What Can Betting Markets Tell Us About Investor Preferences and Beliefs? Implications for Low Risk Anomalies

June 4, 2021

Abstract

An empirical puzzle in financial markets, known as the low-risk anomaly, is that riskier assets earn lower risk-adjusted returns than less risky assets. Theories for this phenomenon focus either on market frictions, such as leverage costs, or non-traditional preferences for lottery-like payoffs. We relate the low risk anomaly to the Favorite-Longshot Bias in betting markets, where returns to betting on riskier "longshots" are lower than returns to betting on "favorites," and provide novel evidence to both anomalies. Synthesizing the evidence, we study the joint implications from the two settings for a unifying explanation. Rational theories of risk-averse investors with homogeneous beliefs cannot explain the cross-sectional relationship between *diversifiable* risk and return in betting markets. Using our unique betting data, we find that non-traditional preferences, rather than incorrect beliefs, are most consistent with the facts. We calibrate a model of reference-dependent preferences, featuring probability weighting and diminishing sensitivity, that is uniquely able to capture the choice to bet and wager sizes. Calibrated parameter values for probability weighting and diminishing sensitivity that capture financial markets facts can simultaneously explain the betting market facts. However, explaining the choice to bet is at odds with loss-aversion, a feature of reference-dependent preferences often used in financial markets applications.

1 Introduction

A major empirical puzzle in financial economics is that riskier assets on average earn lower returns per unit of risk than less risky assets. Starting with Black et al. (1972), a long literature documents the risk-adjusted underperformance of risky stocks in the cross-section, measuring risk in a variety of ways (e.g., idiosyncratic stock price volatility; Ang et al. (2006, 2009) and market beta; Frazzini and Pedersen (2014) and Asness et al. (2020)). The risk-adjusted underperformance of risky assets is also found in other asset classes. Riskier options contracts (out-of-the-money put *and* call options) written on the same underlying stock earn lower risk-adjusted returns than less risky at-the-money contracts (Bondarenko (2014), Ni (2008), Boyer and Vorkink (2014), Frazzini and Pedersen (2020) and Baele et al. (2016)). Similarly, in US Treasuries, corporate bonds, equity indices, currencies, and commodities, riskier assets earn lower risk-adjusted returns (Frazzini and Pedersen (2014)).

The robust finding of "low-risk anomalies" stands in contrast to the predictions of classical models of asset pricing, such as the Capital Asset Pricing Model. Explanations for this puzzle that rely on risk averse investors with correct beliefs lean on capital market frictions, such as leverage constraints, or mismeasurement of risk. In contrast, behavioral explanations have also been proposed that suggest non-traditional preferences (notably, a preference for "lottery-like" payoffs) may play an important role.[1]

The Favorite-Longshot bias (FLB) is another prominent anomaly related to low-risk anomalies. First documented by Griffith (1949), and confirmed as an empirical regularity, the FLB describes a phenomenon found in horse racetrack betting, where the returns to betting on a *longshot* (high payoff with low probability) are substantially lower than the returns to betting on a *favorite* (low payoff with high probability). A central, unresolved debate in the literature is whether this pattern originates from incorrect bettor beliefs of expected outcomes or bettor preferences for lottery-like payoffs.

We draw parallels between these two related phenomena in order to shed light on a possible unifying explanation. Our aim is to see if we can learn something from betting markets that can help resolve the low risk anomaly in financial markets. Certain explanations offered in financial markets, such as those with risk-averse investors having correct beliefs, cannot explain the results in idiosyncratic betting markets. Hence, betting markets reject these theories as a potential unifying explanation. On the other hand, differences between the two markets, including possible segmentation of participants in these markets, may make comparisons challenging. Another issue that often gives pause to comparing the two markets is that

[1] A prominent theory of market frictions, due to Black (1972) and Frazzini and Pedersen (2014), is that excess borrowing costs drive investors seeking higher returns to increase the weight of risky assets in their portfolio, rather than holding an optimally leveraged position in the market portfolio. In equilibrium, this behavior lowers the return on systematically riskier assets. Ang et al. (2009) and Gormsen and Lazarus (2020) suggest low-risk anomalies may be explained by exposure to non-diversifiable risk factors not captured by the Capital Asset Pricing Model. Brunnermeier et al. (2007) and Barberis and Huang (2008) are prominent examples of theoretical models that suggest investors have a preference for idiosyncratically skewed assets.

the average bet has a negative expected return, while the average investment in financial markets offers a positive expected return. We examine what we can learn by comparing the evidence in these two markets, including models that can accommodate an agent who decides to bet despite knowing that betting has a negative expected return.

We find that much can be learned from betting markets to help explain the low risk anomaly in financial markets. A substantial literature in economics and finance, starting with Weitzman (1965), uses betting markets to study decision-making under uncertainty. As Thaler and Ziemba (1988) articulate, betting markets are particularly well-suited to study decision making under uncertainty because bets are gambles on observable, idiosyncratic outcomes that have well-defined termination points, and are not affected by the beliefs and preferences of bettors. These features facilitate the study of how beliefs and preferences influence prices in a cleaner way than is possible in financial markets, where rich dynamics, the exposure of assets to systematic, macroeconomic risk, and the often unobservable nature of the terminal outcomes that assets' payoffs are tied to, confound analysis of the role that beliefs and preferences may play in setting prices.

To understand how betting markets can inform theories of risk and return in financial markets, we study a sample of betting contracts written on 36,609 college and professional basketball and football games. Our setting contains unique features that facilitate direct comparisons to financial markets. First, bets are on the difference in the number of points scored by the two teams playing, an observable quantity. Therefore, we observe the underlying distribution of fundamentals upon which the contracts are written, unlike at the horse racetrack. Second, we observe multiple types of betting contracts on the same outcome, focusing on two in particular: Moneyline contracts and Spread contracts. Both contracts are contingent claims based on the same quantity, the difference in the number of points scored by each team. A Moneyline contract is a fixed-odds bet on which team will win the game (by scoring more points), where the underdog (less favored team) offers a low probability of a high payoff, while the favorite offers a high probability of a low payoff. Spread contracts, on the other hand, bet on the *same outcome*, but are structured as bets with roughly 50-50 odds, where a bet on the favorite (underdog) pays off when the point difference exceeds (falls short of) the *expected* point difference. Importantly, the potential payoffs and riskiness of Spread bets on the favorite and underdog are exactly the same across contracts.[2] Comparing the Moneyline (where risk varies for betting on the favorite versus the underdog) to the Spread contract (where risk is the same for betting

[2]Many studies have focused on Spread contracts in team sports to study the informational efficiency of betting markets (Zuber et al. (1985), Sauer et al. (1988), Gandar et al. (1988), Camerer (1989), Brown and Sauer (1993), Golec and Tamarkin (1991), Gray and Gray (1997), Gandar et al. (1998), and Levitt (2004)). Despite the availability of Moneyline contracts in team sports, they have been studied less extensively, likely in part because Spread contracts are the more popular contract to bet on. Papers that study the pricing of Moneyline contracts in team sports include Pope and Peel (1989), Woodland and Woodland (1994), Kuypers (2000), and Moskowitz (Forthcoming), the latter studying Spread contracts, Moneyline contracts, and a third-type of contract, the Over-Under contract. A unique aspect of our work here is studying the *joint* implications of pricing patterns in Spread and Moneyline contracts.

on the favorite versus the underdog) on the same game's outcome provides a unique contrast between two claims on the same payoff state space. This comparison provides a clean and well-identified test of how idiosyncratic risk matters for pricing, controlling for any other characteristic of each team or any aspect of the match up between the two teams. Everything except for the structure of payoffs is identical across the two contract types.

