# Examining Asset Pricing Behavior through the Lens of Sports Betting Markets

By: John High

Advised: Moritz Lenel

Submitted to Princeton University

Department of Economics

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

Date: 4/13/2023

# Table of Contents

1. Introduction:.....3
2. Literature Review: .....8
3. Data:.....12
   3.1. Introduction to Sports Betting Markets:..... 12
   3.2. Discussion of Data Sources**:** ..... 17
4. Methodology:.....25
   4.1. Motivation ..... 25
   4.2. Returns: ..... 28
   4.3. General Sports Betting Contract Price Movements: ..... 34
   4.4. Defining the Characteristic: ..... 36
      A. Momentum.....37
      B. Value:.....40
      C. Size:.....42
      D. Constructing the Strategy.....43
5. Results:.....46
6. Conclusions.....53
7. Pledge: .....55
8. Bibliography .....56
9. Appendix.....58

# 1. Introduction:

Asset pricing theory is a branch of financial economics that studies the fundamental drivers of prices across various asset classes. The topic is a significant source of study for economists looking to rationalize market behavior and explain why asset prices change. As a result, many different beliefs have emerged to explain various pricing phenomena in financial markets best. This paper focuses on the subsection related to modeling particular investor behavior that exhibits irrationality and leads to inefficiency within a market. In such cases, measuring the magnitude of the pricing movement from such irrational behaviors enables one to make claims regarding behavioral characteristics tendencies to cause either an over or under-reaction in price. However, testing behavioral patterns within financial markets is inherently difficult due to several exogenous variables that add additional noise to asset pricing models. Thus, this paper intends to build upon this subset of the asset pricing literature by implementing the use of sports betting markets to test for behavioral tendencies which result in asset pricing anomalies. The structure of sports betting markets and their underlying contracts eliminates many exogenous factors limiting similar studies of financial securities. Thus, it provides a better structure for focusing on specific behavioral characteristics directly affecting asset prices. Hence, this paper aims to examine evidence of behavioral anomalies in sports betting, which can then be applied to similar asset pricing anomalies in financial markets.

A central idea within asset pricing theory that is critical for understanding this paper's motivation for testing pricing anomalies within sports betting markets is Eugene Fama's Efficient Market Hypothesis (1970). The theory argues that the price of any asset is efficient or at fair market value because all market participants have access to all relevant price information. Accordingly, the return of any asset should reflect its underlying risk, thus making it impossible

to beat the market without taking on additional risk. However, this theory often fails to uphold its strict definition. As a result, economists have devoted vast amounts of work to analyzing and explaining financial returns concerning market efficiency.

Furthermore, these studies on market efficiency must also deal with Fama's Joint Hypothesis Problem, which claims that testing for market efficiency is complicated, if not impossible, because any test must use some form of an asset pricing model to predict returns (1970). Thus, it is impossible to say whether the perceived deviation from the predicted returns is because of market inefficiency or an error in the asset pricing model that predicted the returns. This issue of needing to simultaneously prove the inefficiency of the market and the accuracy of the asset pricing model is a central issue that all studies on this topic within financial markets must grapple with. The more exogenous factors a model must incorporate inherently necessitate more assumptions about the experiment's validity. For this reason, asset pricing studies using financial markets are complex because of additional exogenous variables, including the overall market risk premium and behavioral and rational forces which simultaneously affect one another.

Alternatively, economic researchers have motivated using sports betting markets as a cleaner laboratory for empirical asset pricing analysis for two reasons. First, there is no correlation between sports betting and the standard risk of the traditional financial market and the overall economy (Moskowitz, 2021). While specific individuals betting amounts may decline during a recession, a change in the overall money supply, jobs report, or other economic-related news does not impact the skill or performance of the two teams competing in the underlying contract. For this reason, betting lines are completely idiosyncratic from the baseline risk premia in the economy.

Secondly, the sport betting contracts' structure facilitates examining pricing movements' relationship with terminal values. Unlike equities traditionally used in financial studies, all sports betting contracts are fully resolved after the conclusion of the underlying contest. Thus, we can avoid the Joint-Hypothesis dilemma, as a model is unnecessary to predict the asset's terminal value in sports betting markets. Additionally, the actual playing of the game is inherently independent of the activity in the sports betting market. Whether the price moves in favor of the underdog will not reasonably affect the team's performance during the game. This feature is an essential departure from financial markets, in which investor preferences and behavior can result in market hysteria, panic, euphoria, or other reactions that will undoubtedly drive price movements. As a result, a sports betting contract structure featuring short maturity dates with payoffs independent from the underlying market and individual preferences eliminates additional noise present in financial market analysis. This facilitates the paper's ability to map patterns of behavioral irrationality to movement in contract prices since fewer variables are present that could simultaneously affect the price.

While some argue that there have been examples of a fixed sports game or point shaving, which weaken the claim that contests are conducted independently from the sports betting market, these are scarce examples. Despite media dramatization of specific point-shaving incidents and claims of widespread gambling corruption, they fail to hold up for diverse data periods that cover long durations across multiple sports leagues. Even within college basketball, which is most notorious for point-shaving scandals, the conclusion that winning margins in college basketball were connected to corruption was significantly refuted in favor of factors directly related to the game's activity (Bernhardt & Heston, 2010).

While sports betting markets are structured in a manner conducive to tests of asset pricing theory, the two markets also share vital features, which enables this paper to generalize the results from the asset pricing tests back onto similar phenomena in financial markets. First, in examining individual motives for participating in sports betting markets, individuals are more likely to use bets on sports contests for entertainment. However, this has a negligible effect for a few reasons. First, some studies have shown that entertainment motives exist in the stock market, like sports betting. (Dorn and Sengmueller, 2009). Additionally, the rise in low-cost brokerage services also helped facilitate greater access to retail traders. Ultimately, this led to a rise in social media trading groups and, when fueled by conditions specific to a global pandemic, resulted in the Game Stop trading boom, with many WallStreetBets participants motivated mainly by entertainment or other behavioral factors other than financial investment (Werbach, 2021). Regardless of the entertainment component in the stock and sports market, we know that participants are strictly better off when a bet wins, or stock prices rise, suggesting individual preferences are mainly consistent across markets.

Furthermore, sports betting volumes are traditionally dominated by professionals and other sharp investors using this to generate returns. Peta's Trading Bases even goes so far as to discuss the formulation of his 'baseball hedge fund' that utilized many strategies and methods from traditional Wall Street shops (Peta, 2014). Hence, the prevalence of professionals within sports betting markets is equivalent to the role of hedge funds and trading shops within finance that handle market-making and arbitrage to provide market liquidity.

Lastly, the paper's focus of connecting behavioral theories between markets is strengthened by the nature of how these behavior patterns are identified. Barberis finds that patterns of behavioral theories can be generalized according to generic risky decisions (Barberis,

2018). Since no specificity is related to financial markets, the paper can easily apply the same behavior patterns between markets without additional assumptions. Under this framework, a sports bet is another risky decision affected by the same behavioral principles and models one would face when choosing what stock to buy. Ultimately, this enables a clean analysis of behavior models between sports and financial markets while eliminating the presence of risk-based theories as a potential explanatory variable.

This framework of using sports betting markets to test for asset pricing anomalies to make inferences about similar behavior in financial markets is motivated by Moskowitz (2021). His study provides the foundation for this paper's general structure and procedures. However, I look to provide a novel contribution by incorporating updated data encompassing the recent legalization of sports betting markets across 36 U.S. States. The legalization of the sports betting market in 2020 has drastically increased the size of the market, such that an update to Moskowitz's initial findings will provide a more concrete and direct comparison to financial markets. Furthermore, while the same generic methodology principles are used in this study for consistency purposes when comparing data, I leverage my coding scripts to generate various return distributions, asset pricing tests, and portfolio returns based on assumptions discussed in the Methodology section. By using my source code, I could follow the general framework laid out by Moskowitz but with additional modifications that enhance the analysis in certain situations. All code and data files are provided via Git in Appendix 4.

Thus, the paper's primary focus will provide an updated analysis of the behavioral models chosen by Moskowitz, *momentum*, *value*, and *size*. These behavioral characteristics are extensively used as predictors for financial markets, having been found in many studies to link trading strategies based on these behavioral attributes to abnormal returns. However, since the

validity of these findings is weakened due to the reasons previously motivated, this paper analyses whether trading strategies in sports betting markets based on these behavioral characteristics also exhibit asset pricing anomalies. In defining these behavioral predictors within sports markets, I leverage Moskowitz proposed derivations. Momentum will be derived from various statistics on a team's short-term past performance. Value can be defined according to the sports market's perception of a team's cheapness relative to others. Finally, size will encompass a team's overall market value. Using these characteristics, the paper analyses whether price movements in sports betting contracts from market open to close have any relationship to the behavioral models. Moskowitz concluded that there is significant evidence of momentum effects on the pricing movement of sports betting contracts. Additionally, these movements were typically reversed by the contest's outcome, suggesting a model of overreaction at play. Furthermore, value was weak in the pricing movement, whereas size had no impact.

## 2. Literature Review:

Financial economists have long debated the topic of asset pricing theory; however, it is beyond the scope of this paper to evaluate the literature's wide range of predictors used in generating potential financial returns, which include behavioral and risk-based models, such as the Capital Asset Pricing Model (CAPM) that relates systemic risk with expected returns. As previously mentioned, momentum, value, and size indicators have garnered in-depth research on their potential to predict abnormal returns, especially when prices under or overreact to new market information. Thus, my review will first examine how these predictors have been used in financial markets to understand how best to utilize these predictors for analysis in the sports betting market.

Fama and French (2012) examine these three predictors in their study on international stock returns, finding that there are value premiums across international markets that decrease in size. Additionally, their findings reported the presence of momentum in all markets, except Japan, that were decreasing from small to large stocks. However, their testing mentions that the study is plagued by "bad model problems." This is the manifestation of the Joint-Hypothesis Problem, in which the model must make assumptions about predicting prices. Further, the study mentions that the "models may fail, because we do a poor job constructing value and momentum factors or because it is impossible to capture all value and momentum patterns with factors constructed using simple value and momentum sorts." Understanding the importance of this claim is highly relevant throughout this paper. While I do not have to rely on asset pricing models in the study due to the nature of sports betting contracts, detailed assumptions regarding the various behavioral factors are needed when creating the momentum, value, and size characteristics. This ensures that the trading strategy is more robust and, thus, that the results apply to alternative markets as well, unlike the Fama and French study, which by nature of the limitations of asset pricing models, are unable to apply findings outside of the experiment.

One significant subset of empirical studies on momentum as a financial return predictor is the literature on Post-Earnings-Announcement-Drift. This market phenomenon deals with how asset prices respond to the presence of new information in a market, which is also a pattern that I will look to identify in the pricing of sports betting contracts. The literature on Post-Earnings-Announcement-Drift shows that a significant positive or negative earnings surprise leads to abnormal stock returns in the following trading days in the same direction as the surprise. Thus, for a surprise positive earnings announcement, this theory would expect investors to underreact on the day of the news release and the stock to slowly rise in price over the following days as the

market fully absorbs the information. This prediction is based on most Post-Earnings-Announcement studies, which explain the pricing anomaly as a behavioral irrationality caused by investors underreacting to information in the market. However, alternative testing has shown significant evidence of an abnormal reversal in returns, consistent with overreaction bias, when looking at quarterly earnings changes (Bathke, Mason, Morton, 2019). The discrepancy in findings is another empirical example of the limitations of asset price tests in financial markets, which makes it hard to conclude whether risk-based or behavioral models carry more weight. Regardless, the fact that this anomaly is heavily related to theories of over and underreaction in a similar manner to that of sports betting contracts motivates the importance of drawing comparisons between the two markets.

