in the literature (see Camerer 1989 and Brown and Sauer, 1993 for a discussion). The "hot hand" is a notion that captures the idea that an individual's probability of succeeding at an event changes over time in a deterministic way, such as the individual improving at a task, like shooting a basketball during a game, or even winning games in general.

From our search of the qualitative record on college football betting, we found that the "hot hand" is a very popular strategy:

"Following a team's winning streak is one of the best way of making money in sports gambling, as everybody loves to ride the hot team![13]"

"You need to find out which teams are blazing hot and seemingly can't lose to anybody and the teams that are cold as ice which look like they couldn't beat themselves. It's basic common knowledge and handicapping strategy to try and ride a winner until she bucks you and to stay away from dead beats.[14]"

"I am leaning towards Boise and the points. Right now, the line is sitting at 9.5. Boise has whipped Fresno 4 straight and I think the streak may even continue[15]."

Although this strategy is very popular among bettors, it is well-accounted for by betting houses. Betting houses increase the spread even more prominently for teams that beat the spread multiple weeks in a row. Columns VII-IX of Table 4 indicate that betting on hot teams is not a profitable strategy for bettors, as the probability of winning a bet falls well within the efficient range given by (1) and (2). Bettor beliefs about the likelihood of hot teams continuing to beat the spread seem to be misplaced. Intuitively, perhaps the reason why is that teams win more games after

[13] http://www.atsdatabase.com/blog/blog1.php/2009/09/15/rice-owls-vs-oklahoma-state-cowboys-coll
[14] http://www.locksmithsportspicks.com/college-football-trends/
[15] http://www.againsttheline.com/board/showthread.php?t=1146&highlight=streak

beating the spread, but do not fare better against the spread. Another possible reason is coarse thinking. Bettors may weigh past examples of "hot teams" more heavily in their decision-making process with regard to their betting, while forgetting teams that beat the spread one week and lose to the spread the next week.

In order for this to be the driving force behind the inefficiency, counteracting the hot hand must cause inefficiencies with respect to other strategies. To test this, we analyze how betting houses price time-invariant strategies, such as betting on home or favorite teams, after accounting for the "hot hand." Table 4 indicates that teams are priced differently based on how they perform against the betting line the week before. This provides an additional source of variation both for bettors (bettors can construct a list of teams who performed well or poorly relative to the betting line and devise a strategy based on this information) and for the betting house, which can observe performance compared to the spread and price games based on this rule and other characteristics. Theoretically, betting houses have no incentive to price time-invariant strategies differently from week to week, since bettors do not believe that these characteristics should be priced differently. The only reason these strategies would be priced differently, from the perspective of a betting house, is if they were to occur in conjunction with time-varying beliefs. If bettors believe that a team's performance against the spread is the foremost predictor of future outcomes against the line, then a betting house may price the hot hand strategy more prominently played by bettors according to the criterion in (3) and (4) in order to minimize risk, at the expense of other strategies. By estimating our original model in the light of hot hand pricing, we can isolate whether or not time-invariant strategies are priced differently in the light of time-varying strategies. Additionally, we might be able to learn what drives the magnitude of the inefficiencies that we find in the total sample, where we ignore the presence of hot hand pricing.

Utilizing the additional variation provided from hot hand pricing, we look at how these strategies are priced for teams that are "hot," having beat the spread in at least the previous week. Absent any time-varying characteristics, these strategies should not be priced differently from our total sample: the marginal effects should be the same. If they are not, then the inefficiency for a particular strategy, such as betting on home teams, might arise only when in the presence of a particular time varying strategy. We test to see how betting houses price time-invariant strategies given that they occur in conjunction with a time-varying strategy. To do so, we split our sample into four categories: teams that beat the spread last week, teams that beat the spread in two consecutive weeks, teams that lost to the spread last week, and teams that lost to the spread in either of the last two weeks. We then estimate our baseline model in Table 5 for those different subsamples.