We document three empirical facts in our setting and connect them to evidence in financial markets. First, we show a Favorite-Longshot Bias in our setting. Studying the returns of the cross-section of Moneyline contracts, we find that less risky bets on more favored teams earn higher average returns and more risky bets on less favored teams earn lower average returns. This fact parallels the empirical evidence in the cross-section of stocks: less risky stocks earn higher returns and riskier stocks earn lower returns. Second, we document that in the cross-section of Spread and Moneyline contracts written on the same game, riskier contracts earn lower returns and less risky contracts earn higher returns. This fact parallels an empirical feature of stock options markets: for options contracts written on the same underlying asset, (more risky) out-of-the-money options earn lower returns and (less risky) in-the-money options earn higher returns. Third, we show that in the cross-section of Spread contracts, there is no relationship between how favored a team is and the returns to betting on that team, in contrast with the pattern observed for Moneyline contracts *on the same games*. The Favorite-Longshot bias only exists for the Moneyline contracts, which face different risks, and *not* the Spread contracts, where risk is uniformly set to a 50-50 median outcome. This result does not have a direct parallel in traditional financial markets. It is, however, a unique and important result because it demonstrates that market estimates of the expected point difference, the key piece of information required to price all contracts, do not reflect a Favorite-Longshot bias, indicating that informational inefficiencies are unlikely to explain the Favorite-Longshot bias. Moreover, Spread contracts provide a ubiquitous control for all other features associated with each game − beliefs, team preferences, information shocks, sentiment, etc. – *except* idiosyncratic risk. The lack of a pattern in the pricing of Spread contracts, in contrast with the strong low risk pattern in Moneyline contracts, clearly identifies the role of idiosyncratic risk for pricing. The evidence suggests non-traditional preferences, rather than incorrect beliefs, primarily drive the patterns in the data, in contrast to recent work studying betting at the horse racetrack.[3]

To illustrate our results clearly and relate them to evidence from financial markets, we construct an implied volatility surface for sports betting contracts akin to the implied volatility surface used to analyze

[3]For example, Snowberg and Wolfers (2010) use the prices of exotic bets on the exact order that horses will finish and find that the representative bettor incorrectly reduces compound lotteries. They interpret this result as evidence that the Favorite-Longshot Bias is driven by informational inefficiency. In contrast, we find that market prices reflect the key piece of information required to determine win probabilities and price Moneyline contracts – the expected point difference – and do so in a manner inconsistent with informational inefficiencies driving the Favorite Longshot Bias. This interpretation does not speak to other potential inefficiencies in betting markets, such as the pricing of compound lotteries, a bias for betting on the home team, etc.

stock option prices. We estimate the "implied volatility" of a Moneyline contract as the standard deviation of the distribution of point differentials that would make the contract have the same expected return as the corresponding Spread contract. Strikingly, we observe an implied volatility smile that is qualitatively and quantitatively similar to the famous volatility smile in options markets, with contracts that are "deep in- and out-of-the-money" (very high and very low probabilities of paying off) having higher implied volatilities. Many explanations for the smile in the options implied volatility surface relate to misspecification of the probability distribution of the underlying assets' return. However, our quantitatively similar results in a simpler state space with no dynamics, casts serious doubt on these types of explanations in favor of other explanations, such as risk preferences or heterogeneous beliefs.

Our empirical evidence demonstrates substantial similarity between the cross-section of risk and return in financial markets and betting markets, despite their institutional differences. Putting the evidence from both settings side-by-side informs which theories can simultaneously explain both sets of facts. Theories that rely on the traditional assumptions that investors have correct beliefs, are risk-averse, and evaluate risks in the context of the broader set of risks they face, predict no relationship in equilibrium between *idiosyncratic* risks and expected returns. Only systematic, non-diversifiable risks matter. Hence, classical theories based on traditional preferences cannot explain the relationship between risk and return in the cross-section of betting contracts. The evidence suggests that non-traditional preferences may play an important role.

To gain a better sense of the role that non-traditional preferences may play, we calibrate a reference-dependent preference specification that features rank-dependent probability weighting and diminishing sensitivity. We find that similar parameter values for diminishing sensitivity and probability weighting used to explain the facts in financial markets are also able to explain the Favorite-Longshot bias in betting. Probability weighting induces people to overweight tail events in their decision, and explains the low required return for betting on risky underdogs, as noted in prior work.[4] Diminishing sensitivity, where the magnitude of marginal utility is smaller for gains and losses that are further from the bettor's reference point, also plays an important role, providing an explanation for why bettors choose to wager on negatively skewed, negative expected return contracts on favorites. The potential importance of diminishing sensitivity for explaining bettor behavior is both a new insight in the literature on the Favorite-Longshot bias, as well as a novel economic setting in which diminishing sensitivity, a relatively less studied component of reference-dependent preferences, may be applicable.[5] Relative to previous work studying betting behavior, our specification also

[4]Previous works that emphasize the potential role of probability weighting in explaining the Favorite-Longshot bias include Jullien and Salanié (2000) and Snowberg and Wolfers (2010). Studies of the potential role of probability weighting in financial markets include Kliger and Levy (2009) and Baele et al. (2016), who study the potential role of probability weighting in explaining financial markets facts, and Barberis et al. (2001), Barberis and Huang (2008), Barberis et al. (2016), and Barberis et al. (Forthcoming), who focus on the ability of probability weighting, and Cumulative Prospect Theory more generally, to explain stock price behavior.

[5]In their review chapter on reference-dependent preferences, O'Donoghue and Sprenger (2018) note that "Most applications

endogenizes the choice to bet and the amount that bettors choose to wager, and using our unique data on betting volume, we find that the model is able to explain variation in bet sizes across different contracts. However, the model's ability to explain the results is at odds with loss-aversion, the idea that losses loom larger than gains, which is traditionally assumed in applications of Cumulative Prospect Theory to financial markets. We discuss how the model's ability to explain betting behavior hinges on a particular form of rank-dependent probability weighting, or alternatively on bettors being *loss-tolerant* rather than loss-averse.

We also discuss two alternative belief-based explanations for the Favorite-Longshot bias. The first is that bettors accurately perceive the expected point difference, but misperceive higher moments of the point distribution. This bias may lead them to overestimate the probability that underdogs win and underestimate the probability that favorites win. In the absence of other features, this explanation does not capture bettor decisions to bet a positive and finite amount. This explanation is also at odds with the fact that betting markets do a very good job capturing other features of the point distribution, which we show. The second alternative explanation is that bettors have heterogeneous beliefs, and the marginal bettor choosing to wager on more extreme underdogs has more extreme beliefs, reflected in more negative returns to betting on underdogs.[6] Using volume data on the *actual* proportion of bets placed on each team, we find a much greater proportion of bets are made on the most extreme underdogs (which earn highly negative returns) than on moderate underdogs, inconsistent with the marginal bettor having more extreme beliefs in games with more extreme outcomes. The evidence suggests that while belief heterogeneity could help explain some of the observed patterns, it cannot explain the popularity of extreme underdogs with very negative returns. We conclude that non-traditional preferences likely play a more dominant role in explaining the facts.

A common assumption in economics is that risk preferences are stable across decision contexts. However, as Barseghyan et al. (2018) note, relatively little work has been done in assessing this stability.[7] Our work takes a step in this direction by drawing parallels between anomalies in traditional financial markets and idiosyncratic betting markets. While there are institutional differences between the two markets, betting markets share a number of similar features with financial markets: large transaction volume, widely available

of reference-dependent preferences...ignore the possibility of diminishing sensitivity" and further note that "the literature still needs to develop a better sense of when diminishing sensitivity is important."

[6] Earlier studies focusing on belief heterogeneity in betting markets include Shin (1991, 1992) and Ottaviani and Sørensen (2009, 2010). Gandhi and Serrano-Padial (2015) present the insight that the choice of which horse to bet on in a race is analogous to the choice among horizontally-differentiated products, and use this insight to estimate the degree of belief heterogeneity implied by the Favorite-Longshot bias. Green et al. (2020) suggest that the Favorite-Longshot bias is driven by another form of belief heterogeneity, namely racetracks deceiving gullible bettors and driving a wedge between the beliefs of informed and uninformed bettors. In the asset pricing literature, a non-exhaustive list of papers studying heterogeneous beliefs include Harrison and Kreps (1978); Scheinkman and Xiong (2003); Diether et al. (2002); Geanakoplos (2010); Simsek (2013), and Giglio et al. (2021).

[7] Exceptions include Barseghyan et al. (2011), who examine the choices of a set of households across insurance contexts, and Einav et al. (2012), who study the choices of a set of Alcoa employees across different insurance contexts and 401(k) investment choices. This research focuses on the stability of preferences of specific *individuals* making decisions across different contexts. In contrast, our work is focused on the stability of risk preferences implied by equilibrium market prices across betting markets and traditional financial markets.

information, market making activity, arbitrage activity (from professional bettors and hedge funds), and professional analysts. Preferences related to entertainment and loyalty are secondary motives (as they also are in the stock market, see e.g., Dorn and Sengmueller (2009) and Grinblatt and Keloharju (2009)).[8] Given the similarity of features between the two markets, there is reason to believe that the facts in the two settings could share similar economic drivers. A previous generation of work recognizes the connection between betting and financial markets, primarily focusing on the informational efficiency of betting markets and its implications for theories in financial markets (see e.g., Pankoff (1968), Durham et al. (2005), and Moskowitz (Forthcoming)). Calibrating preferences to match the betting market facts, we find that certain features of non-traditional preferences, namely probability weighting and diminishing sensitivity, may play a role in a unifying explanation for the facts across settings. However, while loss-aversion is thought to play a role in explaining investor behavior in financial markets, it is difficult to reconcile with the decision to bet. This suggests the potential segmentation of market participants in traditional financial markets and betting markets, or alternatively, people displaying loss-aversion when making financial decisions and not doing so when betting.

The rest of the paper proceeds as follows. Section 2 describes the betting markets and our data in more detail. Section 3 presents empirical evidence of the Favorite-Longshot Bias in our setting, and relates it to evidence from financial markets. Section 4 discusses the implications of our results for theories across both betting and financial markets. Section 5 presents a calibration of preferences to explain the betting market facts, and Section 6 analyzes alternative explanations, including erroneous beliefs and belief heterogeneity. Section 7 concludes.