Also, the high-level procedure of constructing trading strategies to test for abnormal returns is consistent with the literature. One such example is a study by Bernard and Thomas (1989), in which they grouped stocks by the size of their respective earnings surprise. They then constructed a portfolio with the groupings in which they went long in stock with positive earnings surprises and short on the ones with negative. This strategy resulted in a statistically significant excess return of 18%. This paper will also construct a similar trading strategy of going long on contracts with a high value of a defined characteristic and short on contracts with low values of the same characteristic. Albeit, by eliminating risk premiums and noise in financial markets, this study's trading strategy's results embody a cleaner framework. This enhances the analysis of behavioral activity, which can then be applied to similar anomalies in financial markets, like Post-Earning-Announcement-Drift.

Previous literature has also leveraged sports betting markets to examine various economic principles that have then been applied to other markets. Some studies focus on the

efficiency of sports betting markets and how it relates to Fama's Efficient Market Hypothesis. In contrast, others look at asset pricing behavior. Chevalier's (1999) study on investor sentiment effect on NFL point spread contract prices helped build the foundations for motivating the effectiveness of sports betting markets for testing asset pricing theories. The study focused on behaviors easily comparable to traditional markets: bets on past winners, expert advice, and team brand recognition. The study did conclude that these behavioral factors did cause a predictable movement in price; however, the study did not determine whether the pricing inefficiency was solely caused by investor sentiment, or a miscalculated opening line set by the Vegas oddsmakers. This study provides an excellent groundwork for sports betting literature that my study leverages to a more extensive dataset, including all four professional sports leagues and various sports betting contracts. Additionally, since the legalization of sports betting across most states in 2020, the efficiency of Sports Markets is much tighter than it was twenty decades ago. Because of the market's increase in size and competition, my study findings can better assume that pricing anomalies stem solely from investor irrationality rather than efficiency.

Moskowitz (2021) contributes significantly to the literature regarding asset pricing theory within sports and financial markets. Like prior researchers, he continues to motivate the claim that asset pricing anomalies are well suited for analysis in the sports betting market because they lack exposure to systemic risk and have terminal values uncorrelated with betting activity. He begins with an analysis on general sports betting contract price movements that are used to define investor behavior in the market. The paper then extends the testing on the financial predictors mentioned above of momentum, value, and size. By constructing well-defined trading strategies to gather returns related to these characteristics in the sports betting market, Moskowitz ultimately concluded that sports contracts demonstrated momentum effects in their

prices but that the transition costs in the form of the sportsbooks' cut or vigorish were too high to offer a profitable strategy.

As mentioned above, the main goal of my paper is to provide an update on Moskowitz's (2021) study using current data, unlike that of Moskowitz, who only examined sports betting contracts up until 2013. This is important because of how the U.S. sports betting market has changed since the overturn of a law in 2018 that had previously outlawed sports gambling outside of Nevada. Since the ruling, 36 states now have sports betting licenses, with many offering an online sportsbook that streamlines the betting process directly onto a user's smartphone. The result of such policies radically changed not only the number of participants in the market but, more importantly, the overall volume. New Jersey, the second largest sports gambling state behind Nevada, has seen an increase in the handle of sports wagers from $16 million in June 2018 to $1.1 billion in January 2023. Furthermore, the total money wagered from 2018 to 2021 was $127 billion, which resulted in nearly $8.9 billion in sportsbook revenue. Additionally, the major players in the U.S. market are now Fanduel and DraftKings, who hold two-thirds of the market together. Hence with vastly different market conditions from 2013, sports betting markets are now even better suited to draw comparisons to financial markets. More maturity and funds suggest a more efficient market, as arbitrage is quickly traded away via more accurate online markets and sharp bettors. As an aside, I also differ from Moskowitz (2021) by using my coding scripts to perform the analysis in this paper. While I will talk through the data and methodology in later sections, a link to the repository with the scripts and data files in various formats, (csv, xtml, Stata, ipynb) is linked in Appendix 4.

# 3. <u>Data:</u>

### *3.1. Introduction to Sports Betting Markets:*

Before giving a detailed account of the data used in this work, I shall provide a brief introduction regarding sports betting markets and the contracts used in this study. At a high level, sports betting markets enable individual agents to buy and sell various contracts which depend on some outcome in the game or season. This study will examine the three most popular and heavily wagered contracts: Point Spread, Moneyline, and Over/Under. To generalize these contracts for any game, let us consider two teams, A and B, whose points scored are denoted as $P_A$, and $P_B$, for A and B, respectively.

The Point Spread then is defined as: $P_A - P_B = X$, with X representing the market's quoted 'Spread' or simply the differential in points between teams. Under this framework, the contract pays out winnings if the team bet upon 'covers' the spread. The term cover is used within sports betting to refer to a team that beats the Point Spread regardless of whether they outright win the game. For example, if the spread between Teams A and B is 5, i.e., $P_A – P_B = 5$, then any wager on the favorite team, A, would only pay off if Team A wins by more than 5 points. In contrast, anyone betting on the underdog would require Team B to win or lose by less than 5 points for the bet to pay off. In this case, if the game concludes with Team A winning by exactly 5 points, neither team has covered the spread, and all bets would be deemed a "push". When a bet pushes, the bettor is returned the principal with no funds lost to the transaction costs. Furthermore, it is a standard procedure that when listing the Point Spread, a negative sign is used to indicate the betting favorite, while a plus sign highlights the underdog. Thus, in the example above, Team A and B would be listed at -5 and +5 respectively.

Additionally, all sports betting contracts are listed with Odds, which are used to determine the payoffs of the contract and are fundamentally linked to the contract's probability of winning. American Odds are formatted with favorites listed as negative numbers (-110, -350),

which means that a better would need to bet the amount listed to win $100. Thus, if a heavy favorite were listed at -1000, one would have to bet $1000 to win $100. Alternatively, underdogs are listed as positive numbers (+100, +350), such that if a bettor will win the amount listed for every $100 bet. Hence betting on an underdog at +350 odds will payout $350 in winnings plus the $100 principal for a $100 bet if the underdog wins.

Hence, since Point Spread contracts are established such that enough points are given to the underdogs to even the playing field between teams, these contracts are designed to create an even probability distribution between winning and losing. Thus, a bet on either team to cover the spread should be equivalent to betting on a fair coin toss. Without market frictions, the odds for both sides of this contract would be listed at +100, implying that a standard $100 bet on either the favorite or underdog to cover the spread would payout $200 when the bet wins, $100 in the case of a push, and $0 when the bet loses, or the team fails to cover. However, the sportsbook charge commission, also known as the vigorish, which is factored into every contract's listed odd. And so, the majority of Point Spread contracts are listed at -110 odds for both the underdog and favorite, which means that to win $100, one must place a $110 dollar bet, with the $10 difference representing the commission. Hence, the payoffs for a standard bet would be: 210 when the team wagered upon "covers" the spread, 110 when the bet results in a push, and 0 if the bet is lost.

Next is the Moneyline contract, which is simply a wager on the team to win the game outright. While the Point Spread contract uses points to balance opposing sides, a Moneyline bet evens the distribution of bettors by adjusting the payout. For example, if team A is favored, then the Moneyline bet will be listed as $-M_A$, implying that a bet of $M_A$ dollars is needed to return $100 in winnings. On the contrary, the underdog will be listed at $M_B$ dollars, which means that a

$100 wager would pay out $M_B$ dollars. Finally, the sportsbook sets the two values such that $M_A$ is greater than $M_B$, with the difference representing the charged commission by the book. Thus, Moneyline contracts prices are simply the odds listed by the book, whereas Point Spread contract prices are given by a combination of Spread and Odds.

This year's past Superbowl provides a good look into how the Point spread and Moneyline bets work in practice. The Philadelphia Eagles are listed as the favorites, hence anyone looking to place a Moneyline wager on the Eagles can do so at -125 odds. Additionally, confident Eagles fans can choose to "lay" the spread and take the Eagles point spread contract at -1.5 for -110 odds, implying that a $110 Point Spread wager on the Eagles would win $100 if the final point differential is greater than 1.5 points in the Eagles favor. Chiefs' supporters then have access to the same point spread contract with the sportsbook allowing them to "take" the points at +1.5 for the same -110 odds. Additionally, if the individual thinks that the Chiefs can win the game outright, they could place a Moneyline wager at +105 odds, implying that a $100 wager will win $105. In this example, the Point Spread has a vigorish of 110-100 = 10, while the Moneyline would be even greater at 125-105 = 20.

Last is the Over/Under contract, which unlike the Spread and Moneyline is not a wager on the contest winner and loser but rather on the combined points scored between opponents. Thus, if the Point Spread is defined as $y = P_A - P_B$, then the Over/Under is similarly defined as $y = P_A + P_B$. The Over/Under contract is traditionally listed by sportsbook as a set number T, implying that those individuals taking the Over need more than T points to receive a payoff, in contrast to those on the Under who need less than T points scored to win. As in the case of the Point Spread, the Over/Under contract is balanced between the two sides such that this is roughly equivalent to a coin toss. Also, we have the same format as the Point Spread with a $110 bet

resulting in a $210 payoff in the case of a win, $110 payoff in the case of a push, and $0 if the bet loses. Again the $10 difference between the amount wagered, $110, and the winnings $100 represents the book vigorish.

For this work, the sports betting data will consist of only these three types of contracts. These submarkets exhibit the greatest volume and quantity of bets across the sports betting industry. Especially since the legalization of sports betting, these markets are structured most similarly to financial markets with significant transaction volume, reactive prices to information, and professional arbitragers. As a result, analysis of behavioral models within this market can be better applied to financial markets when focusing on these three subsections of the sports betting market. It should be noted that since the legalization of sports betting in many states across the United States that there has been a significant rise in Prop betting markets. These are side bets in a contest on much more specific events, including but not limited to the Over/Under on the number of points, yards, or rebounds a particular player will attain in a game, the coin toss or even the Gatorade color dumped on the head coach of the winning team in the Super Bowl. While access to this data may present its own challenges, I see the possibility of future research using these more niche markets to compare pricing activity related to the standard Spread, Moneyline, and Over/Under contracts. Future analysis may examine whether these markets are less liquid and thus more susceptible to mispricing and behavioral theories of over or under-reaction because of their smaller size.

Lastly, in the primer, I would like to briefly discuss how the lines are initially set and adjusted, as this will be a focus when testing price movements in the underlying contracts. While some technology platforms enable agents to trade sports contracts directly with one another, the vast majority of sports betting activity is conducted through Sportsbook entities. Like in financial

markets, where investment banks and other brokers will intermediate trades by offering prices that they will both sell and buy an asset, the Sportsbook acts as the intermediary between bettors by offering both sides prices at which they can enter the various contracts. Thus, when the contracts for a contest are first released, the Sportsbook sets an initial betting line or price with the bookmaker's goal of balancing the odds such that money is placed evenly on both sides. These Sportsbooks employ computer algorithms and mathematical models that factor in various variables, such as team power rankings, weather, and injuries, to best predict the underlying contest outcome and achieve an even balance of betting volume. Thus, an initial line is released to the public incorporating this information. However, betting limits, or amounts that can be wagered upon the contest, are set low to prevent sharp bettors from capitalizing on a potential pricing mistake. Hence, following the opening of the contract, the betting lines will continue to move up and down in reaction to betting volumes. Using the Superbowl example from above, if all the early betting volume came in on the Eagles Point Spread (-1.5), the oddsmakers would adjust the line upwards to -2 and then -2.5 as they attempt to balance out the two sides. In general, Sportsbook will act this way to mitigate risk as they attempt to rake in the vigorish without exposing themselves to excessive risk. Additionally, if new information about the game's potential outcome is released, such as a sports injury, the betting line can move without being affected by volume. A deeper dive into the movement of contract prices with be discussed in the methodology section, as this will motivate the testing procedure used in the study.