The first three columns of Table 5 estimate our model only for the teams that beat the spread last week. Of those teams, home teams are more likely to beat the spread- 3.91 percent more likely, compared with 2.05 percent in our total sample. Time-invariant strategies are priced differently in the presence of time-varying strategies. If those teams are favored, they are neither more likely nor less likely to beat the spread. These columns tell us the source of the underpricing that we find in the market for home teams. Home teams are systematically underpriced by nearly double the magnitude of that found in the total sample. The home teams that are underpriced in our sample share, statistically, the feature of having beat the betting line the week before—the magnitude of the marginal increase in probability is driven specifically by teams priced differently due to a time-varying characteristic. Moreover, in this sample, we do not find evidence that favorites are overpriced, although they are overpriced in the total sample.

Columns IV-VI of Table 6 account only for teams that did not beat the spread in previous weeks. Of those teams, teams that are favored are 3.05 percent less likely to beat the spread, compared with 1.86 percent in our total sample. Again, the magnitude of the marginal effect found in the total sample is explained by pricing of teams that did not beat the spread in the previous week. While we find that the line is systematically lower for these teams, if the team is favored, the line is still systematically high enough to generate a three percent overpricing of favorites. The magnitude of the favorite overpricing found in the total sample is again driven by a subset of teams priced differently because of a behavioral strategy. For teams that did not beat the spread before, home has no predictive power of whether or not a team beat the spread for the subset of teams that did not beat the spread before, but favorite does, and it does with a larger magnitude.

Taken together, Columns I-VI parse out how inefficiency is created in the market. We found that home teams are underpriced and favorites are overpriced in our total data, and Table 6 tells us why- home teams are particularly underpriced for a subset of teams that the betting house prices differently because of a time-varying characteristic (specifically, the hot hand effect), and favorite teams are particularly overpriced for a subset of teams that the betting house prices differently because they lost to the spread. In other words, the market inefficiency of under-pricing home strategies is driven by the betting house's desire to prevent the hot hand effect. Similarly, the market inefficiency for overpricing favorites is driven by a subset of teams that lose to the spread, the opposite of the hot hand.

Columns VII-XII of Table 6 provide corroborating evidence for why betting houses misprice favorites and home teams. In columns VII-IX, we report our basic model specifications for teams that beat the spread in consecutive weeks. Intuitively, bettors might view these teams as being "hotter" than other teams, and Table 4 demonstrates that betting houses add around 4.5 points to the betting line for these teams. We find that home teams are 3.61 percent more likely to beat the spread for this subsample—betting houses continue to misprice home teams by accounting for the hot hand effect. Once again, favorites are statistically insignificant. Columns X-XII report results for teams that lost to the spread in either week (or, potentially, in two consecutive weeks). We find that these teams, if favored, continue to be overpriced.[16]

Table 6 reveals an interesting strategy set available to bettors in this market. A bettor could gather a list of all the teams that beat the spread and only bet on the home teams, or gather a list of all the teams that lost to the spread in the previous weeks and bet against those teams. These conditional strategies are more profitable than unconditionally betting against away and favorite teams.

7. Conclusion

We provide evidence that the college football betting market exhibits market inefficiency. We test for market efficiency directly by using the predicted success of specific betting strategies. We find evidence for statistical inefficiency when pricing home teams, favored teams, and teams that play weak opponents. Moreover, we find evidence for profitable betting against favored teams with strong traditions and on teams that play weak opponents. We find that betting houses systematically overprice and underprice particular subsets of games. These inefficiencies could be exploited by bettors to make significant profit in the betting market.

More interestingly, we find that betting houses' responses to the behavioral biases of bettors are a source of this inefficiency. In this market, betting houses seek to eliminate profitable opportunities that may arise from a team entering a winning streak relative to the betting line by increasing the threshold that this team has to overcome in order to beat the spread the next week.

[16] We provide further evidence of these effects in the appendix, where we demonstrate two results: that standard tests for market efficiency are unchanged by subsample analysis, and that home teams that beat the spread last week are more profitable among home teams. These can be viewed as robustness checks for our main results.

By doing so, the betting house removes one potentially profitable (a widely popular) strategy from bettors, but this leaves open other, less well-known, strategies.

Further research should investigate the link between market function and inefficiency. If, in other markets, market makers respond to behavioral strategies very strongly, then trading strategies that are unrelated to those strategies may become profitable. This may have strong implications for pricing in asset markets where behavioral strategies are prominent (see Hirshleifer 2001, Jegadeesh and Titman 2001, Durham et al. 2005). Indeed, the success of some investors may be due to their ability to uncover the strategies that become profitable when market movers act to take away prominent sources of potential arbitrage.