2 Empirical Setting and Data

Sports betting markets are large, liquid, and active. According to Statista.com, global sports betting markets produced an aggregate gross gaming yield (notional bets taken by betting operators minus winnings/prizes paid out) of nearly $200 billion in 2017 and 50 percent of U.S. adults have made a sports bet (which is higher than stock market participation rates, see e.g., Vissing-Jørgensen (2002)). In the U.S., the American Gaming Association estimates that 4 to 5 billion U.S. dollars are wagered legally each year at Nevada sportsbooks, the only state where it was legal, but the amount bet illegally with local bookies, offshore operators, and other enterprises is roughly 30 times that figure.[9] With the recent U.S. Supreme Court decision overturning the Professional and Amateur Sports Protection Act of 1992 that prohibited state-sanctioned sports betting,

[8] See Peta (2014) for a discussion of the industry of professional gambling and the use of financial tools from Wall Street in the sports betting market.

[9] According to the 1999 Gambling Impact Study, an estimated $80 billion to $380 billion was illegally bet each year on sporting events in the U.S., dwarfing the $2.5 billion legally bet each year in Nevada (Weinberg (2003)).

the expectation is that this market will grow considerably.[10]

We study bets placed on four types of games: National Collegiate Athletic Association (NCAA) Football games, NCAA Basketball games, National Basketball Association (NBA) games and National Football League (NFL) games. For the primary betting contracts we study, a bookmaker controls the "lines" (prices). At the start of betting, bookmakers set an opening line, which is determined by pre-betting market activity from large professional bettors, after which bettors may place bets until the start of the game. Betting volume flows can change the price of bets if bookmakers seek to balance the money being bet on each side. On average, bookmakers generally set lines such that they take no risk, though this may not be true in every game. Without taking risk, they profit from the 'vig,' which refers to the transaction cost implicitly embedded in contract prices that makes betting a negative expected return proposition for bettors.[11] Bettors receive the price at the time they make their bet, even if the line later changes. Betting closes before the start of the game. For some contracts (e.g., the NFL), the time between open and close can be six days, while for others (e.g., the NBA), it may only be a few hours.[12]

Our primary focus is on two types of betting contracts: Moneyline contracts and Spread contracts. The payoffs of Spread and Moneyline contracts are determined by the difference in points scored by each team in the game, $P_A - P_B$, where P_A and P_B are the number of points scored by teams A and B. Moneyline contracts are outright bets on which team will win the game (i.e., a bet on $P_A - P_B \lesseqgtr 0$). Spread contracts are bets on whether the point-differential between the two teams exceeds the "Spread Line" for the game (i.e., a bet on $P_A - P_B \lesseqgtr \bar{x}$, where $\bar{x}$ is the Spread). In an auxiliary test, we also analyze Over/Under contracts. Over/Under contracts are bets on whether the total number of points scored in a game will exceed the line set by the bookmaker (i.e., a bet on $P_A + P_B \lesseqgtr T$). We describe the mechanics and pricing of the three types of contracts in more detail below. For all of our analysis, we use the prices set at the close of betting and assume that all bettors transact at the closing price. In the appendix, we verify that the same pricing patterns persist using opening prices.

2.1 Moneyline Contract

The Moneyline (ML, also known as American Odds) contract is a fixed-odds contract that is a bet on which team wins. The Moneyline contract offers different potential payoffs per dollar wagered on a team depending

[10] Hudson (2014) shows that in the UK, where sports betting is legal, betting has increased annually by about 7%, fueled by online and mobile betting.

[11] Bookmakers may occasionally set prices between the 'informationally efficient' price and the price that balances demand. See Levitt (2004).

[12] This style of betting is different than parimutuel pools, which are the common way in which betting at the horse racetrack is organized in North America (horse betting in the UK is organized with bookmakers, similar to our setting). In a parimutuel pool, prices and odds are not set at the beginning of betting. Rather, each bettor specifies the amount of money they wish to bet, without knowing prices or odds. The prices are only set at the close of betting, and are set such that winning bets are paid out in proportion to the stakes wagered from the losing bets. The literature has studied horse betting both in the parimutuel and the bookmaker context and found similar patterns in the Favorite-Longshot Bias for both.

upon which team is bet in a game. Larger potential payoffs are offered for betting on underdogs (that are less likely to win) and smaller potential payoffs are offered for betting on favorites (that are more likely to win). For example, if a bet of \$100 on Chicago (the favored team) over New York is listed as -165, then the bettor risks \$165 to win \$100 if Chicago wins. Betting on New York (the underdog) the Moneyline might be $+155$, which means risking \$100 to win \$155 if New York wins. The \$10 difference is commission paid to the sportsbook. The payoffs for a \$100 bet on team A over team B on a Moneyline contract listed at $-\$M$ are as follows:

$$\text{Payoff}^{ML} = \begin{cases} M + 100, & \text{if } (P_A - P_B) > 0 \quad (\text{“win”}) \\ \text{Max}(M, 100), & \text{if } (P_A - P_B) = 0 \quad (\text{“tie”}) \\ 0, & \text{if } (P_A - P_B) < 0 \quad (\text{“lose”}) \end{cases} \tag{1}$$

where M is either > 100 or < -100 depending on whether team A is favored or team B is favored to win.

For readers familiar with fixed-odds contracts in other settings, Moneylines can be directly converted to both fractional odds, as quoted in the UK, and decimal odds, as quoted in continental Europe. Positive Moneylines quote the money to be won for a \$100 wager, so for example, an ML of $+400$ would be quoted as 4/1 in fractional odds and as 5 in decimal odds. Negative Moneylines quote the amount of money to be wagered to win \$100, so for example, an ML of -400 would be quoted as 1/4 in fractional odds and as 1.25 in decimal odds.

2.2 Spread Contract

The Spread (S) contract is a bet on a team winning by at least a certain number of points known as the "spread." For example, if Chicago is a 3.5 point favorite over New York, the spread is quoted as -3.5, which means that Chicago must win by four points or more for a bet on Chicago to pay off. The spread for betting on New York would be quoted as $+3.5$, meaning that New York must either win or lose by less than four points in order for the bet to pay off. Unlike Moneyline contracts, Spread contracts offer the same potential payoff for betting on either team in a game for a given wager. In the next section, we show empirically that Spreads are set to make betting on either team roughly a 50-50 proposition or to balance the total amount bet on each team. The typical bet is \$110 to win \$100. So, the payoffs for a \$110 bet on team A over team B on a spread contract of $\bar{x}$ points are:

$$\text{Payoff}^{S} = \begin{cases} 210, & \text{if } (P_A - P_B) > \bar{x} \quad (\text{“cover”}) \\ 110, & \text{if } (P_A - P_B) = \bar{x} \quad (\text{“push”}) \\ 0, & \text{if } (P_A - P_B) < \bar{x} \quad (\text{“fail”}) \end{cases} \tag{2}$$

where "cover, push, and fail" are terms used to define winning the bet, a tie, and losing the bet, respectively. For half-point spreads, ties are impossible since teams can only score in full point increments.

2.3 Over/Under Contract

Finally, the Over/Under contract (O/U), is a contingent claim on the total number of points scored ($y = P_A + P_B$). Sportsbooks set a "total", T, which is the predicted total number of points the teams will score combined. Bets are placed on whether the actual outcome of the game will fall "over" or "under" T. The payoffs are similar to the Spread contract in that a bet is for \$110 to win \$100. For example, wagering on the over contract in a game earns the following payoffs:

$$\text{Payoff}^{OU} = \begin{cases} 210, & \text{if } (P_A + P_B) > T \quad (\text{“over”}) \\ 110, & \text{if } (P_A + P_B) = T \quad (\text{“push”}) \\ 0, & \text{if } (P_A + P_B) < T \quad (\text{“under”}) \end{cases} \tag{3}$$

Bookmakers set lines such that there is a 50/50 chance of either side of the contract paying off, or the dollar volume on both sides of the contract is approximately 50/50.

2.4 Betting Contract Risk

Betting on spread contracts (on either side) amounts to taking a gamble with an approximately 50% probability of paying off. Betting on Moneyline contracts on the favorite amounts to taking a gamble with a greater than 50% probability of paying off, while betting on the underdog amounts to taking a gamble with a less than 50% probability of paying off. Per dollar wagered, the lines for each contract are set such that the potential payoff for a winning bet is decreasing in the probability of the contract paying off. Bets on the underdog in the Moneyline, which offer a low probability of a high payoff, are the riskiest contract type (measuring risk in terms of idiosyncratic variance or skewness), with risk decreasing in the probability the underdog wins. Moneyline contracts on the favorite, which offer a high probability of a low payoff, are the least risky contract, with contract risk decreasing in the probability the favorite wins the game. Spread bets on the favorite and underdog face the same risk, since they are 50-50 bets or bets on the median outcome, which is intermediately risky compared with Moneyline contracts on the favorite and the underdog. Additionally, since Spread contract payoffs are set to be the same across games, there is no (or negligible) cross-sectional variation in Spread contract risk across games (all are approximately 50-50 bets). However, Moneyline contracts exhibit significant cross-sectional heterogeneity in risk across games as there are significant differences in the relative quality of the two teams playing across games.