### *3.2. Discussion of Data Sources:*

The sports betting data used for this paper's first analysis was collected from OddsWarehouse.com. OddsWarehouse.com provides historical data on sports betting contracts for the NFL, NBA, NHL, and MLB from 2006-2023. The dataset also includes each contract's

basic game and team information (Home/Away Team, Date, Points Scored). While one of Moskowitz's (2021) sources, Covers.com, extends further back decades than this dataset, this only encompassed Point Spread contracts. Most of his data was sourced from SportsInsights.com, which began in 2005. Thus, my data sufficiently overlaps with the time frame that Moskowitz used to develop his cross-sectional variance. Most importantly, the OddsWarehouse data encompasses all four leagues in their most recently played seasons. This ensures that a large portion of my dataset incorporates the recent legalization of sports betting in U.S. states.

Specifically, the dataset contains opening and closing prices for the Point Spread, Moneyline, and Over/Under contracts. The NFL is the only league with data on all three contracts. Unlike the NFL and NBA, in which contests are higher scoring and thus often separated by several points, MLB and NHL games' lower scoring nature results in much tighter point differentials. For example, a single possession in an NFL game can result in 8 points; while scoring 8 goals or runs throughout an entire NHL or MLB is rare. Thus, the Point Spread in the MLB is always set at -1.5 points for the favorite, while NHL Point Spread contracts are either placed at -0.5 or -1.5. As a result, Moskowitz (2021) decided to drop the Point Spread from his NHL and MLB, citing it as a secondary market with litter cross-sectional variance. For this reason, although OddsWarehouse does not include the Point Spread contract data for the NHL and MLB, I see no need to acquire supplemental data for these markets. However, the OddsWarehouse NBA data only provides opening and closing lines for the Point Spread and Over/Under contracts. There is no Moneyline data, which Moskowitz (2021) did include. Since this paper contains Moneyline data for the other three leagues, I again do not see a need to augment the dataset with NBA Moneyline values. To supplement the sportsbetting odds and

prices OddsWarehouse also includes basic team information, home field, and game outcomes for both the regular season and playoff games.

OddsWarehouse.com sources its data from several of the largest sportsbooks in the United States and Europe, which helps to eliminate potential outliers. This follows the same procedure as Moskowitz (2021), in which he utilized the average of all sportsbooks' lines when they differed. Because this study is more interested in looking at potential pricing anomalies due to behavior rather than a test of market efficiency, the focus is on the prevailing line to the public. Additionally, with the increased role of electronic trading and modeling, betting lines for the major markets, Point Spread, Moneyline, and Over/Under, almost move identically across the various books (Moskowitz, 2021). However, as online sports gambling has continued to grow with the increased legalization in the United States, there has been growing literature on "Line Shopping," in which bettors look for discrepancies in sports betting lines in small and relatively illiquid prop betting markets. One prominent platform, OddsJam, offers a subscription of positive expected value bets derived from comparing various Sports Book lines amongst one another and going long on the contracts in which a given price exceeds the fair value odds provided by the market. I should note that these opportunities are almost exclusively found in specific Prop markets (Kevin Durant Over/Under 2.5 Rebounds) and do not present themselves in the large Point Spread, Moneyline, and Over/Under contracts that this paper examines. A potential follow-up study could look at these relatively illiquid markets and analyze whether Momentum, Value, and Size behavioral predictors have more significant effects than larger markets. Nevertheless, for this paper, the OddsWarehouse.com data follows the same general collection methods as Moskowitz (2021), ensuring that comparisons between studies are on the

same basis, yet with updated data from 2013-2023 that will factor in the recent maturation and scale of the market.

Table 1 summarizes the data on the betting contracts from OddsWarehouse, with the mean, standard deviation, and distribution broken into the 1st, 10th, 25th, 50th, 75th, 90th, and 99th percentiles. This table reflects clean datasets from executing basic Stata Do-files. I removed contracts in which the underlying game was delayed or postponed, and outliers created by misentering data fields. Note that the specific lines provided are the closing prices from the home team's perspective, which validates why the mean for the Moneyline and Point Spread contracts are negative. In general, it is considered an advantage to be the home team; hence Sportsbooks will factor this advantage into the price leading to a home team favorite on average. While the notion of home team advantage is a mainstay in modern sports literature and a component of the betting world, its impact is most undoubtedly variable across the different leagues and stadiums. Whether this variation is accurately incorporated into each contract's final price is hard to determine, but I will consider this potential noise to be accurately priced in for now. Regardless, Table 1 summary statistics are presented in the same manner as Moskowitz's (2021), provided in Table 2 for comparison.

**Table 1:**
**Summary Statistics of Sports Betting Contracts from OddsWareHouse.com**

| | mean | sd | p1 | p10 | p25 | p50 | p75 | p90 | p99 |
|---|---|---|---|---|---|---|---|---|---|
| Panel A: NFL, 2006 — 2023; 4,051 Games; 12,077 Betting Contracts | | | | | | | | | |
| Point Spread | -2.2 | 6.0 | -16.0 | -10.0 | -6.5 | -3.0 | 3.0 | 6.0 | 12.5 |
| Moneyline | -127.0 | 318.8 | -1,150.0 | -431.0 | -260.0 | -155.0 | 129.0 | 214.0 | 500.0 |
| OverUnder | 44.6 | 4.9 | 34.0 | 38.5 | 41.5 | 44.5 | 47.5 | 51.0 | 56.0 |
| Panel B: NBA, 2006 — 2023; 21,392 Games; 42,692 Betting Contracts | | | | | | | | | |
| Point Spread | -2.9 | 6.4 | -15.5 | -11.0 | -7.5 | -4.0 | 2.5 | 6.0 | 11.0 |
| OverUnder | 207.2 | 20.2 | 178.5 | 188.0 | 195.5 | 206.5 | 218.0 | 227.0 | 239.0 |
| Panel C: MLB, 2009 — 2022; 34,018 Games; 67,518 Betting Contracts | | | | | | | | | |
| Moneyline | -69.5 | 135.6 | -296.0 | -195.0 | -158.0 | -124.0 | 105.0 | 135.0 | 200.0 |
| OverUnder | 8.4 | 1.1 | 6.5 | 7.0 | 7.5 | 8.5 | 9.0 | 9.5 | 11.5 |
| Panel D: NHL, 2008 — 2023; 18,444 Games; 36,888 Betting Contracts | | | | | | | | | |
| Moneyline | -88.3 | 135.3 | -350.0 | -210.0 | -167.0 | -131.0 | -100.0 | 130.0 | 200.0 |
| OverUnder | 5.6 | 0.5 | 5.0 | 5.0 | 5.5 | 5.5 | 6.0 | 6.0 | 6.5 |

Despite varying time frames, the overall summary statistics of the data align with the values Moskowitz (2021) reported. For instance, the mean of the Point Spread, Moneyline, and Over/Under contracts for the NFL in my dataset are -2.2, -127.0, and 44.6, respectively; Moskowitz (2021) reported -2.6, -160.6, and 42.3 for his NFL data. Both statistics agree in terms of sign and only differ slightly in absolute terms. The NBA varies somewhat in the values reported by the Over/Under contract. Moskowitz reported a mean final price of 196.1 for the NBA Over/Under contract from 1999-2013, whereas my NBA data, which incorporates games from 2006-2023, report a mean of 207.2. However, this intuitively makes sense, as basketball has changed rapidly over the past decade as sports analytics have placed a greater winning likelihood on a team's ability to make three-pointers. This phenomenon enabled the success of

the Golden State Warriors, who thrived on the three-point-making ability of Steph Curry to lead them to multiple championships. As a result of higher-scoring NBA games, the Sportsbooks adjust the Over/Under betting lines to reflect these underlying changes. Thus, this demonstrates one example of how an update to Moskowitz's original data may present a new and exciting analysis. In contrast, the MLB and NHL data exhibit very little difference between the two data sets. For both leagues, the differences between the mean and standard deviation reported for the Moneyline and Over/Under contracts are negligible, which makes sense given the nature of scoring for these two sports. Also, because the scoring differential is tighter in hockey and baseball, the sportsbooks price the favorite close to neutral for these two leagues. In my dataset, the average Moneyline spread was -69.7 and -93.6 for the MLB and NHL, respectively, whereas the NFL and NBA reported values of -160.6 and -220.0. This example helps demonstrate the overall variance in the data, which increases the robustness of the analysis compared to historical studies of sports betting markets.

**Table 2:**
**Moskowitz Summary Statistics of Sports Betting Contracts**

| | mean | sd | p1 | p10 | p25 | p50 | p75 | p90 | p99 |
|---|---|---|---|---|---|---|---|---|---|
| Panel A: NFL, 1985 — 2013; 7,035 Games; 10,775 Betting Contracts | | | | | | | | | |
| Point Spread | -2.6 | 6.0 | -15.5 | -10.0 | -7.0 | -3.0 | 2.5 | 5.5 | 11.5 |
| Moneyline | -160.6 | 208.0 | -700.0 | -370.0 | -270.0 | -174.0 | -112.0 | 144.0 | 264.0 |
| OverUnder | 42.3 | 4.8 | 32.5 | 36.5 | 38.5 | 42.5 | 45.5 | 48.0 | 54.5 |
| Panel B: NBA, 1999 — 2013; 18,681 Games; 38,939 Betting Contracts | | | | | | | | | |
| Point Spread | -3.4 | 6.0 | -15.0 | -10.5 | -7.5 | -4.5 | 1.5 | 5.0 | 10.0 |
| Moneyline | -220.0 | 438.8 | -2200.0 | -565.0 | -315.0 | -172.0 | 107.0 | 177.0 | 333.0 |
| OverUnder | 196.1 | 11.4 | 172.0 | 182.5 | 188.0 | 195.0 | 203.5 | 211.0 | 226.0 |
| Panel C: MLB, 2005 — 2013; 23,986 Games; 47,964 Betting Contracts | | | | | | | | | |
| Point Spread | -1.5 | 0.0 | -1.5 | -1.5 | -1.5 | -1.5 | -1.5 | -1.5 | -1.5 |
| Moneyline | -69.7 | 128.9 | -265.0 | -187.0 | -154.0 | -124.0 | 104.0 | 130.0 | 173.0 |
| OverUnder | 8.7 | 1.1 | 6.5 | 7.5 | 8.0 | 9.0 | 9.5 | 10.0 | 11.5 |
| Panel D: NHL, 2005 — 2013; 9,890 Games; 19,764 Betting Contracts | | | | | | | | | |
| Point Spread | -1.5 | 0.0 | -1.5 | -1.5 | -1.5 | -1.5 | -1.5 | -1.5 | -1.5 |
| Moneyline | -93.6 | 120.7 | -280.0 | -201.0 | -165.0 | -133.0 | -105.0 | 120.0 | 158.4 |
| OverUnder | 5.6 | 0.4 | 5.0 | 5.0 | 5.5 | 5.5 | 6.0 | 6.0 | 6.5 |

I gather additional sports and team data to track various aspects of a team concerning Momentum, Value, and Size predictors. I begin by leveraging the extensive database of Rodney Fort, a prominent figure in the sports and economics industries. Fort has extensively researched the relationship between a sports team's revenue, payroll, and location with actual on-field performance. In doing so, he continues to upload and maintain a google drive compiled with information from various sources. Thus, for annual MLB team valuations and revenue, I use Fort's drive, an aggregate of data provided by Forbes.com's annual listing of MLB team values,

revenue, and operating income. I clean the data within Excel using a variety of functions and format the team string values to be compatible with the OddsWarehouse MLB dataset for 2009-2022. This is important because this dataset will be used to derive measures of each team's *size*, as defined in the following section.

Additionally, I compiled Payroll data for the NBA and MLB; however, to maintain consistency across the entire period of sports betting contracts in the OddsWarehouse database, I used external sources linked from Fort's google drive since his data did not encompass the past few years. Since the past two years are the most important in providing new analysis on sports betting markets, sourcing consistent data for the entire period is of paramount concern. Thus, for the NBA, I used the salaries database from *hoopshype.com* to gather each team's annual payroll from 2006-2022. Likewise, for the MLB, I used *thebaseballcube.com* for team payroll data from 2009-2022. These values are based on the opening-day salaries of all rostered players.