References

Against the Line. "Boise @ Fresno: College Football Message Boards." http://www.againsttheline.com/board/showthread.php?t=1146&highlight=streak (accessed September 25, 2009)

Avery, Christoper and Judith Chevalier. Identifying Investor Sentiment from Price Paths: The Case of Football Betting. *Journal of Business,* 72(4): 493-521, 1999.

Barberis, Nicolas and Richard Thaler. A Survey of Behavioral Finance. *NBER Working Paper 9222,* 2002.

Boyd, Jimmy. "College Football Trends | NCAA Football Betting Trends" http://www.locksmithsportspicks.com/college-football-trends/ (accessed September 25, 2009)

Brown, William and Raymond Sauer. Fundamentals or Noise? Evidence from the Professional Basketball Betting Market. *Journal f Finance,* 48(4): 1193-1209, 1993.

Bröder, Arndt and Benjamin Scheibehenne. Predicting Wimbledon 2005 Tennis Results by Mere Name Recognition. *International Journal of Forecasting.* 23(3): 415-426, 2007.

Brown, William and Raymond Sauer. Does the Basketball Market Believe in the Hot Hand? Comment. *American Economic Review,* 83(5): 1377-1386, 1993.

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

Cowgill, Bo, Justin Wolfers, and Eric Zitzewitz. *Using Prediction Markets to Track Information Flows: Evidence from Google. mimeo.*

Dare, William and Steven Holland. Efficiency in the Betting Market: Modifying and Consolidating Research Methods. *Applied Economics,* 36(1): 9-15, 2004.

Dare, William and Scott McDonald. A Generalized Model for Testing the Home and Favorite Team Advantage in Point Spread Markets, *Journal of Financial Economics.* 40: 295-318, 1996.

Durham, Gregory, Michael Hertzel, and J. Spencer Martin. The Market Impact of Trends and Sequences in Performance: New Evidence. *Journal of Finance.* 60(5): 2551-2569, 2005.

Ellison, Glenn and Richard Holden. A Theory of Rule Development. *mimeo,* 2008.

Fair, Ray and John Oster. College Football Rankings and Market Efficiency. *Journal of Sports Economics,* 8(1): 3-18, 2007.

Fama, Eugene. Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance.* 25(2): 383-417, 1970.

Gandar, John, Richard Zuber, Thomas O'Brien, and Ben Russo. Testing Rationality in the Point Spread Betting Market. *Journal of Finance,* 43(4): 995-1008, 1988.

Gil, Ricard and Steven Levitt. Testing the Efficiency of Markets in the 2002 World Cup. *Journal of Prediction Markets.* 1(3): 255-270, 2007.

Golic, Joseph and Maury Tamarkin. The Degree of Inefficiency in the Football Betting Market. *Journal of Financial Economics,* 30: 311-323, 1991.

Gray, Phillip and Stephen Gray. Testing Market Efficiency: Evidence From the NFL Sports Betting Market. *Journal of Finance.* 52(4): 1725-1737, 1997.

Hirshleifer, David. Investor Psychology and Asset Pricing. *Journal of Finance.* 56(4): 1533-1597, 2001.

Jegadeesh, Narasamhan and Sheridan Titman. Profitability of Momentum Strategies: An Evaluation of Alternative Explanations. *Journal of Finance.* 56(2): 699-720, 2001.

Kilby, Jim, Jim Fox, and Anthony Lucas. Casino Operations Management, Second Edition. Wiley: Hoboken, 2002.

Levitt, Steven. Why are Gambling Markets Organized so Differently from Financial Markets? *Economic Journal.* 114: 223-246, 2004.

Logan, Trevon. Whoa, Nellie! Empirical Tests of American College Football's Conventional Wisdom. Forthcoming*, Applied Economics.*

Lo, Andrew. Long Term Memory in Stock Prices. *Econometrica.* 59(5): 1279-1313, 1991.