We exploit both of these risk differences: 1) for a given game, we compare Moneyline favorite versus underdog against the Spread bet on the favorite versus the underdog and 2) we compare Moneyline favorites versus underdogs across games that vary in the difference in win probabilities between the two teams (i.e., ex ante competitive contests versus lopsided contests) and hence the risk of betting, but where Spread bets across games hold risk constant.

2.5 Data

The data come from SportsInsights.com, beginning in 2005 and ending in May 2013 for all four sports we analyze. There are 36,609 total games in the sample: 21,982 NCAA Basketball games, 8,392 NBA Basketball games, 4,392 NCAA Football games and 1,843 NFL Football games. Each game contains two betting contracts – the Moneyline and Spread contract – based on the point difference between the two teams, and a third contract based on the total points scored by both teams combined (Over/Under contract).

The betting lines are drawn from the Las Vegas legalized sportsbooks and online betting sportsbooks, where all bookmakers offer nearly identical closing lines on a given game. The data include all games from the regular season and playoffs/post-season. The data for all games include the team names, start and end time of game, final score, and the opening and closing betting lines across all contracts on each game. The betting lines are taken from the Las Vegas legalized sportsbooks and online sportsbooks and are transactable quotes. For the majority of contests, all bookmakers offer nearly identical lines in the database. On the rare occasion when lines differ (less than 1% of the time), that contract is removed from the sample. Results are robust to using the highest or lowest line, or an average of the lines when there is a discrepancy. In addition, the data also include information on the proportion of bets placed on the two teams in each contract from three sportsbooks: Pinnacle, 5Dimes, and BetCRIS. These three sportsbooks are collectively considered the "market setting" sportsbooks that dictate pricing in the U.S. market.[13]

The data offer some unique advantages relative to previous studies on sports betting. The data are more comprehensive, covering multiple sports over a decade (most studies cover a single sport over a few years), and uniquely provide multiple contracts on the same outcome of the same game – namely Moneyline and Spread contracts, which are key to our identification strategy to isolate the relevance of risk to betting prices. In addition, the data contain the actual betting lines/prices to compute realized returns to betting. Many studies (Gandar et al. (1988), Avery and Chevalier (1999), Levitt (2004)) use newspaper-sourced, composite, or even survey-based betting lines that are not real transaction prices. Using real returns from transactable prices allows us to quantify economic magnitudes and compare them to real returns in financial markets related to the same phenomenon.

[13] In sports betting parlance, market setting means that if one of the three big sportsbooks moves their line, other sportsbooks will follow, "moving on air," even without taking any significant bets on the game.

3 Empirical Analysis of Betting Contract Returns

We document three novel empirical facts in betting markets that serve as the cornerstone for the analysis, and relate each of these facts to analogous facts in financial markets.

3.1 Three Facts about Betting Contract Returns

First, we document the Favorite-Longshot Bias in our unique setting and data. We find that in the cross-section of Moneyline contracts across games, riskier bets (bigger underdogs) earn significantly lower average returns relative to less risky bets (bigger favorites). We sort Moneyline contracts into deciles based on the Moneyline, with decile 1 corresponding to extreme underdogs and decile 10 corresponding to extreme favorites.[14] Figure 1 plots the average return in each decile. The average returns are monotonically decreasing in risk. Less risky contracts earn significantly higher average returns than more risky contracts. The biggest underdogs (riskiest contracts) in decile 1 have average returns of -22.60%. Contracts in decile 10, which correspond to betting on the biggest favorites (the safest contracts), earn an average return of -0.51%. The fact that all returns are negative is due to the vig or transaction cost in betting markets, which at 5% is substantially punitive that average returns are negative.

The pattern of returns by risk has a direct parallel in financial markets. In the cross-section of stocks, riskier stocks tend to earn lower risk-adjusted returns than less risky stocks. In the second panel of Figure 1, we plot the CAPM alphas (the risk-adjusted returns) of US stocks sorted into decile portfolios based on their risk, measured by market beta (estimated using daily return data over the prior year as in Frazzini and Pedersen (2014)), at each point in time, drawing our numbers from Frazzini and Pedersen (2014). The pattern in the plot is similar to the pattern in sports betting. Stocks with the lowest betas substantially outperform, earning a monthly CAPM alpha of 52 basis points (bps) per month, while stocks with the highest betas underperform substantially, earning a CAPM alpha of -10 bps per month.

Second, we document that in the cross-section of contracts written *on the same game*, riskier contracts earn lower average returns and less risky contracts earn higher average returns. Figure 2 plots the average return of Spread and Moneyline contracts on the favorite and the underdog in each game, including error bars corresponding with 95% confidence intervals for the average return for each game. The riskiest type of contract (betting on the underdog in the Moneyline) earns an average return of -8.5 percent. The least risky contract (betting on the favorite in the Moneyline) earns an average return of -1.5 percent. The two intermediately risky contracts (which have the same risk), the spread contracts on the underdog and favorite, each earn average returns of -4.1 and -4.8 percent, which are not statistically different from each other. This

[14] We obtain nearly identical results by sorting on the Spread line rather than Moneyline, since the lines have a nearly perfectly monotonic relationship within each sport. We discuss this fact further when interpreting the results.

result shows a clear and clean test that idiosyncratic risk matters for pricing. While the Moneyline favorite versus Moneyline underdog face different risks, the Spread favorites and underdogs do not. Yet, all other characteristics associated with the favorite and underdog teams – expected point outcomes and distribution, sentiment, information, etc. – should be embedded in *both* contract types, as should any other types of preferences (such as a preference to bet on the home team or a preference to bet on big market teams). The only difference is the risk in payoffs and payoff amounts. Hence, the fact that the Spread favorite and Spread underdog provide the same average returns, while the Moneyline returns are decreasing in risk, identifies clearly that risk is being priced in this market *unrelated to any other feature of the game*. The key feature of this clean identification is that the two contracts are written on the same outcome of the same game, but where one contract varies the riskiness of payoffs while the other equates the risk of bets on both sides of the contract as well as across games.

The result above has a parallel in options contracts, though the interpretation is much cleaner in betting markets. Riskier (out-of-the-money) options earn lower risk-adjusted returns than less risky (at-the-money) options written on the same underlying stock price. In the second panel of Figure 2, we plot the delta hedged excess returns of one-month maturity individual equity options and equity index options, sorted into five buckets based on the absolute value of their delta, drawing our numbers from Table III of Frazzini and Pedersen (2020). The portfolios are ordered in terms of "moneyness", from "deep out of the money" (the riskiest option contracts) to "deep in the money" (the least risky option contracts). The plot shows that delta-hedged returns are decreasing from the deep out of the money portfolio (monthly delta hedged excess returns of -17.26% and -29.23% for equity options and index options) to the deep in the money portfolio (monthly delta hedged excess returns of -3.36% and -1.26%). Frazzini and Pedersen (2020) attribute this return pattern to the embedded leverage of the options contracts, where investors seeking leveraged exposure buy the (riskier) out of the money options, thus lowering their equilibrium return. This explanation, however, cannot explain the sports betting results because the sports contracts are idiosyncratic. Leverage exposure is about systematic risk exposure that is compensated in equilibrium. There is no such exposure in betting markets, hence an alternative explanation must be present.

Third, we document that in Spread contracts, where there is no cross-sectional heterogeneity in the risk of contracts, the average returns of different contracts do *not* appear to vary with the likelihood that the favorite wins the game. We sort spread contracts into deciles based on the Moneyline, with decile 1 corresponding with extreme underdogs and decile 10 corresponding with extreme favorites. This is the exact same sort as before, but now instead of betting on the Moneyline, we bet on the Spread contract on the *same games*. Figure 3 plots the average return of contracts in each decile. The returns in the deciles range from -2.8% to -6.3%, but the returns are not statistically different across any of the deciles and there is no systematic pattern. To the extent there is a pattern, it is in the opposite direction of the Favorite-Longshot Bias, and can be resolved completely by comparing bets made on home teams across games and bets made on away teams across games (see Appendix Figure C.8).[15] The lack of any return pattern indicates that it is the *risk* of the contract that matters to bettors in the favorite-longshot bias. For example, this result rules out that favorites versus longshots are mispriced, since such mispricing should show up in the Spread contract as well as the Moneyline contract. Similarly, this result also rules out that non-monetary preferences for certain characteristics of betting contracts (e.g., a preference to bet on the home team or on specific teams) drive the Favorite-Longshot bias, since such preferences should also be reflected in a Favorite-Longshot bias in Spread contracts. Put differently, the only difference between Figure 3 (Spread contracts) and Figure 1 (Moneyline contracts) is the risk of the contract. Everything else remains exactly the same because these are two different contracts on the same outcome of the game. Hence, we have a controlled experiment betting on the same outcome from the same contest, where the only variation is the risk of the contract. The stark difference in the results points to risk, in this case idiosyncratic risk, being the only difference between the two bets. This result on Spread contracts does not have a clear parallel in financial markets. However, it is a useful placebo test for our setting that rules out a host of potentially confounding factors.