Payroll data is only gathered for the MLB and NBA leagues since these two leagues do not enforce hard-cap salary restrictions. Unlike the NHL and NFL, which enforce strict salary cap levels that result in nearly equal team payrolls, the MLB and NBA enable teams to exceed the target salary cap. This results in a more significant discrepancy in the payroll of teams, a key measure I use to map *value* to an individual contract. The NBA currently operates under a very complex soft cap system, in which by leveraging many possible exceptions, teams can exceed the 'salary cap.' For example, under the Larry Bird Exception, a team may exceed the salary cap if they re-sign a current three-year consecutive player who is also a free agent. While there are many others, the details of every exception are irrelevant. What matters is that teams that exceed the cap must pay a luxury tax, which is then redistributed to both teams below the cap and the league. The MLB also uses a luxury tax system that fines teams exceeding the cap. This fine is

determined as the percentage of the excess amount, which also increases for teams who happen to be consecutive offenders. As a result, the MLB exhibits by far the broadest range in team payrolls of the four professional sports leagues. For example, in 2022, the L.A. Dodgers recorded a league-high payroll of ~$285.5 million, far exceeding the ~$230 million target level. In contrast, the league low was held by the Oakland Athletics at $32.5 million. This discrepancy motivates the use of the MLB and NBA for evaluating the significance of the value characteristics, unlike the NHL and NFL, whose payroll structure does not present similar opportunities for analysis.

Lastly, I ensure that teams with changed names or locations are treated as one entity for all data sources. While a team may change stadium, name, or logo, the underlying past performance of the team remains and thus will remain a factor in the pricing of the franchise's underlying sports betting contract regardless of its current identity. This guarantees that all teams have a thorough database of past data for contracts in the most recent years, which is necessary to rank contracts according to defined behavioral characteristics derived from various team measures. Furthermore, I remove 'all-star' and other one-off contracts from the datasets. While betting lines are provided for these All-Star Games that pit conference players against each other, these only occur once per year and are irrelevant to the study.

# 4. Methodology:

## *4.1.Motivation*

This section of the paper will motivate various Asset Pricing Tests and provide a detailed account of the procedure for generating the analysis. First, I will provide an underlying guide to sports contract pricing movements and lay out a few hypotheses to explain why the factors affect prices. In the sports betting primer, I discussed how prices for the various contracts are

determined by sportsbooks, concluding with a short explanation of the movement in contract prices from the betting market's opening to close. We determined that, like financial asset prices, these sports betting contracts undergo price movements due to rational and irrational factors, which may or may not depend on new information in the market. However, regardless of the movement in price, each contract exhibits a terminal value after the outcome of the game, which can then be used to calculate the contract's return.

Thus, I define a general contract with three defined returned horizons as such:

$P_0$: Opening Price; $P_1$: Closing Price; $P_T$: Game Outcome

$$R_{open:close} = \frac{P_1 - P_0}{P_0};\ R_{close:end} = \frac{P_T - P_1}{P_1};\ R_{open:end} = \frac{P_T - P_0}{P_0}$$

In this framework, $P_0$ and $P_1$ are simply the opening and closing prices provided by the Sportsbook, while $P_T$ represents the payout of the contract determined by the outcome of the game. Now, since the goal of this paper is to examine the behavioral factors affecting sports betting contract prices, I first look to see if general movements in prices have any significative predictive ability on terminal values. We can test this by examining the relationship between $R_{open:close}$ and $R_{close:end}$. or more formally by running the general regression for each contract *i* as:

$$R_{i,close:end} = \alpha + \beta_1 R_{i,open:close} + \epsilon_i \qquad (1)$$

Based on this equation, a few predictions depending on the size and magnitude of the beta coefficient emerge. First, consider that price movements from the Open-to-Close are due to new information in which all market agents rationally react to the news. Thus, those who placed a wager at the open are disadvantaged to those who placed a wager at the close once the new information was priced in. In this situation, because all information is rationally incorporated into the price, we know that the expectation of the final price $P_T$ should equal the price at the

close of betting $P_1$. While the price has changed moved $P_0$ to $P_1$, this change does not result in the ability to predict Rclose:end better. Thus, when prices change due to rational responses, we expect that $\beta_1 = 0$.

Alternatively, let us examine market conditions where prices move for purely non-information reasons. Here, any contract price change is driven entirely by irrational factors independent of any relevant information to the underlying contest. In this state of the world, those placing a wager at betting close are disadvantaged compared to those who placed a wager at the open. Hence, the price at the close will be a worse predictor of the game's outcomes because it incorporates noise. In the long run, the conclusion of the game should result in a final price $P_T$ that results in a reversion back to the truth since the playing out of a sports contest is independent of market activity. This highlights the usefulness of sports betting markets as studies using financial assets would have to make assumptions about how the model incorporates market activity into terminal values. Ultimately, we predict that in this state of the world, $\beta_1 = -1$ as Ropen:close will be inversely related to Rclose:end upon completing the sports contest.

Finally, we examine the scenario in which prices may move due to informational reasons, but due to behavioral activity, the price does not entirely reflect the new information. Many tests of the Efficient Market Hypothesis deal in some manner with this phenomenon, often leading to claims of over/underreaction behavioral models in markets. Like the second prediction, the price at the close also incorporates some noise. Again, because the outcome of a sports contest is independent of market activity, we expect that the final price will only incorporate the game-specific price information and not the noise. Thus, there will be some measure of predictability on the Close-to-End return by the movement in price or Open-to-Close return. When agents overreact to new market news, we predict the reversion will be negative as the price comes back

down by the game's conclusion. Formally, we expect $\beta_1 < 0$ in this state of the world. Alternatively, if investors underreact to the presence of new information, we expect a positive return prediction as the closing price increases to its efficient price, $P_T$. In this setting, we formalize the prediction that $\beta_1 > 0$. Overall, general price movements in sports betting contracts are summarized under the following predictions related to equation 1:

*Prediction 1: When prices move according to information and markets respond rationally then $\beta_1 = 0$*

*Prediction 2: When prices move according to non-information reasons then $\beta_1 = -1$*

*Prediction 3: When prices move according to information and markets respond irrationally then $\beta_1 > 0$ demonstrates an underreaction while $\beta_1 < 0$ demonstrates an overreaction.*

## *4.2.Returns:*

I will briefly define the framework for gathering the return distributions used for the regression in Equation 1. For this paper, I calculate actual betting returns that a better would realize on the market in real-time. I do this to compare the Moskowitz study with my updated data that incorporates the recent legalization of betting activity. Other studies have examined the changes in sports betting lines from the open to close as a potential indicator of a game's outcome; however, the data used in prior studies did not represent actual betting prices across a diverse range of contracts and sports. Since the end goal of this paper is to draw comparisons regarding behavioral models to financial markets, the use of real-time returns enables a direct comparison between the trading strategies constructed in this paper with those developed for financial markets. Unlike point returns that are only specific to sports betting markets, actual returns also better encompass the probabilistic distribution of the game's outcome. Hence, I can more easily distinguish examples of contract mispricing as a byproduct of the behavioral predictors in my testing.

Thus, to calculate actual returns, I first compute each contract's payouts using the betting lines/prices from the Open and Close. I normalize the investment in each contract or stake to a standard wager of $100. I created a Stata Do-File script to determine the winner of the Spread, Moneyline, and Over/Under contracts for every game in the dataset. I then leverage the following two equations to calculate the Payouts for these winners based on whether the odds for the contract were given as an underdog or favorite. Payouts for bettors on the losing side of the wager are simply $0, while a push results in a payout of $100 for both sides.

$$\text{Underdogs: } \frac{Odds}{100} * S\text{take} + \text{Stake} = \text{Payout}$$

$$\text{Favorite: } \frac{100}{-1 * Odds} * S\text{take} + \text{Stake} = \text{Payout}$$

Note that the Odds must be multiplied by negative ones for favorites because American Odds are given as -XXX in the dataset. With these payouts, the return is simply the percentage change from the initial investment in the contract to the payout. Thus, using the opening line for each league across the Point Spread, Moneyline, and Over/Under, the return from the Open-to-End, $R_{i,\ open:end}$ represents the return that a better receives from placing a wager on the opening line for contract *i*. Similarly, the return from the Close-to-End, $R_{i,\ close:end}$ would be the return from a wager on the same contract, *i*, but placed at the closing line. Using these two returns, I calculate the return from Open-to-Close as the difference between the Open-to-End return with the Close-to-End return. In practice, this represents the actual return an agent would receive if they placed a wager on the opening line and then placed a wager on the opposite side at the close to unwind the position. I now have returns on every contract needed to run regression 1 defined above.

I want to discuss the limitations of this calculation though briefly. For Moneyline wagers, in which the odds given translate directly into the contract's current price, it makes sense to calculate the Open-to-Close return in this manner. However, for Point Spread and Over/Under contracts because the prices are listed in terms of points, which are dependent on the game outcome and do not translate directly into financial prices, the Open-to-Close return may only partially demonstrate price activity in the market. I will use an example to motivate this.

Consider an NFL game between the New England Patriots and Buffalo Bills. The Point Spread market opens by listing the Patriots as a 5.5-point home favorite against the Bills at +100 odds (assume no vigorish in the example). The market closes with the Patriots as a 7.5-point favorite at +100 odds. The game concludes with the Patriots winning by 8 points. In this example, the return from Open-to-End would be 1 since the Patriots won by more than 5.5 points, resulting in a payout of $200 for a $100 wager. Additionally, the Close-to-End return if betting on the Patriots would be 1 since the final point differential of 8 exceeded the closing spread of 7.5 points. Thus, the return from the Open-to-Close in this scenario would be 0. We can compute this in two ways.

(1) $R_{open:close} = R_{open:end} - R_{close:end} = 1-1 = 0$

(2) An agent places a wager on the Patriots at -5.5 points and then unwinds this position at the close by taking the Bills at +7.5 points. The Patriots covered the first bet paying out $200; however, since the Bills did not cover, the second bet would return $0. Thus, the combined strategy would have invested $200 and returned $200 for a return of 0%.

However, we can see that the contract price from the Open-to-Close increased by two points. Thus, there is an implied greater market value of the 5.5 contracts, given that the closing

market price placed another 2-point premium onto the Patriots. In financial markets, one would be able to trade out of this asset and collect a return from the increase in market value; yet because of the Sportsbooks structure, a bettor cannot enjoy this same benefit for Point Spread markets. Upon examining the Moneyline contract on this game, we see that the Moneyline price for the Patriots would increase such that the return from the Open-to-End would be greater than the return from the Close-to-End (For a 5.5 spread, the Moneyline odds are placed at -245, while a 7.5 spread has Moneyline odds of -345). Thus, calculating the returns of this contract gives the following if betting on the Patriots:

$$\mathrm{R}_{\mathrm{open:end}} = \frac{100}{-245} * 100 + 100 = 140.82 \Rightarrow 40.82\%$$

$$R_{close:end} = \frac{100}{-345} * \ 100 + 100 = 128.99 \Rightarrow 28.99\%$$

$$R_{open:close} = 40.82\% - 28.99\% = 11.83\%$$

As a result, the return from the Open-to-Close for the Moneyline is slightly positive, which is reflective of the movement in price. For this reason, I consider the Moneyline returns the most robust test but keep the Over/Under and Point Spread contracts to maintain a diverse cross-section of data, which can be compared to Moskowitz's findings.