Matuszewski, Erik. "Vegas Betting Cops Help NCAA Find Suspicious Basketball Wagers." http://www.bloomberg.com/apps/news?pid=20601079&sid=aDptJsOc5pGo&refer=home
March 19, 2009.
Mau, Ronald. Mispricing of Rules and Regulations: The College Football Wagering Market. *mimeo, University of Kansas.* 2004.
Mullainathan, Sendhil, Joshua Schwartzstein and Andrei Shleifer. Coarse Thinking and Persuasion. *mimeo, Harvard University.* 2006.
Ottaviani, Marco and Peter Sorenson. Aggregation of Information and Beliefs in Prediction Markets. *mimeo, Northwestern University.* 2007.
Ottaviani, Marco and Peter Sorenson. Aggregation of Information and Beliefs in Prediction Markets: Lessons for Asset Pricing. *mimeo, University of Copenhagen.* 2008.
Paul, Rodney, Andrew Weinbach, and Patrick Coate. Expectations and Voting in the NCAA Football Polls. *Journal of Sports Economics.* 8(4): 412-424, 2007.
Paul, Rodney, Andrew Weinbach, and Chris Weinbach. Fair Bets and Profitability in College Football Gambling. *Journal of Economics and Finance.* 27(2): 236-242, 2003.
Raymond, Ron. (2009) "Rice Owls vs. Oklahoma State Cowboys- College Football Betting." http://www.atsdatabase.com/blog/blog1.php/2009/09/15/rice-owls-vs-oklahoma-state-cowboys-coll (accessed September 25, 2009)
Rhode, Paul and Koleman Strumpf. Historical Presidential Betting Markets. *Journal of Economic Perspectives.* 18(2): 127-142.
Sauer, Raymond, Vic Brajer, Stephen Ferris, and M. Wayne Marr. Hold Your Bets: Another Look at the Efficiency of the Gambling Market for National Football League Games. *Journal of Political Economy.* 96(1): 206-213, 1988.
Snowberg, Erik, Justin Wolfers, and Eric Zitzewitz. Partisan Impacts on the Economy: Evidence from Prediction Markets and Close Elections. *Quarterly Journal of Economics.* 122(2): 807-829, 2007.
Vlastakis, Nikolaos, George Dotsis, and Raphael Markellos. How Efficient is the European Football Betting Market? Evidence from Arbitrage and Trading Strategies. *mimeo,* 2007.
Wolfers, Justin and Eric Zitzewitz. Prediction Markets. *Journal of Economic Perspectives.* 18(2): 107-126, 2004.
Wolfers, Justin and Eric Zitzewitz. Interpreting Prediction Market Prices as Probabilities. *mimeo, Stanford GSB.* 2007.
Woodland, Linda and Bill Woodland. Market Efficiency and the Favorite-Longshot Bias: The Baseball Betting Market. *Journal of Finance.* 49(1): 269-279, 1994.
Zuber, Richard, John Gandar and Benny Bowers. Beating the Spread: Testing the Efficiency of the Gambling Market for National Football League Games. *Journal of Political Economy.* 93(4): 800-806, 1985.

Appendix

A1. Data Appendix

We use betting line data from the years 1985-2003 for 119 Division I-A teams. If a team was originially in Division I-AA and entered Division I-A, then we only include data after the team entered Division I-A. Data was taken from Jim Feist's betting workbook and was used in Paul et. al, 2003. For a given team, we have data on the date of the game, the opponent, the betting line, the location of the game, and the week of the season. In addition, for the Logan sample, we have data on the poll rank before and after the game and the win/loss record of the opponent for the season for a sample of 25 of the most prominent teams over that period, listed in Table A1. The Logan sample methodology is described in Logan (2009). In addition to the existing data, we define the following variables:

- Home favorites are defined as teams that are both playing at home and favored for a particular game. We assign a team one if it meets both criteria and zero otherwise. We define away favorites, away underdogs, and home underdogs in the same manner.
- ``Beat Spread'' is a variable set equal to one if a team beats the spread. If a team does not beat the spread, it is set equal to zero. We do not assign values for pushes.
- ``Margin Above Spread'' refers to the difference between the actual margin of victory and the margin of victory predicted by the betting line. For example, if a team is favored to win by 10, but wins by 3, then we define the margin above spread to be -7. If the team is favored to win by 10, but wins by 17, then we define the margin above spread to be 7.
- ``Opponent Strength'' is defined to be the number of wins minus the number of losses for a team's opponent at the end of the season.
- ``Week of Season'' is the poll week of the season. The first poll corresponds to preseason rankings, thus we start with week 0, which corresponds to the first poll before any games are played.