3.2 Quantifying the Results: The Betting Implied Volatility Surface

Each of the betting contracts we study is a bet on the difference between the number of points scored by each team, with the bet paying off if the point-differential exceeds a certain threshold, and offering zero payoff otherwise. Here, we use data on the underlying point-differentials to synthesize and quantify our results in terms of an *implied volatility surface*, which is analogous to the implied volatility surface studied in options markets. Framing our results in this way, we can quantify the return differences we observe in terms of observable fundamentals, and compare the results in betting markets with evidence from options contracts in equity markets.

To construct the implied volatility for each contract, we use two ingredients. The first is an estimate of the objective probability that a contract pays off, p. The second is the market "implied probability" of a contract paying off, $\hat{p}$, embedded in the contract price. Using p and $\hat{p}$, we calculate the implied volatility of a contract, $\hat{\sigma}$, as the value that satisfies Equation (4):

$$F\left(F^{-1}(p;\sigma);\hat{\sigma}\right) = \hat{p} \tag{4}$$

where F is an assumed cumulative distribution function of the point differential minus the spread line, and σ is the standard deviation of F. For each sport, we use the Normal CDF for F, and we estimate

[15] The pattern is explained by the fact that there are systematically lower returns for betting on the home team.

σ using Maximum likelihood. We calculate the implied probability, $\hat{p}$, directly from contract prices as the probability of a contract payoff that sets the contract's expected return equal to the expected return of a Spread contract with 50% chance of paying off. To estimate the objective probability, p, we run a kernel regression of Moneyline contract returns on the Moneyline and Spread line of the contract, and use this regression to estimate the expected return of each Moneyline contract. We back out p from the expected returns and the price of each contract. We find that this procedure produces reliable estimates of win probabilities. We further discuss our methodology and present evidence of the accuracy of the estimated win probabilities in Appendix A.[16] Although our approach uses the Normal CDF to calculate implied volatilities, this is done purely to capture the magnitudes in an interpretable way. Our calculations do not rely on the assumption that the point-differential minus spread line follows an identical normal distribution across games. Rather, our methodology imposes no parametric assumption on the true distribution of the point-differential minus spread line, but simply uses the Normal distribution as a convenient tool to express the magnitude of the pricing patterns we observe in terms of standard deviation.[17]

For this exercise, we exclude contracts where p is between 0.45 and 0.55, because our methodology is not particularly well-suited to handle win probabilities very close to 50%. Note that Equation (4) has an undefined solution at $p = 0.5$. This restriction excludes 31% of contracts. However, our interest is primarily in studying contracts with low and high values of p.

We compute implied volatilities for each Moneyline contract. For each sport, we sort contracts into one of 90 equally spaced bins based on the estimated win probability of the contract (i.e., one bin for contracts with win probability between 0% and 1%, a bin for win probability between 1% and 2%, etc.). Figure 4 plots a scatterplot of the average estimated win probability versus the average implied volatility of each bin for Moneyline contracts in each sport. We refer to these plots as "implied volatility surfaces", using the options market terminology, as they relate the moneyness of the contract (the probability of the contract paying off) with the implied volatility in the contract. For each sport, the figure reveals a striking volatility *smile* or *smirk*, that is reminiscent of the smile and smirk observed in options implied volatility surfaces. Implied volatilities are higher on average for contracts that have a high probability of paying off and for contracts that have a very low probability of paying off. There also appears to be some asymmetry in the plot around

[16] Our approach is broadly similar to the Lowess smoothing approach that Snowberg and Wolfers (2010) take to map from horse betting odds to probabilities. Additionally, we omit non-payoff contract characteristics in these regressions. While non-payoff characteristics have some explanatory power, they do not appear related to our main quantity of interest here, the relative performance of riskier versus less risky contracts.

[17] An alternative way to construct implied volatilities is to calibrate the point-spread distribution using the assumption that the point spread minus spread line is *iid* Normal across games, and use the calibrated distribution to find the implied volatility that matches the implied win probability for a game. This alternate method heavily relies upon the assumption of normality, whereas the method we use only uses it to quantify the magnitude of implied probability distortions in terms of the standard deviation of a Normal distribution. We empirically analyze how well the Normal distribution fits the data in Appendix C.1. The fit is reasonable, though the non-parametric estimates do a substantially better job of capturing win probabilities.

50% in the implied volatilities for favorites versus underdogs, suggesting that bettors treat gains and losses differently.

To provide context for the magnitude of the implied volatility surface smile, we compare it to the implied volatility smile in options markets. Panel A of Table 1 displays the average implied volatilities for contracts expected to pay off approximately 5-15% of the time and 85-95% of the time. We compare these implied volatilities with the sample volatility of the point-spread distribution to calculate a measure of the implied volatility "premium" for the contracts. Contracts expected to pay off 5% to 15% of the time have implied volatilities that are 1.4% (for NBA contracts), 12.4% (NFL), 11.3% (NCAAB), and 12.8% (NCAAF) higher than the true sample volatility, with an average premium of 8.6% across sports. Contracts expected to pay off 85% to 90% of the time have implied volatilities that are 12.4% (NBA), 17.4% (NCAAB), 23.7% (NCAAF), and 24.8% (NFL) higher than the true sample volatility, with an average premium of 19.6% across sports.

For comparison, Panel B of Table 1 presents the simple average implied volatility for "standardized" 10 (and -10) delta and 90 (and -90) delta call (and put) options with one month maturity for a set of 13 equity indices. These correspond roughly to payout probabilities of 10 and 90%, respectively. Panel C presents the same quantities averaged across all individual equity options from the OptionMetrics IvyDB Implied Volatility Surface file.[18] The options at these delta values are deep out-of-the-money (low probability of paying off) and deep in-the-money (high probability of paying off). Because the implied volatility of options embed a variance risk premium, we capture the magnitude of the options smile by comparing the implied volatilities of the out-of-the-money options with implied volatilities of 50 (and -50) delta options, which have payoff probability of approximately 50%. The results presented in Panel B show that index options have a pronounced smirk, which is consistent with the intuition that index options are often used to hedge against the possibility of a market crash. Deep out-of-the-money put options and deep in-the-money call index options have a substantial implied volatility premium (on average 43.6% and 48% above the implied volatility of at-the-money options). Deep in-the-money put options and out-of-the-money call index options have implied volatilities that are slightly greater than the implied volatility of the at-the-money options (on average 4.5% and -0.9% below at-the-money options). The results presented in Panel C suggest that equity options have a more symmetric smile. Deep out-of-the-money call and put options have implied vols that are, on average, 17.0% and 26.7% higher than at-the-money calls and puts. Deep in-the-money options have implied vols that are, on average, 21.1% and 12.4% higher than at-the-money calls and puts.

The direct comparison between the implied volatilities of betting contracts and options contracts from financial markets is imperfect. There are many reasons for volatility smiles and smirks to exist in options

[18] We follow Frazzini and Pedersen (2020) in selecting our indices. OptionMetrics constructs implied volatilities for "standardized" delta values and maturities by interpolating the prices of traded options. The values produced, accordingly, do not correspond with any single traded option. See the appendix for more details.

markets, which primarily come from the difficulty in assessing the true probability distribution of underlying asset returns. Implied volatility calculations for option prices assume that stock prices are approximately log-normally distributed. The true distribution of asset prices is often thought to be fatter tailed, coming from different features of the asset return process, such as stochastic volatility and jumps. Exposure to aggregate risk may also play a role in the implied volatility smile. For example, deep-out-of-the money put options provide a hedge against market crashes, where the implied volatility smirk in index options may also, in part, be driven by risk premia associated with market crashes.

However, the volatility smile can also come from other features, such as non-traditional preferences or belief heterogeneity. These features are qualitatively consistent with the delta-hedged under-performance of risky, out-of-the-money options, as argued by Boyer and Vorkink (2014). The similar magnitude of the implied volatility smile in options on single name stocks provides more quantitative evidence of the similarity in the pricing of out-of-the-money betting contracts and options. The evidence further supports that non-traditional beliefs or preferences may play a role in options markets. Our evidence in betting markets is particularly useful because the simple state space and idiosyncratic nature of contracts makes the probability distribution easy to estimate, so misspecification is unlikely to drive our results. For instance, aggregate crash risk cannot be embedded in these contracts, and misspecification of the volatility process seems unlikely. Hence, our results point to other explanations driving this phenomenon, which we spend the rest of the paper exploring.