I implement a final procedure to obtain gross and net return distributions. Gross returns enable the construction of trading strategies without transaction costs, which are much higher for sports betting markets. This ensures that comparisons between markets are on similar grounds regardless of market frictions. Thus, to strip away the sportsbook's transaction fees, or vigorish, from the prices of each contract, I must first calculate each contracts underlying implied probability of success according to the following equations.

$$\text{Implied Probability for Favorite Odds} = \frac{1}{\frac{\text{Odds}}{100} + 1}$$

$$\text{Implied Probability for Underdog Odds} = \frac{1}{\frac{100}{\text{Odds}} + 1}$$

For Moneyline contracts, I calculate the implied probability of both a favorite and an underdog. In contrast, for Point Spread and Over/Under contracts, since these are designed to balance the proportion of bets placed on the contest, both sides are traditionally listed as favorite odds of -110, as explained earlier. For these contracts, only the favorite odds calculation is necessary. Given each outcome's implied probability, I then calculate the true probability that the respective outcome occurs through the equation below. With the true probabilities, I then update the American Odds in the dataset to reflect both outcome's real market price excluding frictions.

$$\text{A True Prob.} = \frac{\text{A Implied Prob}}{\text{A Implied Prob} + \text{B Implied Prob}}$$

Table 3 provides net return distributions for each contract type within the four professional sports leagues. Returns for the Point Spread are calculated from the home team's perspective, while the Over/Under lists returns from betting solely on the over. However, both home and away returns are included for Moneyline contracts since these contracts encompass inherent leverage. Consider the case of an extreme underdog. A Moneyline bet on the underdog to win outright has the chance to return a high multiple of invested capital, albeit at a low probability. As a result, the standard deviation is universally much more significant for Moneyline returns.

In general, the return distribution provides the baseline for analyzing and testing asset price movements related to the regression and prediction laid out above. We can see that

universally the mean return distributions for each league are negative, albeit more so from the Close-to-End than the Open-to-Close. Because the Open-to-Close returns are calculated from the difference between the Open-to-End and Close-to-End returns, and all the Close-to-End returns are negative, we know that the Open-to-End mean return is the most negative. This implies that there may be some pricing inefficiency when the initial betting prices are released. More importantly, since there is evidence that the return distributions move from the Open-to-Close, we know there is some pricing movement. We can now run the regression leveraging the returns to test whether the movement in price exhibits rational behavior.

**Table 3:**
**Net Return Distributions**

| | Point Spread | | Over/Under | | Moneyline | |
|---|---|---|---|---|---|---|
| | Open-to-Close | Close-to-End | Open-to-Close | Close-to-End | Open-to-Close | Close-to-End |
| | | | NBA | | | |
| Mean | -1.8% | -5.3% | -1.4% | -4.0% | | |
| Stdev | 35.0% | 95.6% | 37.7% | 95.9% | | |
| | | | NFL | | | |
| Mean | -2.7% | -7.2% | -1.8% | -5.5% | -0.6% | -3.4% |
| Stdev | 43.5% | 96.2% | 35.0% | 96.6% | 22.5% | 115.3% |
| | | | MLB | | | |
| Mean | | | -1.7% | -5.0% | -0.3% | -2.1% |
| Stdev | | | 26.0% | 95.1% | 9.5% | 101.6% |
| | | | NHL | | | |
| Mean | | | -1.2% | -2.6% | -0.4% | -2.7% |
| Stdev | | | 16.8% | 95.6% | 8.9% | 101.8% |

I also include gross return distributions in Table 4 below. This demonstrates that the mean returns are less negative across the board when removing the transaction costs from the

contracts. Interestingly, the Moneyline contract for the MLB and NHL reports an average return of 0.0%, implying that these are efficient markets. This makes logical sense, considering that MLB and NHL contracts tend to be more closely decided in points and thus will contain less noise in the contract's pricing. However, because other leagues still report a mean negative return, there is some inefficiency in the pricing of contracts that I will analyze first for general sports betting contracts and then again when constructing strategies based on behavioral characteristics.

**Table 4:**
**Gross Return Distributions**

| | Point Spread | | Over/Under | | Moneyline | |
|---|---|---|---|---|---|---|
| | Open-to-Close | Close-to-End | Open-to-Close | Close-to-End | Open-to-Close | Close-to-End |
| | | | NBA | | | |
| Mean | -0.9% | -1.7% | -0.5% | -0.5% | | |
| Stdev | 36.3% | 98.9% | 39.2% | 99.3% | | |
| | | | NFL | | | |
| Mean | -1.3% | -4.5% | -0.4% | -2.5% | 0.3% | -0.8% |
| Stdev | 45.0% | 99.0% | 36.2% | 99.6% | 17.6% | 117.6% |
| | | | MLB | | | |
| Mean | | | -0.3% | -2.2% | 0.0% | 0.0% |
| Stdev | | | 26.9% | 97.9% | 9.5% | 103.7% |
| | | | NHL | | | |
| Mean | | | -0.9% | 0.1% | 0.0% | 0.0% |
| Stdev | | | 17.0% | 98.4% | 8.9% | 104.5% |

### *4.3. General Sports Betting Contract Price Movements:*

Before defining the three behavior characteristics, I will perform a baseline analysis on sports betting contract price movements according to Equation 1. Using the Net Returns

provided in Table 3, I can now estimate the Close-to-End returns using the calculated Open-to-Close returns to test the predictions for general pricing movements. For the Point Spread and Over/Under, only the Home and Over contract returns are included since the payoffs are distributed evenly for opposing sides. Contrarily, Moneyline returns incorporate both the return for the Home and Away team. Lastly, the returns are estimated using the Net Returns provided earlier, which incorporate the transaction costs. Because these contracts are structured as straight bets that depend only on a single outcome, the vigorish is applied once. Thus, the regression findings do not differ when using gross vs. net returns. Future analysis may examine whether this is consistent for Parlay wagers, which include a combination of multiple contracts, thus increasing the net transaction cost for the final contract as the vigorish is applied many times.

**Table 5:**
**Regression Results on General Sports Betting Contract Return Distributions**

| | Point Spread | Moneyline | OverUnder |
|---|---|---|---|
| | All Sports | | |
| $\beta_1$ | -0.579 | -1.192 | -0.662 |
| *t*-stat | (-39.96) | (-35.42) | (-58.44) |
| Observations | 29,439 | 104,365 | 81,629 |
| | NFL | | |
| $\beta_1$ | -0.598 | -0.883 | -0.651 |
| *t*-stat | (-17.86) | (-15.75) | (-15.41) |
| Observations | 4,041 | 8,080 | 4,041 |
| | NBA | | |
| $\beta_1$ | -0.568 | | -0.564 |
| *t*-stat | (-31.06) | | (-33.25) |
| Observations | 21,357 | | 21,357 |
| | MLB | | |
| $\beta_1$ | | -1.117 | -0.701 |
| *t*-stat | | (-27.35) | (-35.88) |
| Observations | | 67,479 | 33,746 |
| | NHL | | |
| $\beta_1$ | | -1.353 | -1.086 |
| *t*-stat | | (-22.75) | (-26.45) |
| Observations | | 36,886 | 18,444 |

Most importantly, Table 5 demonstrates a consistent negative coefficient for β1, providing evidence of a reversal in contract pricing from the Open-to-Close. Given the provided t-stats, these coefficients are all statistically significant, and thus we can eliminate both Predictions 1 and 2, as the coefficients reject being equal to either 0 or 1. This dictates that the market does not act solely on rational or irrational factors. Instead, the finding most aligns with the subset of Prediction 3 related to overreaction models. This is because when agents overreact to news in the market, we expect β1 to be negative, as the closing price will reverse to reflect the fundamental value of the contract more accurately. In support of this prediction, the coefficient for the Point Spread contracts reports a value of -.579, implying that for every dollar change in the Open-to-Close period, .58 cents of value are reversed when the contract pays out following the conclusion of the game. The Moneyline contracts results also report a negative but with a more significant magnitude coefficient of -1.19 for the entire dataset. This is interesting as it suggests that the initial pricing of the contract may not be fully efficient. Not only is the Open-to-Close price movement fully reversed in this scenario, but the price reverts more than the movement altogether. Finally, the Over/Under is consistent with the Point Spread contracts.

### *4.4.Defining the Characteristic:*

Having defined general price movements for sports betting contracts, I now develop the framework related to the three behavioral characteristics, momentum, value, and size. I will then construct trading strategies to test for return predictability and price changes in sports betting contracts related to each characteristic. I will motivate and leverage various principles from Moskowitz (2021) to achieve this; however, all produced code and data are of my own creation, which can be referenced in the Appendix. I form trading strategies at a high level by calculating

a weighted index value to represent every contract's measure of the tested characteristics, Momentum, Value, and Size.

*A. Momentum*

I define the momentum characteristic to be consistent with financial market momentum predictors by calculating three past performance indicators to find a team's current momentum characteristic value. Specifically, I look at a team's winning percentage, point differential, and the percentage of games that the team has covered in the point spread. To complete this task, I converted the sports betting contracts and game information data into a Python Pandas dataframe for each respective league. Concurrently, I define a dictionary with each unique team name representing a key. The dictionary values for each team are then broken into four separate dictionaries with the keys, 'wins', 'spread', 'returns', and 'games'. The values for 'wins', 'spread', and 'returns' are empty lists that will function as First-In-First-Out Queues. The value for 'games' is simply an integer. Thus, as I iterate through each contract, I identify the two teams in the underlying contest and populate the result of the game, each team's point differential, and the betting results (which team successfully covered the spread) into the dictionary. The values for wins and betting returns are zeroes and ones to indicate either success or failure for the two criteria. In contrast, the point differential value is the *away team score – home team* score or vice versa when evaluating from the perspective of the Home Team. Finally, the value for games is an integer that increments by one.

The dictionary only contains these game statistics for a team's most recent N games, with N varying between leagues to accommodate the number of games in a season. For the NFL, which has the shortest number of games played, I choose a lag of 3 games to represent a team's past performance for roughly one month, given that bye weeks are included in a season. For the

NBA and NHL, I increase this to 4 games, which represents roughly 5% of each season and matches the analysis that Moskowitz performed. Finally, I choose nine games for the much longer MLB season as this represents three complete series. Generally, a baseball team will perform three contests in a row against an opponent to form a series. This matches with three game rotations used for the NFL and is roughly 5% of the 162-game baseball schedule. For all leagues, I do not include games in the previous year when looking at the past N games. While financial literature differs on a consensus for a momentum horizon, time horizons of less than a year are generally agreed upon. Additionally, because sports betting contracts have short maturity dates, it is reasonable to assume that momentum measures would be shorter than financial markets (Moskowitz, 2021). Contracts with less than N games are also dropped when looking at the returns for the final strategy since the momentum measures have not been fully populated yet. Thus, to show an example, in the NBA dataset, with N = 4, the dictionary key-value pair may look like such:

Boston Celtics: {‘Wins’: [0,1,1,0], ‘Spread’: [5, -3, 13, 2], ‘Returns’: [1, 1, 0, 0], ‘Games’: 6}

Thus, as I iterate through the data frame, for each contract *i*, I use the populated dictionary of the team’s recent game performance to calculate each contract’s underlying Home and Away teams’ winning percentage, total spread differential, and percentage of games covered over the past N games. This is executed by iterating through the length of each list in the dictionary of team values and summing the individual values. I then pop out the statistics from each team’s least recent game in the dictionary data structure and append the current contract *i’s* outcome (winner, spread, betting winner) to the dictionary so that the values can be used when each respective team’s next contract arises.

Once this iteration is complete, I have each team's specific momentum measures for every sports betting contract in the dataset. At a high level, I used the dictionary data structure to create three additional vectors in the Sports Betting dataframe. These hold Momentum values of Wins, Point Differential, and Betting Returns for each team at time *t* for contract *i.* To equally weigh the Wins, Point Differential, and Betting Returns, I normalize each measure to have a mean zero with unit variance. I then compute the team's composite Momentum characteristic for each contract *j* as the sum of the scaled measures.