A2. Standard Test of Market Efficiency

Section A.2 provides estimation results for the standard model for efficiency in the literature:

$$MOV_i = \beta_0 + \beta_1 LINE_i + \varepsilon_i \quad \text{(A1)}$$

where MOV_i is the margin of victory (or margin of defeat) for team *i* and $LINE_i$ is the number of points that team *i* is favored to win (or lose) by the betting line. The traditional test for efficiency is that the constant, β_0 is zero and the coefficient on the betting line, β_1 is statistically indistinguishable from 1. The prevailing logic in the literature is that the betting line contains all of the information needed to predict the margin of victory, on average. If this is true, the betting line will accurately and consistently predict the margin of victory. Our results for this particular test mirror those found in this literature, and are detailed in the appendix.

We reproduce the standard test of market efficiency using both the total sample and the Logan sample. Using both sources, we find corroborating evidence for these previous results. Columns I and II of Table A2 show that the coefficient on betting lines is statistically indistinguishable from 1. In the total sample, however, playing at home adds one additional point to margin of victory that is not captured by the betting line, which suggests that while the betting line may predict margin of victory on average it excludes information that influences the actual margin of victory. In the Logan sample the predictive power of the line is slightly weaker, and it diminishes significantly once additional variables are added to the specification. For example, column V indicates that opponent strength, playing at home, and AP rank are all important indicators of the actual margin of victory. Also, the inclusion of these additional game characteristics causes the betting line to be statistically different from zero.[17]

[17] If the betting line includes all of the information that is used by bettors, the inclusion of these characteristics would cause a multicolinearity problem.

One could argue that the results of standard tests, where we show that the predictive power of the betting line declines when additional game characteristics are added to the specification, is simply due to colinearity. If the betting line captures these game characteristics including them in the regression results in biased estimates. But note that such an argument also implies that, conditional on the betting line, these game characteristics would have no predictive power, as they have been captured by the betting line.

A3. Robustness

We separate our robustness checks into two categories—robustness related to strategies we did not report, and robustness related to our estimated marginal effects. First, we test to see whether or not profitable strategies exist for betting on teams from particular conferences. Essentially, this test allows us to see whether or not some obvious lack of information causes betting houses to price games imprecisely. Second, we test to see whether or not time plays a role in our estimation; perhaps our results are driven either by variation in betting house behavior between years or by variation in betting house behavior during a given year. If either of these arguments is true, then the probabilities we estimate for the total sample do not capture an unconditional increase or decrease in the profitability of a particular strategy. Rather, our results would then capture either an anomaly or the fact that betting houses are initially imprecise in their pricing but improve over time. Neither of those explanations would correspond with market inefficiency.

A.3.1 Conference Effects

First, we test to see whether or not being a member of a particular conference influences the likelihood of a team beating the spread. It may be the case that bettors are poorly informed about the quality of a team from a particular conference, and that betting houses are aware of this and exploit it. For example, conferences that do not regularly have games televised on national television have less information available regarding team quality, and subsequently teams from this conference may be mispriced. If this is true, then we might expect it to be profitable to bet on teams or against teams from a particular conference. We estimate the marginal effects of being a member of a particular conference in Table A3. Some teams, such as Florida State, Arkansas, and Penn State were in multiple conferences during the course of our sample. However, in any given year, a team is only in one conference. We assign conference dummies based on the conference that a team was a member of in a given year. We find that there are not any significant effects on the probability of beating the betting line. The conference that a team is a member of has no bearing on whether or not a team outperforms the betting line.

A.3.2 Time and Season Fixed Effects

Our results could be biased due to particular years being heavily mispriced. For example, favorites could have been mispriced more heavily in a given year; the results from that year could in turn lead to both the magnitude and the significance of the inefficiencies we find in the market. We test for this possibility by including year fixed effects in the model, which allow us to control for any sensitivities related to year. If a particular year drives the results of our model, then including that year as a fixed effect would mitigate the inefficiency result that we find.