4 Interpreting the Evidence

We bring new evidence to the Favorite-Longshot Bias literature. We begin by discussing the implications of our results for explanations proposed for the Favorite-Longshot Bias, and then discuss the implications for a unifying theory that can explain the facts in financial markets and betting markets.

4.1 Implications for Theories of the Favorite-Longshot Bias

A central debate on the Favorite-Longshot Bias is whether it represents an informational inefficiency (see Thaler and Ziemba (1988) for a discussion of various explanations for the Favorite-Longshot Bias). Is the Favorite-Longshot Bias a result of the market incorrectly assessing the outcomes of sporting events, representing a mispricing? Or, is the pattern a reflection of preferences for lottery-like payoffs among bettors? In either case, is there a profitable opportunity for informed parties to exploit? Although earlier studies have found profitable systematic betting opportunities, more recent evidence indicates that the Favorite-Longshot Bias does not represent a profit opportunity, as the transaction costs from exploiting it exceed expected profits. In our sample, we also do not find reliable evidence of a profitable betting opportunity from the

Favorite-Longshot Bias after transactions costs.[19] The prohibitive transactions costs in betting markets may prevent arbitrage from eliminating price distortions, allowing them to persist and be detectable in the data.

The debate is still open, however, as to whether mispricing or preferences are driving the Favorite-Longshot Bias. Snowberg and Wolfers (2010) use the prices of exotic bets on the order in which horses finish as evidence that the market does not properly reduce compound lotteries, representing a misperception of horse win probabilities. Taking a non-parametric approach, Chiappori et al. (2019) find evidence that preferences and probability misperceptions both play a role in the FLB in horse racing.

4.1.1 Information Inefficiencies

We present evidence against the Favorite-Longshot Bias being a market-level informational efficiency. The key piece of evidence rejecting this interpretation is the third empirical fact we document: the absence of a FLB in Spread contracts *in the same games* where we do observe a Favorite-Longshot Bias in Moneyline contracts. This result indicates that market prices are accurate in forecasting the expected difference in points scored in the game. To the extent that market prices might be inaccurate in forecasting the point difference, these differences, which are embedded in contract prices, are not systematically optimistic about underdogs and pessimistic about favorites, which is required to reconcile the Favorite-Longshot Bias. This point is notable because which team wins and which team loses is entirely determined by the point difference of the game. Accordingly, the expected point difference is the key piece of information required to determine win probabilities and appropriate prices for Moneyline contracts. The fact that the expected point differences from the Spread lines on each game do not display a Favorite-Longshot Bias suggests that incorrect assessments of the expected outcome of the game are unlikely to explain the Favorite-Longshot Bias in the Moneyline contract. Hence, information inefficiencies do not appear to be driving the FLB.

4.1.2 Market Segmentation

An alternative explanation for the results is that the two markets (Moneyline and Spread) are informationally segmented. To address this alternative explanation, we show that Spread contract lines span nearly all relevant information regarding win probabilities embedded in Moneyline contracts. This result supports the claim that expected point differences are the key piece of information required to determine win probabilities. We run the following regression,

$$\mathbb{I}_{\text{home win},i} = \alpha_s + \beta_{\text{Spread,s}} \text{SpreadLine}_i + \beta_{\text{ML,s}} \log\left(\frac{1+y_{h,i}}{1+y_{a,i}}\right) + \epsilon_i \tag{5}$$

[19] Early studies find evidence of profitable betting strategies, such as Hausch et al. (1981). More recent studies (Levitt (2004) and Snowberg and Wolfers (2010)) do not find evidence of profitable betting opportunities. Admittedly, we did not attempt to devise betting strategies more complex than simple sorts of contracts based on the expected outcomes of games.

where for game i, $\mathbb{I}_{\text{home win},i}$ is a 0/1 indicator if the home team won the game, SpreadLine_i is the Spread line for game i, and $y_{h,i}$ and $y_{a,i}$ are the payoffs per dollar wagered associated with winning bets on the home and away teams in the game. The quantity $\log\left(\frac{1+y_{h,i}}{1+y_{a,i}}\right)$ is termed "the Moneyline ratio" and captures how favored the home team is versus the away team as implied by the Moneyline, with smaller values corresponding with the home team being more favored.[20] $\beta_{Spread,s}$ and $\beta_{ML,s}$ are separate regression coefficients for each sport, s. We standardize the independent variables within sports to have zero mean and unit standard deviation for ease of interpretation.[21]

We estimate three versions of the regression: 1) including only the Spread line as an independent variable, 2) including only the Moneyline ratio as an independent variable, and 3) including both. Panel A of Table 2 reports the results from the regressions. The first two columns show that the Spread Line and Money lines have very similar return predictability for wins, with nearly identical, highly-significant regression coefficients (ranging from -0.18 for the NFL to -0.22 for NCAA Basketball, indicating that a one-standard deviation change in the independent variable corresponds with an 18 to 22 percent increase in win probability for the home team). The third column shows the results from the multivariate regression. The coefficient on the Spread line remains significant for three of the four sports (insignificant for the NFL), while the coefficient on the Moneyline ratio is only significant for NCAA football. The R^2 for the multivariate regression is 20.87%, indicating that the multivariate regression adds little explanatory power over the univariate regressions, and that the Money lines and Spread lines capture the same predictive information for wins. In fact, the Moneyline and Spread are so correlated that including both in the regression creates a multicollinearity problem, which is why the significance of the coefficients varies. The F-statistic comparing the multivariate regression with the univariate Spread regression is 0.0367, implying we cannot reject the univariate regression in favor of the multiple regression, consistent with the Spread line and Moneyline capturing common information for wins.

We also plot a binned scatterplot of the Moneyline ratio versus home win percentages in Appendix Figure C.1, both including and not including a control for the Spread Line. The figure shows that win probabilities are monotonically decreasing in the Moneyline, but the second figure shows that the Spread Line captures almost all of the Moneyline's explanatory power. It is also apparent that alternative functional forms for the relationship between win probability and Moneyline implied probability are unlikely to explain the results.

Second, we show that changes in the Spread line and the Moneyline ratio from the open to the close of betting are highly correlated. In Panel B of Table 2, we show regression results of the open-to-close

[20] For readers familiar with decimal odds, the Moneyline ratio is simply the log of the ratio of the decimal odds on the two teams. The Moneyline ratio is 0.98 correlated with the Spread line across games in our sample.

[21] We estimate regression coefficients separately by sport because the point spread distributions differ across sports. Additionally, the transformation that we apply to Moneyline contracts is done to make the independent variable more linear. Alternative formulations have *worse* predictive power for wins.

changes in the Moneyline ratio on the open-to-close changes in the Spread line for each sport. The regression coefficients range from 0.53 (for NCAA Football) to 0.63 (NBA), with t-stats ranging from 34.08 to 114.49. Betting lines move because either volume demand on one side of the contract exceeds demand to take the other side, or information about game outcomes arrives between the open and close of betting (e.g., an injury to a key player). The results in Panel B of Table 2 show substantial commonality in Moneyline and Spread line responses to the arrival of information, indicating that these markets are highly integrated and not segmented informationally.

The near complete degree of overlap in the information captured by Spread lines and Moneylines regarding win probabilities, and the high correlation between changes in the Moneyline ratio and Spread line across games from open to close of betting, is inconsistent with Moneyline and Spread markets being informationally segmented. Segmentation is also not consistent with the fact that the same bookmakers are setting prices in the Spread and Moneyline markets simultaneously, and there is an active arbitrage market between them.

4.2 Implications for a Unifying Theory of the Facts

Given the analogous facts in both sports betting and financial markets, what types of unifying theories are able to simultaneously explain the findings in both settings?

Traditional asset pricing models in finance assume investors are risk-averse, have correct beliefs, and evaluate the risk of any asset in the context of their wealth portfolio. Under these assumptions, people diversify away idiosyncratic risk, meaning that the only priced risks in financial markets should be systematic, non-diversifiable risk, where equilibrium returns should compensate investors for bearing those risks. Theories for low-risk anomalies that rely on these assumptions suggest that capital market frictions, such as leverage constraints (Black et al. (1972) and Frazzini and Pedersen (2014)), can explain the facts, or alternatively suggest that low-risk anomalies may reflect compensation for some undiversifiable risk.[22] However, these theories are not able to explain the existence of any patterns in the cross-section of returns of sports betting contracts, which are bets on purely idiosyncratic risk. A unifying explanation of the facts requires departing from the traditional paradigm.

Our results provide some guidance for the types of 'behavioral' explanations that can explain the data. For example, incorrect market forecasts of the expected point difference do not appear to explain the Favorite-Longshot Bias, as shown by the lack of a bias in Spread contract returns, which suggests that a theory of excessive optimism of fundamentals is unlikely to provide a unifying explanation.