Lastly, with the team's Momentum characteristic values, I calculate the actual Momentum *Char* for each contract *i*. When we consider Point Spread and Moneyline contracts, these are inherently bets on a team's performance. Thus, in defining a momentum characteristic for a Point Spread or Moneyline contract, a contract exhibiting a high Momentum value would occur when the favorite has recently performed well, and the underdog has performed poorly in the same period. Hence, for every game's Point Spread and Moneyline contracts, the Momentum Index Value (Char) used to formulate a trading strategy is calculated by taking the difference between the favorite and underdog team's Momentum characteristics. Contrarily, because the Over/Under contract depends on the point performance of both teams, the Momentum Index Value for these contracts is calculated by adding the two team's Momentum measures. I should note that while winning percentage and betting performance make logical sense for tracking Momentum in Point Spread and Moneyline, as these contracts depend directly on a team's playing performance, their ability to measure Momentum for Over/Under contracts is undoubtedly weaker (Moskowitz, 2021). The Over/Under contract is more dependent on a team's ability to score points than a team's ability to win; thus, measures of a team's winning percentage are only marginally applicable in determining the expected number of points for the underlying

contest. However, to accurate Moskowitz's study, I include the three momentum measures for all contracts with the caveat that the Over/Under values likely exhibit more noise. Thus, under this framework, I have mapped a Momentum Index Value to each type of contract for every contest across all four leagues. These are then used to determine which contracts are included in the trading strategy algorithm defined at the end of this section.

*B. Value:*

In defining value measures to calculate a sports betting contract's underlying exhibit of *value,* I leverage the work of Moskowitz and prior studies on value in financial markets. At a high level, a portfolio focused on value is best thought of going long on cheap assets while concurrently shorting expensive assets (Asness, Moskowitz, Pederson, 2013). Thus, in applying this theory to sports betting contracts, the measures must reasonably apply notions of cheapness or expensiveness to each contract. To perform this task, I leverage the measures laid out by Moskowitz, long-term past performance, contract fundamental-to-market ratio, and team talent-to-market ratio.

Long-term past performance is motivated by financial literature in which Debondt and Thaler tested for overreaction in the stock market and found that trading strategies comprised of long-term prior losers outperformed historical winners. Additionally, the negative of these long-term returns is correlated with other value metrics used in predicting financial returns. As a result, long-term returns should similarly be an accurate metric for determining value in sports betting markets (Asness, Moskowitz, and Pederson, 2013).

Contract fundamental-to-market ratio is similarly motivated by the prevalence of assigning the ratio of an equity's book value of equity to the market value of equity as a stock's

underlying value. Thus, I leverage the famous sports analytics predictor, the Pythagorean Win Expectation Formula, to determine a sports betting contract's fundamental value. This is consistent with Moskowitz's calculations for assigning fundamental value to a contract.

$$\text{Win Expectation} = \frac{(\text{Points Scored})^{\gamma}}{(\text{Points Scored})^{\gamma} + (\text{Points Against})^{\gamma}}$$

In this equation, Points Scored represents the underlying team's average points scored over the desired time horizon, with Points Against as the average points scored on the team for the same period. According to the literature, the exponent, γ, varies for each sport (MLB: 1.83; NBA: 13.91; NFL: 2.37; NHL: 2.11). Thus, I can calculate each team's fundamental win expectation for any contract, given the team's scoring history. I then compute the difference between the home and away teams' expected win probabilities to determine the contract's overall expected value. Then for Point Spread contracts, I multiply this value by the total number of points expected in the contest, which is simply the value of the Over/Under contract. This results in an expected point difference or E(P). Finally, I divide by the actual price listed in the dataset, P, which for Point Spread contracts is the market spread. The result, E(P) / P, represents the Point Spread contract's ratio between the expected fundamental value and the market price P.

Lastly, the team talent-to-market ratio seeks to provide a final measure of cheapness by relating another market fundamental to the price of the actual contract. Moskowitz motivates this through Murphy's (2007) study as such "Although labor markets for athletic talent are not perfectly efficient, they do correlate well with marginal productivity." This implies that we can use a team's payroll to make logical inferences about the quality of a team (In general, higher payroll teams tend to perform better). Thus, a contract in which a high payroll team plays against a low payroll team will offer an investor a good value proposition if the market price is relatively

small or irrationally favors the low payroll team. I calculate the talent-to-market ratio as the difference between the team's payrolls divided by the contract's price or the Point Spread.

To derive these three values measures I implement a similar procedure as expressed in the Momentum section by coding a Python script to assign a Value Index Value to each contract. I iterate through the OddsWarehouse dataframe of contracts to populate a dictionary data structure; however, the dictionary of team information now holds information pertinent to the value measures. Thus, for any given team, I store their points forced over the past 40 games, points scored against over the past 40 games, the one-year cumulative lag of their Point Spread betting returns, and the cumulative Point Spread betting returns in the current season. The first two measures are stored as lists and treated as a FIFO queue, as explained above, while the other two measures are single integer values. In addition, I also converted the data containing the team's Payroll information into a separate dataframe. I then use this dataframe during the iteration to map the difference between the Home and Away teams' payroll to each contract. Thus, after completing this iteration, for every contract, I have the Home and Away teams' average points for and against, the Home team's one-year lag of Point Spread returns, and the Home team's payroll difference. I then use these contract-specific team values to perform various vectorized operations to calculate the three measures of value, negative of one-year returns, fundamental-to-market ratio, and team talent-to-market ratio. From here, I implement the same procedure as the Momentum section by normalizing the measures to have mean zero with unit variance. I then sum up these three variables to create a single equally weighted Value Index mapped onto every contract.

*C. Size:*

I use two measures for the final behavioral predictor, size, annual team market value, and total revenue. These measures represent the overall value of the asset and local market consistent with size predictors used in financial markets. Since these measures apply directly to the team and are not a function of previous game or contract performance, I do not need to perform dictionary operations with Python as an intermediary step. Thus, I converted the annual Team Valuation and Revenue database, sourced from Ford's database and *Forbes.com*, into a Python dataframe to use with the OddsWarehouse dataset. As I iterate through the OddsWarehouse dataframe, I use the current contracts team name to perform a lookup operation in the dataframe containing franchise value and revenue. I then map these values to the OddsWarehouse dataframe in new columns for both the Home and Away team. and normalize these values so both measures have a mean 0 with unit variance. Home and Away team size values are computed as the sum of these two normalized variables. Lastly, consistent with Momentum and Value definitions, I compute each contract's overall Size Index Value as the difference between the Home and Away team's size values for Point Spread and Moneyline contracts, while the Over is the sum of the two teams' size measures.

### *D. Constructing the Strategy*

Having calculated a single value to represent every contract *j's* underlying characteristic, *Momentum, Value, and Size* concerning both the Point Spread / Moneyline contracts and the Over/Under, I now formulate trading strategies by ranking the contracts within each category at every time interval, *t*. With rankings on every contract *j* for various time intervals *t*, I then decide to go long on contracts in the top quintile and short the contracts in the bottom quintile. Algorithmically, I perform the ranking calculation by using the *groupby* function on the dataframe of sports betting contracts and characteristic values. I use

the time interval to create the subgrouping and then use the rank function with the characteristic as the parameter. Since the Point Spread and Moneyline contracts use the difference between the Home and Away team's characteristic values, these two contracts can share the same rankings for deciding which contracts to use in the trading strategy. In contrast, a separate ranking is created for the Over/Under contracts.

To scale the number of wagers according to the number of games played on a given day, I also use the groupby function to count the number of contracts for every time interval. I then use indicator variables to encode whether a specific contest qualifies for going long or short on the Point Spread/Moneyline and then again if the contest qualifies for entering a long or short position for the Over/Under. Using these indicator variables, I assign a weight to each contract based on the following equation:

$$\mathrm{w_i} = \frac{1\left(char_{i,t-1} \epsilon Q5\right)}{\sum_{i=1}^{N} 1\left(char_{i,t-1} \epsilon Q5\right)} - \frac{1\left(char_{i,t-1} \epsilon Q1\right)}{\sum_{i=1}^{N} 1\left(char_{i,t-1} \epsilon Q1\right)}$$

Here, we have assigned a weight for every contract, *i*, at time *t*, given that there are N contracts with characteristic values. Note that the characteristic looks at *t-1* since these predictors are based on records to predict the underlying contest. In essence, this motivated my use of a dictionary data structure to keep track of each team's *t-1* performance while examining contracts played at time *t*.

Additionally, a limitation of this trading strategy is the assumption that we can enter short positions for sports betting contracts as if they were financial market securities. Under this assumption taking a short position while already long on the underlying contract would neutralize the exposed risk. However, in practice, individual bettors cannot access such abilities to short sports betting contracts, given that intermediary sportsbooks broker all contracts. A

better could take the opposite side of a contract (bet on the favorite at the current Point Spread if they are currently on the underdog); however, we know that this strategy fails due to high friction costs from the sportsbook vigorish. Thus, for this reason, because I only want to evaluate the feasibility of these trading strategies to evaluate asset pricing behavior, I allow for the theoretical shorting of contracts. Additionally, all returns are calculated from a gross perspective excluding transaction costs, so that comparisons with financial market returns are accurate.

Thus, given the weight for each contract, I calculate four different trading returns across all N contracts at every time. The first two, Open-to-Close and Close-to-End, are simply a weighted sum of the return distributions provided in the previous section of this study.

$$\text{Open-to-Close: } \mathrm{R}_{\text{open:close,t}} = \sum_{i=1}^{N} w_i r^i_{open:close,t}$$

$$\text{Close-to-End: } \mathrm{R}_{close:end,\mathrm{t}} = \sum_{i=1}^{N} w_i r^i_{close:end,t}$$

From a coding perspective, I encode both equations into the Trading Strategy script by performing vectorized operations on the dataframe with the indicator variables to create a new column of 'weights'. I then multiply this weight by the return distribution, which I previously calculated and presented in return table 4, to find the weighted return on every contract. I then use the groupby function with the ‘date’, or *t*, as the parameter to split the dataframe into groups and summarize these weighted returns. After this operation, I now have the daily returns of both the Open-to-Close and Close-to-End strategies. Lastly, I combine these two returns to form two other returns:

$$\text{Open-to-End: } \mathrm{R}_{\text{open:close,t}} + \mathrm{R}_{close:end,\mathrm{t}} \qquad \text{Trading Strategy: } \mathrm{R}_{\text{open:close,t}} - \mathrm{R}_{close:end,\mathrm{t}}$$

Because these two strategies are combinations of the Open-to-Close and Close-to-End returns that fundamentally track movement in the pricing of sports betting contracts, it follows that the Open-to-End and Trading Strategy returns test rationality in the market. If the Open-to-End return is the highest, then we know there is some positive movement in the contract price from the Open-to-Close since the Close-to-End return is smaller. Additionally, since the strategy continues to go long in the asset, an abnormal positive Open-to-End return shows that the market is underreacting to some information. The contract does not reach its fundamental value until the game's outcome, consistent with financial behavioral theories of underreaction such as PEAD. In contrast, if the final return, Trading Strategy, is the highest, we can infer that behavioral overreaction is present. This is because the strategy benefits from the initial movement in the contract price from the Open-to-Close; however, since it goes short on the underlying asset at the close, the strategy only exhibits abnormal returns when the price reverses from the Close-to-End.

I compute each strategy's mean and standard deviation with each contract's four daily portfolio returns. I then scale these values by the number of games in a season and the constant 2/5 since the strategy only incorporates two out of five quintiles in the dataset. This provides annualized mean and standard deviation of returns that facilitate comparison with financial markets. Lastly, I compute each portfolio's Sharpe Ratio by dividing the mean by the standard deviation. I also provide *t*-statistics by leveraging Python's Statsmodel.formula.api package.

# 5. Results:

The following section provides tables reporting the mean portfolio returns, *t*-statistics, and Sharpe ratios of the above-formulated trading strategies. All returns presented are on a percentage basis calculated without sportsbook vigorish or transaction costs. I include

Moskowitz's (2021) tables in the appendix and discuss how the similarities and differences between the results affect this paper's analysis.