We report our results for the richer set of covariates in the Logan data; we find that the coefficients on both favorite and opponent strength retain statistical significance. These results are reported in the first three columns of Table A4. We do not find any evidence that betting houses priced teams significantly less efficiently in a given year.

We also test to see whether or not our results are sensitive to the week of the season that the game was played in. Perhaps betting houses become better at pricing teams over the course of a season after the quality of a team is more precisely revealed. If this is true, then games that were priced at the beginning of the season would be priced less efficiently, and the results that we find for the total sample might be driven by the inefficiencies found at the beginning of the year. To test this hypothesis, we include week of season fixed effects in our basic model for the Logan sample. We find that including week of season fixed effects does not significantly alter our estimated marginal effects. We do not find any evidence of a week of season bias driving our results.

Table 1
Summary Statistics

	Total Data					Logan Data				
Variables	**N**	**Mean**	**Standard Deviation**	**Min**	**Max**	**N**	**Mean**	**Standard Deviation**	**Min**	**Max**
Home	23224	0.5089563	0.4999305	0	1	7144	0.521277	0.4995821	0	1
Favorite	23224	0.5318182	0.4989977	0	1	4796	0.710801	0.4534379	0	1
Beat Spread	22440	0.5088271	0.4999328	0	1	4821	0.51649	0.4997798	0	1
Margin of Victory	23224	0.3825353	22.10814	-81	81	7161	8.802123	20.59927	-77	81
Betting Line	22440	-0.013347	14.85041	-59	59	4897	7.819379	13.90563	-42	55
Home Underdog	23224	0.1909662	0.393071	0	1	4796	0.092786	0.2901621	0	1
Home Favorite	23224	0.3432225	0.4747953	0	1	4796	0.42035	0.4936666	0	1
Away Favorite	23224	0.2044006	0.4032716	0	1	4796	0.237907	0.425846	0	1
Away Underdog	23224	0.2951688	0.4561284	0	1	4796	0.158465	0.3652149	0	1
Rank in AP Poll (before game)						5451	10.54302	6.560637	1	25
Rank in AP Poll (after game)						5330	10.46023	6.523023	1	25
Opponent Strength						6669	0.900135	5.231009	-12	13

See data appendix for variable definitions.

Table 2
Marginal Effects in Betting Line Models

	Total Data			Logan Data						
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)
Home	0.0205***		0.0269***	0.006				0.0159	0.02	0.0173
	(0.01)		(0.01)	(0.01)				(0.02)	(0.02)	(0.02)
Favorite		-0.0186***	-0.0254***		-0.0638***			-0.0684***	-0.182***	-0.177***
		(0.01)	(0.01)		(0.02)			(0.02)	(0.02)	(0.03)
Opponent Strength						-.0119***			-0.0190***	-0.0163***
						(0.01)			(0.00)	(0.00)
Week of Season										0.0001
										(0.00)
Rank in AP Poll (before game)							0.0001			-0.00155
							(0.01)			(0.00)
Team Clustered Standard Errors			X							X
Observations	22370	22370	22370	4805	4720	4530	2980	4704	4429	2691
Psuedo R-Squared	0.003	0.00025	0.000742	0.003	0.00242	0.0113	0	0.0261	0.0248	0.0187

Robust standard errors in parentheses (*** $p<0.01$, ** $p<0.05$, * $p<0.1$).
The dependent variable is Y = 1 if team i beat the spread and 0 otherwise.
We report the coefficients as the marginal effects of the probit regressions.
For dichotomous variables, such as home or favorite, the effects represent the change in probability from 0 to 1.