Relaxing the standard assumptions about correct beliefs and standard preferences, a few potential theories stand out in their ability to accommodate the betting market facts. Non-traditional preferences may be able

[22] As an example of the latter, Gormsen and Lazarus (2020) suggest that low-beta stocks have shorter duration cash flows than high-beta stocks, which may expose them to undiversifiable macroeconomic risks, as posited by Lettau and Wachter (2007).

to accommodate the results from sports betting markets and financial markets. For example, under *narrow framing*, agents evaluate gambles in isolation from the other risks they face elsewhere (Kahneman (2003)). This assumption can explain the existence of patterns in the cross-section of asset returns where there are no differences in exposure to aggregate risk. There is also evidence of narrow framing in financial markets (Barberis and Huang (2001) and Barberis et al. (2006)). Assuming narrow framing, a preference for lottery-like payoffs, for example, generated by the probability weighting component of Cumulative Prospect Theory, can explain less risky assets earning higher returns than riskier assets in both markets.[23] Note that while the literature on the Favorite-Longshot bias sometimes assumes that bettors have risk-loving Expected Utility preferences (a convex utility function), this assumption is fundamentally inconsistent with the evidence from financial markets, where investors demand compensation for risk, just not to the extent suggested by the standard models. A unifying, preference-based explanation suggests departing from the Expected Utility paradigm.

On the beliefs side, bettors may have heterogeneous beliefs, which can potentially rationalize the Favorite-Longshot bias if bettors sort into contracts based on their beliefs.[24] Survey evidence in financial markets also suggests the presence of substantial belief heterogeneity, which has been shown to explain pricing patterns in the cross-section of stocks. An alternative, belief-based explanation is that market participants accurately perceive the expected point difference, but they mistakenly believe that the distribution of point differences has more dispersion than it does in reality (e.g., they overestimate the variance of the point-difference distribution). Bettors with such mistaken beliefs would properly price Spread contracts, but would underestimate the probability of the favorite winning and overestimate the probability of the underdog winning in a manner consistent with the Favorite-Longshot Bias in Moneyline contracts.

We explore these potential explanations in the subsequent sections. In Section 5, we study a preference specification that features rank-dependent probability weighting and diminishing sensitivity, and endogenizes the choice to bet and the amount that bettors choose to wager. We find that the Favorite-Longshot bias is explained by similar parameter values of probability and diminishing sensitivity that can explain the financial markets facts. However, our results also highlight differences in the preferences that are able to explain the betting markets facts and the financial markets facts, particularly with respect to the treatment of losses.

[23] Probability weighting may deliver the patterns in betting returns in a way that the rational paradigm cannot, without narrow framing. As Barberis and Huang (2008) show, an individual that engages in probability weighting, and evaluates risks in the contexts of all other risks he faces may still demand positively, idiosyncratically *skewed* assets with negative returns (such as bets on extreme underdogs). Such assets can increase the overall skewness of the individual's portfolio, which the individual finds desirable.

[24] In addition to heterogeneous beliefs, bettors may also have heterogeneous preferences. For example, Chiappori et al. (2019) examine preference heterogeneity as an explanation for the Favorite-Longshot Bias at the horse race track and Andrikogiannopoulou and Papakonstantinou (2016) analyze heterogeneity in preferences by examining the behavior of individual bettors. Chiappori et al. (2019) find that allowing for preference heterogeneity has limited additional explanatory power for the Favorite-Longshot Bias. We do not explore preference heterogeneity, but recognize that it could also play a role in explaining these pricing patterns.

We then consider belief-based explanations in Section 6.

5 Preferences and the Favorite-Longshot Bias

In this section, we focus on the preferences of a bettor that has correct beliefs and is indifferent between wagering the optimal amount (given her preferences) on all betting contracts available. While there are a wide-range of bettors with potentially different motivations for betting (for example, recreational bettors versus professional 'sharps'), our focus is on a 'typical' or average bettor, who understands that betting is a negative expected return proposition and endogenously chooses to participate.

5.1 Preference Specification

Consider a wager of b on a simple binary outcome. The bet pays off with probability $(1-p)$, in which case the bettor receives yb dollars. With probability p, the bet does not pay off, and the bettor loses the b she wagered. We consider a bettor with reference-dependent preferences, who evaluates this bet by computing

$$V = (1-\eta)\underbrace{\left[p(-b) + (1-p)yb\right]}_{\text{Expected Utility}} + \eta \underbrace{\left[w(p)v(-b) + (1-w(p))v(yb)\right]}_{\text{Non-Expected Utility}} \tag{6}$$

where $w(\cdot)$ is a *probability weighting* function and $v(\cdot)$ is the *value* function. As advocated by Kőszegi and Rabin (2006), the formulation in Equation (6) has a "traditional" Expected Utility term, in addition to a non-Expected Utility term, with $\eta \in (0,1)$ governing the relative weight of the two in bettor's preferences.

The Expected Utility term in Equation (6) corresponds with a risk-neutral utility function. Given that bettors are unlikely to wager substantial portions of their wealth on any given bet, and that risk aversion is thought to be small for lotteries with stakes that are small relative to total wealth, risk-neutrality is a natural candidate for the Expected Utility component of preferences in our setting.

The non-Expected Utility term in Equation (6) corresponds with rank-dependent utility (Quiggin (1982)), and captures the reference-dependent features of the bettor's preferences. We use the functional forms proposed by Tversky and Kahneman (1992) for the value function, $v(\cdot)$, and the probability weighting function, $w(\cdot)$.

$$w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}} \tag{7}$$

and

$$v(x) = \begin{cases} x^\alpha & x \geq 0 \\ -\lambda(-x)^\alpha & x < 0 \end{cases} \tag{8}$$

where $\alpha \in (0,1), \gamma \in (0.28,1)$ and $\lambda \geq 1$.[25]

The value function is computed over gains and losses relative to a reference point. Here, we assume that the reference point is zero, which is the bettor's payoff if they do not take the bet.[26] $v(\cdot)$ is also concave over gains and convex over losses, with α governing the concavity and convexity of the value function over gains and losses. These characteristics of the value function lead the bettor to exhibit *diminishing sensitivity*, where for both gains and losses, the marginal effect of an additional gain or loss is smaller the further gains and losses are from the reference point. The value function also allows the bettor to display *loss-aversion* ($\lambda > 1$), where she may be more sensitive to small losses than small gains; however, as we discuss further, loss-aversion appears to be at odds with the decision to bet. The features of the value function can be drawn in contrast with the general properties of utility functions under Expected Utility, where people are thought to evaluate decisions based on their terminal wealth (rather than based on gains or losses) and the utility function is assumed to be concave everywhere (rather than concave over gains and convex over losses).

The bettor also uses transformed probabilities (captured by $w(\cdot)$), rather than objective probabilities, in the non-Expected Utility term of Equation (6). Probability weighting is motivated by evidence from psychology in the laboratory and field that people tend to systematically overweight low-probability events and underweight high-probability events (Fehr-Duda and Epper (2012), and Barberis (2013) summarize the evidence). The probability weighting function, and specifically, the parameter γ, govern how much the bettor transforms probabilities, with smaller values of γ corresponding with more over-weighting of tails of the distribution. The standard interpretation of probability weighting is not that the bettor does not know objective probabilities but rather that the transformed probabilities represent the bettor's "decision weights" that give additional weight to tail events in her utility function.

We highlight two points regarding the form non-Expected Utility preferences take in our implementation. First, the form of non-Expected Utility in Equation (6) can be considered a special case of Cumulative Prospect Theory (Tversky and Kahneman (1992)). Under CPT, bettors probability weight for gains is expressed by the function $w^+(\cdot)$ and for losses with the function $w^-(\cdot)$; our implementation corresponds with rank-dependent probability weighting of the form $w^+(1-p) = 1 - w^-(p)$. Our implementation of probability weighting is commonly used in other work (e.g., see Barseghyan et al. (2013) and Barseghyan

[25] γ is sometimes assumed to be in the range $(0,1)$, but it is only increasing everywhere for $\gamma \geq 0.279$. Values less than that can admit negative decision weights. See Ingersoll (2008).

[26] This is often referred to as "status-quo" reference dependence, where potential gains and losses are measured relative to the agent's wealth in the status quo, as specified in Tversky and Kahneman (1979). Other work assumes different reference points. For example, Kőszegi and Rabin (2006, 2007) propose that the reference point is an individual's *expected* outcome. In a more dynamic environment than assumed here, the reference point is sometimes assumed to be an individual's wealth in some past period and combined with an assumption about framing (e.g., an individual may compute potential gains and losses for an asset with respect to the time at which the individual purchased the asset). See Grinblatt and Han (2005), Barberis and Xiong (2009), Imas (2016), and Barberis et al. (Forthcoming) for analysis of this idea as it relates to stock price behavior, and Andrikogiannopoulou and Papakonstantinou (2019) for evidence from individual bettors. See O'Donoghue and Sprenger (2018) for a broader discussion of reference dependence.

et al. (2018)), but does have certain properties that substantially differ from those of other commonly used formulations of probability weighting (such as assuming $w^{+}(\cdot) = w^{-}(\cdot)$), particularly in how losses are treated. Second, and related, our implementation of probability weighting relies on rank-dependence, which has faced recent empirical challenges (Bernheim and Sprenger (2020)). We discuss these points, and how they relate to our results, in more detail below.