Table 6 reports the four Momentum returns for the NFL, NBA, MLB, and NHL across Point Spread, Moneyline, and Over/Under contracts. The formulated Trading Strategy (1)-(2), which goes long on momentum contracts at the Open and then cancels the position to go short the contract from the Close-to-End, produces a consistent positive annual return across all panels. This is generally a result of a positive return from the Open-to-Close and a significant negative return from the Close-to-End. The consistent positive return from the Open-to-Close shows that a bettor who identifies a contract pitting a high-momentum team against a low-momentum team can expect to earn a return simply by holding the asset during the betting period before the contest is played. I note that while this provides useful analysis, this Open-to-Close return is infeasible to obtain in real betting markets, given that exiting a sports betting contract prior to the beginning of a game only pays back the bettor's initial principle and would not factor in an increase in the price of the sports betting contract price.

More importantly, we see that for many of the panels, there is a sharp correction in prices from the period of the Close-to-End consistent with the overreaction subcomponent of Prediction 3. This demonstrates that the positive return from the Open-to-Close is due to bettors overreacting to past performance defined by the Momentum characteristic, which ultimately results in a pricing reversion at the game's outcome. Since column 4, Trading Strategy, seeks to benefit from this predicted behavioral theory, it reports the greatest average return. Panel 3: NFL Moneyline reports an annual return of 7.45% with a significant *t*-stat of 2.77 for the Trading Strategy. This is consistent with the data presented by Moskowitz (2021), in which his Trading Strategy for all contracts reported an average return of 6.34%.

**Table 6:**
**Momentum Trading Strategy Returns**

| | (1) Open-to-Close | (2) Close-to-End | (1)+(2) Open-to-End | (1)-(2) Trading Strategy |
|---|---|---|---|---|
| | | Panel 1: NFL Point Spread | | |
| Mean | -0.64 | -2.86 | -3.50 | 2.23 |
| *t*-stat | -0.56 | -0.92 | -1.13 | 0.63 |
| Sharpe | -0.22 | -0.36 | -0.45 | 0.25 |
| | | Panel 2: NFL Over/Under | | |
| Mean | -1.72 | -5.17 | -6.89 | 3.45 |
| *t*-stat | -1.71 | -1.66 | -2.24 | 1.00 |
| Sharpe | -0.67 | -0.65 | -0.87 | 0.39 |
| | | Panel 3: NFL Moneyline | | |
| Mean | 4.40 | -3.05 | 1.35 | 7.45 |
| *t*-stat | 4.24 | -1.33 | 0.59 | 2.77 |
| Sharpe | 1.66 | -0.53 | 0.23 | 1.09 |
| | | Panel 4: NBA Point Spread | | |
| Mean | 6.67 | 1.18 | 7.84 | 5.49 |
| *t*-stat | 3.25 | 0.32 | 2.11 | 1.17 |
| Sharpe | 1.26 | 0.12 | 0.82 | 0.45 |
| | | Panel 5: NBA Over/Under | | |
| Mean | -1.89 | -3.81 | -5.70 | 1.92 |
| *t*-stat | -1.32 | -1.01 | -1.52 | 0.45 |
| Sharpe | -0.51 | -0.39 | -0.58 | 0.17 |
| | | Panel 6: MLB Over/Under | | |
| Mean | 0.87 | -10.07 | -9.20 | 10.94 |
| *t*-stat | 0.53 | -1.64 | -1.50 | 1.67 |
| Sharpe | 0.32 | -1.00 | -0.91 | 1.02 |
| | | Panel 7: MLB Moneyline | | |
| Mean | 6.67 | 1.18 | 7.84 | 5.49 |
| *t*-stat | 2.83 | -1.66 | -0.43 | 2.47 |
| Sharpe | 1.72 | -1.01 | -0.26 | 1.50 |
| | | Panel 8: NHL Over/Under | | |
| Mean | 0.33 | -2.78 | -2.45 | 3.10 |
| *t*-stat | 0.47 | -0.91 | -0.81 | 0.98 |
| Sharpe | 0.20 | -0.39 | -0.35 | 0.42 |
| | | Panel 9: NHL Moneyline | | |
| Mean | 3.96 | -3.97 | -0.01 | 7.92 |
| *t*-stat | 3.50 | -1.47 | -0.00 | 2.54 |
| Sharpe | 1.49 | -0.63 | -0.00 | 1.09 |

The fact that the Moneyline data reported in Panels 3, 7, and 9 exhibits the greatest *t*-stats and Trading Strategy returns is further supported by the prior motivation that the Moneyline contract is more robust than Point Spread markets since it can better reflect the implied increase in the value of the contract. Hence, these panels show that the Trading Strategy and Open-to-Close returns are significantly positive. Additionally, for these panels, the Open-to-End return exhibits small *t*-stats, which implies that these returns are not significantly different from 0. Accordingly, an Open-to-End return indistinguishable from zero suggests an efficient market in which the opening price accurately prices all market information. We infer then that bettors are overreacting according to the Momentum measure, which presents an opportunity for the Trading Strategy to achieve a significantly positive return as the market price moves due to irrational behavior. This pattern is consistent with Moskowitz's (2021) findings that the Open-to-End return is not statistically different from 0 for most of the examined markets.

The findings in Table 6 generally align with Moskowitz's results in Appendix 1. This makes sense, given that Momentum is a strong behavioral predictor routinely shown to cause investor overreaction in asset prices. However, since this paper incorporates data over the past years, in which the sports betting market structure has developed more similar to that of financial markets, the implications of this conclusion are even greater. Having found evidence of momentum premium in the market, I expect that similar phenomena in financial markets like the Post-Earnings-Announcement-Drift may be more closely linked to models of overreaction rather than the traditional models suggesting that investors underreact to the earnings surprise. Additionally, this finding also strengthens claims in financial markets that Momentum premium is a function of behavior irrationality and not a byproduct of faulty asset pricing models. Hence, the study by Fama and French (2012) discussed in a prior section can arguably make a better

claim that their findings of momentum premium in international markets are accurate, given that this study significantly linked the behavioral predictor to asset pricing movements.

Table 7 reports the four average portfolio returns, *t*-statistics, and Sharpe ratios for Value strategies in the MLB Moneyline and NBA Point Spread markets. Panel 1 and 2 report insignificant Trading Strategy returns of 0.10% (*t*-stat value of 0.02) and -4.66% (*t*-stat value of -1.02) for the MLB Moneyline and NBA Point Spread markets, respectively. While Moskowitz also reports statistically insignificant returns for his Value Trading Strategy column, the significantly positive returns in the NBA Point Spread Market differ. I find a 15.45% (*t*-stat of 4.27) for the Open-to-End return, suggesting that the Market is not accurately incorporating Value premium into sports betting contracts at the open. In fact, bettors appear to underreact to Value premiums since the price does increase from the Open-to-Close; yet the Close-to-End return still exhibits a significant positive return of 10.03%. These findings then suggest a contradiction with the results from the trading strategies linked to Momentum in this paper and the concluding findings from Moskowitz's paper.

**Table 7:**
**Value Trading Strategy Returns**

| | (1)<br>Open-to-Close | (2)<br>Close-to-End | (1)+(2)<br>Open-to-End | (1)-(2)<br>Trading Strategy |
|---|---|---|---|---|
| | | Panel 1: MLB Moneyline | | |
| Mean | 1.57 | 1.47 | 3.05 | 0.10 |
| *t*-stat | 3.05 | 0.24 | 0.50 | 0.02 |
| Sharpe | 1.87 | 0.15 | 0.31 | 0.01 |
| | | Panel 2: NBA Point Spread | | |
| Mean | 5.37 | 10.03 | 15.45 | -4.66 |
| *t*-stat | 2.68 | 2.78 | 4.27 | -1.02 |
| Sharpe | 1.03 | 1.07 | 1.64 | -0.39 |

Some explanations for this discrepancy include changing market conditions, data anomalies, faulty Value measure definitions, and the assumptions I took when constructing my Python trading strategies. The first theory connects to my use of updated data incorporating the legalization of sports betting markets and various sports trends since 2013. While the overall sports betting market has changed drastically in the past two to three years, it is also true that the structure of the NBA and its teams has evolved over the past decade. With a greater emphasis on one or two superstars, a league identity of NBA 'super teams' began with Lebron James moving in free agency from the Cleveland Cavaliers to the Miami Heat in 2010. This trend of major stars moving freely during free agency has continued in the past decade, causing the league to exhibit less parity and providing little opportunity for teams to improve without a significant player acquisition (Zeko, 2021). Thus, since the Value measures are defined by using a team's longer-term past performance, this lagging measure may not fully incorporate the market's current perception of the team's Value when the sports betting contract is released. For instance, a team that has historically underperformed the league but now trades for three high-profile players would now be considered an expensive team by the market, yet my Value measures would not immediately reflect this change given that the actual long-term performance of the franchise has not changed.

Interestingly, in the MLB, in which the same high-profile teams, L.A. Dodgers, Boston Red Sox, and New York Yankees, are the ones purchasing expensive all-star players, the returns are statistically insignificant, consistent with that of Moskowitz. This suggests that the changing conditions of the NBA market may result in a data anomaly due to how I have defined my measures of Value. Future analysis may take a deeper dive to examine the extent to which any Value premiums are present. The free agency nature of the league may be such that Value fails to

emerge as concretely as other leagues, which implement Salary Caps and are less dependent on a single player.

Nonetheless, since the findings from this paper's Momentum Trading Strategies generally align with that of Moskowitz, if Value premiums are present in the market but not accurately reflected in the trading strategy, this likely stems from the defined Value measures and not the code execution. The assumptions and procedures for creating my novel coding scripts generated seemingly realistic values consistent with market expectations. The Value measure definitions more generally demonstrate a limitation of this paper in that loose assumptions is needed to create a Value characteristic that is fundamentally equivalent to those used to equate value in financial markets. For this reason, more robust future research on each league's intricacies is necessary to create Value measures that can better model variations within sports leagues. As such, the paper finds inconclusive returns related to the Value characteristic.

**Table 8:**
**Size Trading Strategy Returns**

| | (1)<br>Open-to-Close | (2)<br>Close-to-End | (1)+(2)<br>Open-to-End | (1)-(2)<br>Trading Strategy |
|---|---|---|---|---|
| | | Panel 1: MLB Moneyline | | |
| Mean | -0.31 | 2.37 | 2.06 | -2.69 |
| *t*-stat | -0.92 | 0.66 | 0.58 | -0.74 |
| Sharpe | -0.60 | 0.43 | 0.37 | -0.48 |
| | | Panel 2: MLB Over/Under | | |
| Mean | 1.16 | -3.90 | -2.74 | 5.06 |
| *t*-stat | 1.12 | -1.05 | -0.74 | 1.27 |
| Sharpe | 0.73 | -0.68 | -0.48 | 0.83 |

Lastly, Table 8 provided above reports average portfolio returns, *t*-statistics, and Sharpe ratios for Size strategies in the MLB Moneyline and Over/Under markets. Consistent with Moskowitz's results in Appendix 3, all Size returns in Table 8 are statistically insignificant from

0. Since this paper uses more current data from a larger and more professional sports betting market, it makes intuitive sense that Size Premiums, which were already shown not to provide any return predictability, continue to be accurately priced in by the market. Especially within the framework of sports markets, we know that size is generally a slow-moving characteristic (Moskowitz, 2021). Thus, investors are not reacting to new information regarding Size premiums like they would to Momentum factors. Recent studies in financial markets (Alquist et al., 2018) have similarly concluded that size premiums do not exhibit significant evidence of abnormal returns for equities. Hence, I conclude that Size premiums exhibit little to no predictability for sports betting and financial markets.