Table 3: Marginal Effects of Interaction Strategies: Full Model

	Total Data								Logan Data			
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)	(XI)	(XII)
Home Favorite	0.011				-0.001	.042***			-0.0429**			
	(0.01)				(0.01)	(0.01)			(0.02)			
Home Underdog		0.0192***						0.001		0.185***		
		(0.01)						(0.01)		(0.04)		
Away Favorite			-0.0435***				-.0411***				-0.0306	
			(0.01)				(0.01)				(0.02)	
Away Underdog				0.009								0.105***
				(0.01)								(0.03)
Home					.0262***		0.004					
					(0.01)		(0.01)					
Favorite						-.0446***		-.0151*				
						(0.01)		(0.01)				
Opponent Strength (season)									-0.0115***	-0.0127***	-0.0112***	-0.0127***
									(0.0019)	(0.0019)	(0.0019)	(0.0019)
Week of Season									0.00005	-0.00006	0.00004	-0.00004
									(0.0018)	(0.0018)	(0.0018)	(0.0018)
Rank Before Game (AP Poll)									0.00003	-0.00052	0.00001	-0.00044
									(0.0014)	(0.0015)	(0.0014)	(0.0015)
Observations	22370	22370	22370	22370	22370	22370	22370	22370	2691	2705	2691	2705
Psuedo R-Squared	0.0001	0.0002	0.0009	0	0.0009	0.0003	0.0009	0.0003	0.0105	0.0098	0.0139	0.0119

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

We report the coefficients as the marginal effects of the probit regressions.

For dichotomous variables, such as home or favorite, the effects represent the change in probability from 0 to 1.

Table 4: Determinants of Point Spreads

	Profitability of Hot Hand Strategy Dependent Variable: Beat Spread			Betting House's Reaction to the Hot Hand Dependent Variable: Betting Line					
	(I)	(II)	(III)	(IV)	(VI)	(VI)	(VII)	(VIII)	(IX)
Beat Spread (one week ago)	0.001	0.004	0.004	2.245***	2.073***	2.072***			
	(0.007)	(0.008)	(0.008)	(0.208)	(0.231)	(0.219)			
Beat Spread (two weeks ago)		0.007	0.009		2.556***	2.542***			
		(0.008)	(0.008)		(0.231)	(0.235)			
Beat Spread (three weeks ago)			0.002		2.681***	2.645***			
			(0.008)		(0.231)	(0.219)			
Home						7.277***			
						(0.208)			
Margin Above Spread (one week ago)							.091***	.087***	.082***
							(0.007)	(0.007)	(0.008)
Margin Above Spread (two weeks ago)								.108***	.102***
								(0.007)	(0.008)
Margin Above Spread (three weeks ago)									.101***
									(0.007)
Team Clustered Standard Errors			X						X
Observations	19820	17481	15283	19820	15283	15283	19820	17481	15283
R-Squared				0.006	0.022	0.085	0.01	0.023	0.034
Psuedo R-Squared	0.0001	0.00005	0.00008						

Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1).
Columns (I-III) report the coefficients as marginal effects of the probit regression where beat spread is the dependent variable.
Columns (IV-IX) report regression results where the dependent variable is the betting line.

Table 5: Strategy-Based Tests for Efficiency Among Population Subsets

	Team Beat Spread Last Week			Team Lost to Spread Last Week			Team Beat Spread Consecutive Weeks			Team Lost to Spread Either Week		
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)	(XI)	(XII)
Home	0.0391***		0.0438***	-0.00082		0.00678	0.0361**		0.0407***	0.0106		0.0169**
	(0.01)		(0.01)	(0.01)		(0.01)	(0.015)		(0.016)	(0.008)		(0.009)
Favorite		-0.0082	-0.0191*		-0.0305***	-0.0321***		-0.0125	-0.0215		-0.0219***	-0.0260***
		(0.010)	(0.011)		(0.010)	(0.010)		(0.016)	(0.016)		(0.008)	(0.009)
Observations	9782	9782	9782	10038	10038	10038	4258	4258	4258	14587	14587	14587
Psuedo R-squared	0.0011	0.0001	0.0014	0.0001	0.0007	0.0007	0.0009	0.0001	0.0013	8.17E-05	0.000345	0.00054

Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1).
The dependent variable is Y = 1 if team i beat the spread and 0 otherwise.
Specifications (I-III) include only teams that beat the spread last week.
Specifications (IV-VI) include only teams that did not beat the spread in the previous week.
Specifications (VII-IX) include only teams that beat the spread in consecutive weeks.
Specifciations (X-XII) include only teams that did not beat the spread in one or both of the last two weeks.
We report the coefficients as the marginal effects of the probit regressions. For dichotomous variables, such as home or favorite, the effects represent the change in probability from 0 to 1.

Table A1

Teams Used In Logan Sample

Alabama	Michigan
Arkansas	Nebraska
Auburn	Notre Dame
Boston College	Ohio State
Brigham Young	Oklahoma
California	Penn State
Colorado	Stanford
Florida	Tennessee
Florida State	Texas
Georgia	Texas A & M
Iowa	UCLA
LSU	USC
Miami (FL)	

Table A2
Standard Tests for Market Efficiency

	Total Data		Logan Data					
	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)
Constant	-0.00866	-0.861***	1.334***	0.434	2.838***	0.781	2.017***	10.04***
	(0.11)	(0.16)	(0.27)	(0.33)	(0.37)	(0.56)	(0.73)	(1.36)
Betting Line	1.003***	0.986***	0.941***	0.920***	0.690***	0.921***	0.865***	0.507***
	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.06)
Home		1.704***		2.071***	2.956***	2.090***	2.912***	4.398***
		(0.23)		(0.48)	(0.48)	(0.48)	(0.62)	(0.65)
Opponent Strength					-0.980***			-1.191***
					(0.06)			(0.09)
Week of Season						-0.0345		-0.0809
						(0.04)		(0.05)
Rank in AP Poll (before game)							-0.0858*	-0.299***
							(0.05)	(0.07)
Team Clustered Standard Errors		X						X
Observations	22440	22440	4894	4878	4585	4859	3014	2787
R-squared	0.462	0.464	0.401	0.404	0.45	0.405	0.363	0.414

Robust standard errors in parentheses (*** p<0.01, ** p<0.05, * p<0.1)
Dependent Variable: Margin of Victory

Table A3

Strategy-Based Tests for Efficiency: Conference Tests

Conference	(I)	(II)	(III)	(IV)	(V)	(VI)	(VII)	(VIII)	(IX)	(X)	(XI)	(XII)
ACC	-0.00318		0.00123									
	(0.01)		(0.01)									
Big 10	-0.00442			-0.000109								
	(0.01)			(0.01)								
Big 12	-0.0119				-0.00798							
	(0.02)				(0.02)							
SEC	0.000213					0.00502						
	(0.01)					(0.01)						
MAC	-0.0178							-0.0147				
	(0.01)							(0.01)				
Sunbelt	-0.0228						-0.0187					
	(0.03)						(0.03)					
MWC	0.000571	0.00499										
	(0.03)	(0.02)										
WAC	0.00271											0.00774
	(0.01)											(0.01)
Big East	0.000299										0.00488	
	(0.02)										(0.02)	
Big 8	6.19E-05											
	(0.02)											
CUSA	-0.0136								-0.00963			
	(0.02)								(0.02)			
Independent	-0.00818											
	(0.01)											
SWC	-0.004									0.000344		
	(0.02)									(0.02)		
Observations	22370	22370	22370	22370	22370	22370	22370	22370	22370	22370	22370	22370
Psuedo R-squared	0.00011	0.00000	0.00000	0.00000	0.00001	0.00001	0.00001	0.00005	0.00001	0.00000	0.00000	0.00001

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

The dependent variable is $Y = 1$ if team i beat the spread and 0 otherwise.

Table A4

Sensitivity Checks

	(I)	(II)	(III)	(IV)	(V)
Home	0.0157	0.02	0.0175	0.0216	0.0201
	(0.0151)	(0.0158)	(0.0200)	(0.0159)	(0.0201)
Favorite	-0.0685***	-0.183***	-0.177***	-0.183***	-0.178***
	(0.0166)	(0.0187)	(0.0281)	(0.0187)	(0.0281)
Opponent Strength		-0.0190***	-0.0163***	-0.0193***	-0.0164***
		(0.0017)	(0.0021)	(0.0017)	(0.0021)
Week of Season			0.0000869		-0.00943
			(0.0018)		(0.0160)
AP Rank			-0.00149		-0.00161
			(0.0015)		(0.0015)
Year fixed effects?	X	X	X		
Week of Season fixed effects?				X	X
Observations	4697	4429	2691	4428	2691
Psuedo R-Squared	0.00431	0.027	0.0225	0.0264	0.0203

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

The dependent variable is Y = 1 if team i beat the spread and 0 otherwise.