One topic for future research that may be interesting would be whether there are any short-term Size premiums present for professional expansion teams. Given that these new franchises do not have prior season performance for investors to examine, agents may alter betting behavior to reflect superstitions about the market and the corresponding size of the new team. For instance, the NHL added the Vegas Golden Knights in 2017, who went to the league finals, losing to the Washington Capitals in Game 5 of the Stanley Cup. Without prior seasons to place a measure of Value on the team, it would be interesting to see if any Size premiums were significant for the Golden Knights during their opening season. This may enable an analysis of certain stock IPOs in which investors may incorporate irrational notions of fundamental Value regarding the ‘size’ of the IPO.

# 6. Conclusions

Using the previous work of Moskowitz (2021) as motivation and underlying guide, this paper conducts tests of asset pricing anomalies within sports betting markets. Leveraging novel data incorporating the legalization of sports betting markets across the majority of U.S. States the

paper examined the behavioral predictors of Momentum, Value, and Size in relationship to the movement of price in sports betting contracts. Implementing novel assumptions and coding scripts to construct the various behavioral measures and corresponding trading strategies, I find a significant Momentum premium, insignificant Value results, and evidence of no Size premiums within the Point Spread, Moneyline, and Over/Under markets for the four professional U.S. sports leagues. The Momentum and Size findings are consistent with those of Moskowitz, suggesting that despite a more professional and larger sports betting market, there is still a significant prevalence of agents overreacting to Momentum characteristics. In applying this to the financial literature, the study supports claims of overreaction activity in financial markets despite an increased presence of exogenous factors independent of sports betting markets. Overall, the paper accomplishes the goal of replicating the high-level work of Moskowitz and provides a starting point for future analysis of certain anomalies in various sports betting markets.

# 7. **Pledge**:

This paper represents my own work in accordance with University Regulations

Signature: John High

# 8. Bibliography


Allen, S. (n.d.). *What Sports Betting Teaches Us about Financial Markets*. Yale Insights. Retrieved November 17, 2022, from https://insights.som.yale.edu/insights/what-sports-betting-teaches-us-about-financial-markets

Alquist, R., Israel, R., & Moskowitz, T. (2018). Fact, Fiction, and the Size Effect. *Journal of Portfolio Management*, *45*, 34–61. https://doi.org/10.3905/jpm.2018.1.082

Barberis, N. (n.d.). *Psychology-based Models of Asset Prices and Trading Volume*.

Bathke Jr, A. W., Mason, T. W., & Morton, R. M. (2019). Investor Overreaction to Earnings Surprises and Post-Earnings-Announcement Reversals. *Contemporary Accounting Research*, *36*(4), 2069–2092. https://doi.org/10.1111/1911-3846.12491

Bernhardt, D., & Heston, S. (2010). Point Shaving in College Basketball: A Cautionary Tale for Forensic Economics. *Economic Inquiry*, *48*(1), 14–25. https://doi.org/10.1111/j.1465-7295.2009.00253.x

Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, *25*(2), 383–417. https://doi.org/10.2307/2325486

Feddersen, A., Humphreys, B. R., & Soebbing, B. P. (2020). Casual bettors and sentiment bias in NBA and NFL betting. *Applied Economics*, *52*(53), 5797–5806. https://doi.org/10.1080/00036846.2020.1776833

Fink, J. (2022, February 28). *Post Earnings Announcement Drift (PEAD)—Analyzing Alpha*. https://analyzingalpha.com/post-earnings-announcement-drift

Gandar, J. M., Dare, W. H., Brown, C. R., & Zuber, R. A. (1998). Informed Traders and Price Variations in the Betting Market for Professional Basketball Games. *The Journal of Finance*, *53*(1), 385–401. https://doi.org/10.1111/0022-1082.155346

Gandar, J., Zuber, R., O'Brien, T., & Russo, B. (1988). Testing Rationality in the Point Spread Betting Market. *The Journal of Finance*, *43*(4), 995–1008. https://doi.org/10.2307/2328148

Hill, D. (2019, June 5). *The Rise and Fall of the Professional Sports Bettor*. The Ringer. https://www.theringer.com/2019/6/5/18644504/sports-betting-bettors-sharps-kicked-out-spanky-william-hill-new-jersey

Levitt, S. D. (2004). Why are Gambling Markets Organised so Differently from Financial Markets? *The Economic Journal*, *114*(495), 223–246. https://doi.org/10.1111/j.1468-0297.2004.00207.x

McConnell, M. (n.d.). *Signal Theory and Earnings Surprise*. Retrieved November 17, 2022, from https://people.duke.edu/~charvey/Teaching/BA453_2004/MCM/A1webfile_MID.htm

*MLB 2022 Luxury Tax Payrolls Exceed $5 Billion For First Time; Six Clubs Pay Penalties*. (n.d.). Retrieved April 9, 2023, from https://www.forbes.com/sites/maurybrown/2023/01/19/mlb-2022-luxury-tax-payrolls-exceed-5-billion-for-first-time-six-clubs-pay-penalties/?sh=6e7c7f06efed

Moskowitz, T. J. (2021). Asset Pricing and Sports Betting. *The Journal of Finance*, *76*(6), 3153–3209. https://doi.org/10.1111/jofi.13082

Murphy, Christopher, 2007, Performance and pay in the NBA, Unpublished thesis, University of Wisconsin.

Peta, J. (2014). *Trading Bases: How a Wall Street Trader Made a Fortune Betting on Baseball* (Reprint edition). Berkley.

Preciado, D. (2022, July 5). *States Where Sports Betting Is Legal*. Forbes Betting. https://www.forbes.com/betting/sports-betting/legal-states/

Rathmanner, D. (2022, September 22). 20+ Surprising Sports Betting Statistics (for 2022). *Odds Assist*. https://oddsassist.com/sports-betting/resources/sports-betting-statistics/

Shorin, G. (n.d.). *Team Payroll Versus Performance in Professional Sports: Is Increased Spending Associated with Greater Success?*

Silverman, A. (2022, January 18). 2021 Was a Banner Year for Sports Betting Participation. *Morning Consult*. https://morningconsult.com/2022/01/18/sports-betting-trends/

*Understanding the NBA Salary Cap: It Is Rocket Science | News, Scores, Highlights, Stats, and Rumors | Bleacher Report*. (n.d.). Retrieved April 8, 2023, from https://bleacherreport.com/articles/224051-understanding-the-nba-salary-cap-it-is-rocket-science

Werbach, K. (n.d.). *Yes, Trading Is a Game, Just Not the One Many Investors Signed Up For*. Retrieved April 9, 2023, from https://www.barrons.com/articles/yes-trading-is-a-game-just-not-the-one-many-investors-signed-up-for-51613052836

Zeko, M. (2021, May 24). *Super Teams: The Story of Basketball | New University | UC Irvine*. https://newuniversity.org/2021/05/24/super-teams-the-story-of-basketball/

# 9. Appendix

**Appendix 1:**
**Momentum Trading Strategy Returns**

| | (1) Open-to-Close | (2) Close-to-End | (1)+(2) Open-to-End | (1)-(2) Trading Strategy |
|---|---|---|---|---|
| | | Panel A: All Sports, All Contracts | | |
| Mean | 2.97 | -3.37 | -0.40 | 6.34 |
| *t*-stat | 7.75 | -1.63 | -0.19 | 2.96 |
| Sharpe | 1.33 | -0.33 | -0.04 | 0.60 |
| | | Panel B: All Sports, Point Spread Contract | | |
| Mean | 2.20 | -4.89 | -2.69 | 7.10 |
| *t*-stat | 2.63 | -1.73 | -0.95 | 2.31 |
| Sharpe | 0.61 | -0.41 | -0.22 | 0.54 |
| | | Panel C: All Sports, Moneyline Contracts | | |
| Mean | 1.38 | -4.96 | -3.58 | 6.34 |
| *t*-stat | 3.44 | -1.24 | -0.95 | 2.31 |
| Sharpe | 1.01 | -0.40 | -0.29 | 0.50 |
| | | Panel D: All Sports, Over/Under Contracts | | |
| Mean | 4.95 | -1.01 | 3.94 | 5.96 |
| *t*-stat | 6.81 | -0.32 | 1.27 | 1.82 |
| Sharpe | 1.80 | -0.09 | 0.36 | 0.51 |
| | | Panel E: NFL, All Contracts | | |
| Mean | 1.22 | -3.72 | -2.50 | 4.94 |
| *t*-stat | 2.88 | -2.25 | -1.51 | 2.82 |
| Sharpe | 0.45 | -0.32 | -0.21 | 0.41 |
| | | Panel F: MLB, All Contracts | | |
| Mean | 5.38 | 3.98 | 9.37 | 1.40 |
| *t*-stat | 6.57 | 0.74 | 1.76 | 0.26 |
| Sharpe | 1.59 | 0.18 | 0.43 | 0.06 |
| | | Panel G: NHL, All Contracts | | |
| Mean | 0.54 | -2.72 | -2.18 | 3.26 |
| *t*-stat | 1.58 | -0.63 | -0.50 | 0.75 |
| Sharpe | 0.44 | -0.17 | -0.13 | 0.20 |

**Appendix 2:**
**Value Trading Strategy Returns**

| | (1) Open-to-Close | (2) Close-to-End | (1)+(2) Open-to-End | (1)-(2) Trading Strategy |
|---|---|---|---|---|
| | | Panel A: All Sports, All Contracts | | |
| Mean | 0.29 | 2.13 | 2.42 | -1.84 |
| *t*-stat | 0.53 | 0.71 | 0.81 | -0.59 |
| Sharpe | 0.14 | 0.21 | 0.24 | -0.18 |
| | | Panel B: All Sports, Point Spread Contract | | |
| Mean | 0.09 | 1.55 | 1.64 | -1.46 |
| *t*-stat | 0.09 | 0.47 | 0.50 | -0.41 |
| Sharpe | 0.02 | 0.13 | 0.14 | -0.11 |
| | | Panel C: All Sports, Moneyline Contracts | | |
| Mean | 0.26 | 1.83 | 2.08 | -1.57 |
| *t*-stat | 0.42 | 0.28 | 0.32 | -0.24 |
| Sharpe | 0.20 | 0.15 | 0.17 | -0.13 |
| | | Panel D: All Sports, Over/Under Contracts | | |
| Mean | 0.42 | 2.72 | 3.14 | -2.30 |
| *t*-stat | 0.39 | 0.56 | 0.65 | -0.45 |
| Sharpe | 0.16 | 0.25 | 0.28 | -0.20 |

**Appendix 3:**
**Size Trading Strategy Returns**

| | (1) Open-to-Close | (2) Close-to-End | (1)+(2) Open-to-End | (1)-(2) Trading Strategy |
|---|---|---|---|---|
| | | Panel A: All Sports, All Contracts | | |
| Mean | 0.51 | -2.62 | -2.11 | 3.13 |
| *t*-stat | 0.97 | -0.89 | -0.72 | 1.03 |
| Sharpe | 0.25 | -0.26 | -0.21 | 0.30 |
| | | Panel B: All Sports, Point Spread Contract | | |
| Mean | -0.06 | -0.76 | -0.81 | 0.70 |
| *t*-stat | -0.06 | -0.23 | -0.25 | 0.20 |
| Sharpe | -0.02 | -0.06 | -0.07 | 0.05 |
| | | Panel C: All Sports, Moneyline Contracts | | |
| Mean | 0.53 | -9.30 | -8.77 | 9.83 |
| *t*-stat | 0.75 | -1.42 | -1.34 | 1.48 |
| Sharpe | 0.37 | -0.75 | -0.71 | 0.78 |
| | | Panel D: All Sports, Over/Under Contracts | | |
| Mean | 0.78 | 3.13 | 3.91 | -2.35 |
| *t*-stat | 0.78 | 0.69 | 0.86 | -0.49 |
| Sharpe | 0.30 | 0.28 | 0.35 | -0.20 |

## Appendix 4:

All code and datafiles can be accessed at https://github.com/jhigh7113/Thesis.git