5.2 Betting Decision and Equilibrium Prices

We consider the betting decision of the bettor introduced in the previous subsection. She chooses the optimal amount to wager, b^*, to satisfy her first- and second-order conditions. Her first order condition is given by

$$0 = (1-\eta)\left(-p + (1-p)(y)\right) + \eta\left(-w(p)\lambda\alpha b^{*\alpha-1} + (1-w(p))\alpha y^{\alpha} b^{*\alpha-1}\right) \tag{9}$$

which yields an expression for the optimal wager:

$$b^{*\alpha-1} = \frac{1-\eta}{\alpha\eta}\left(-\frac{-p+(1-p)y}{-w(p)\lambda + (1-w(p))y^{\alpha}}\right) \tag{10}$$

where the right hand side of the expression must be positive for the optimal wager to exist. The bettor's second-order condition is given by

$$0 > \eta(w(p)\alpha(1-\alpha)\lambda b^{*\alpha-2} - w(1-p)\alpha(1-\alpha)y^{\alpha} b^{*\alpha-2}). \tag{11}$$

Since $b^* > 0$ (bettors can't short), and $\eta, \alpha > 0$, the second-order condition reduces to $w(p)\lambda - (1-w(p))y^{\alpha} < 0$. Jointly, the first- and second-order conditions imply that, under the assumptions of the model, a finite, positive betting amount for a contract exists when the returns to betting on the contract are negative, but where the bettor receives non-Expected Utility value from wagering on the contract.[27]

Labeling each contract $i = 1, \ldots, n$, we denote V_i as the utility that the bettor receives from wagering the optimal amount on contract i, b_i, which offers a payoff of y_i with probability $1 - p_i$. Following the literature on estimating preferences from aggregate betting data, we assert the equilibrium condition that the bettor is indifferent between all contracts offered at the equilibrium prices.[28] We implement this approach by asserting that each $V_i = V_s$, where V_s is the utility a bettor derives from wagering the optimal amount b_s on a spread contract that offers a potential payoff of $y_s = \frac{100}{110}$ per dollar wagered and has a 50% probability of paying

[27]Including risk-aversion in the Expected Utility component of preferences, as done in financial market applications, relaxes the requirement of negative expected returns.

[28]This condition can be motivated, for example, by following Jullien and Salanié (2000) and assuming that the bettor is part of a group of homogeneous bettors (who can co-exist with other bettor types), and that there is a bet placed on each contract offered on each game by at least one member of the group.

off. This condition can be expressed as,

$$
\begin{aligned}
0 &= V_i - V_s \\
&= (1-\eta)\left[p_i(-b_i) + (1-p_i)y_i b_i)\right] + \eta\left[w(p_i)v(-b_i) + (1-w(p_i))v(y_i b_i)\right] \\
&- ((1-\eta)\left[0.5(-b_s) + 0.5y_s b_s)\right] + \eta\left[w(0.5)v(-b_s) + (1-w(0.5))v(y_s b_s)\right]) \\
&=^* \frac{\eta^{\frac{1}{1-\alpha}}}{(1-\eta)^{\frac{\alpha}{1-\alpha}}}\left(\alpha^{\frac{\alpha}{1-\alpha}} - \alpha^{\frac{1}{1-\alpha}}\right)\left(\frac{((1-w(p_i))y_i^{\alpha} - w(p_i)\lambda)^{\frac{1}{1-\alpha}}}{(p_i-(1-p_i)y_i)^{\frac{\alpha}{1-\alpha}}} - \frac{((1-w(0.5))y_s^{\alpha} - w(0.5)\lambda)^{\frac{1}{1-\alpha}}}{(0.5-0.5y_s)^{\frac{\alpha}{1-\alpha}}}\right)
\end{aligned}
\tag{12}
$$

where the * step comes from substituting the optimal betting amounts for contracts into the expression. We also impose the additional condition, which is *not* imposed in the literature on the Favorite-Longshot bias, that V_i must be positive, implying that the bettor must derive positive utility from betting in the first place.

Equation (12) shows that in the interior of the range $(0,1)$, η, the relative weight the bettor places on the non-Expected Utility component of her preferences, does not matter for satisfying the equilibrium condition. It only enters into the equation (and utility at the optimum) via a multiplicative scaling factor $\frac{\eta^{\frac{1}{1-\alpha}}}{(1-\eta)^{\frac{\alpha}{1-\alpha}}}$. η plays a role in the amount of betting that a bettor chooses and the level of utility she derives from betting, but it *does not* affect her decision to bet on the extensive margin, and is not pinned down by her indifference between different contracts.[29]

5.3 Evaluating the Parameters of Interest

Our general approach to study the preference parameters of interest, $\Theta = (\alpha, \gamma, \lambda)$, is as follows. For a given choice of Θ, we construct the set of contract payoffs, denoted $\{y_{i,\Theta}\}_{i=1}^n$, implied by Θ and the loss probabilities across games, $\{p_i\}_{i=1}^n$. We evaluate a given Θ by observing how the expected returns constructed using the implied payoffs compare with the expected returns observed in the data. We treat the loss probabilities as known, and use the estimated probabilities from Section 3.2 (the estimation methodology for these probabilities is discussed in more detail in Appendix A).[30]

We rewrite Equation (12) to provide a fixed point expression for the implied payoff that satisfies the equilibrium condition, $y_{i,\theta}$, for each contract i.

$$
y_{i,\Theta} = \frac{p_i - A_s^{\frac{\alpha-1}{\alpha}}\left((1-w(p_i))y_{i,\Theta}^{\alpha} - w(p_i)\lambda\right)^{\frac{1}{\alpha}}}{1-p_i} \tag{13}
$$

where $A_s \equiv \frac{((1-w(0.5))y_s^{\alpha} - w(0.5)\lambda)^{\frac{1}{1-\alpha}}}{(0.5-0.5y_s)^{\frac{\alpha}{1-\alpha}}}$. Given parameters Θ, and the loss probability p_i, we use Equation (13) to calculate the implied value of $y_{i,\Theta}$ for each contract (if it exists) by iterative methods. A solution

[29]One way to further study the value of η would be with additional data on the wealth of bettors and the dollar amounts that they choose to bet relative to their wealth.

[30]Our approach conceptually differs from the approach pioneered by Jullien and Salanié (2000) and developed further on the Favorite-Longshot bias (such as Gandhi and Serrano-Padial (2015)). In that approach, the probability of the winning horse is written as a function of the observed payoffs and the parameters of interest, and the objective is to find the parameter values that best match the empirically observed win probabilities, which is done via maximum likelihood.

need not exist for $y_{i,\Theta}$ for all assumed parameter values and loss probabilities.

We evaluate a given choice of Θ by constructing the mean squared error of the expected returns implied by Θ versus the expected returns estimated from the data.

$$\text{MSE}_\Theta = \frac{1}{n}\sum_{i=1}^{n}\left(\underbrace{(1-p_i)y_{i,\Theta}-p_i}_{\text{Implied Expected Return}} - \underbrace{((1-p_i)y_i-p_i)}_{\text{Realized Average Return}}\right)^2 \tag{14}$$

Note that this expression corresponds with evaluating the weighted mean squared errors of payoffs y_i, where the weight on observation i is $(1-p_i)^2$. This type of weighting of errors is desirable, as we expect low payoff probability contracts with high y_i to also have larger magnitude errors (and vice-versa for high payoff probability contracts with low y_i). When a given value $y_{i,\Theta}$ does not exist, we substitute a squared error of 1. Contracts must have negative expected returns in our model, and the expected return is bounded below at -100% (a bettor loses their wager with probability 1). Hence, this treatment has the economic interpretation of Θ yielding a "maximally incorrect" estimate of the expected return when it does not provide a feasible solution for $y_{i,\Theta}$.

5.4 What Parameters Can Fit the Data?

To get a sense of what parameter values may be able to fit the data, we form a grid of values for $(\alpha,\gamma) \in [0.5,1]\times[0.5,1]$ (which broadly correspond with the range of values for α and γ found in experimental settings), with values on the grid spaced in intervals of 0.01. For each (α,γ) pair on the grid and $\lambda \in \{1, 1.25\}$, we compute the proportion of contracts for which the model is able to provide implied payoffs that satisfies the equilibrium and optimality conditions for the given choice of parameters. We also compute the mean squared error of expected returns implied by the model for each set of parameters.[31] We plot the surface of valid contract proportions and the negated MSE values in Figure 5.

The top left panel in the figure plots the proportion of contracts for which the model is able to calculate valid payoffs for a given (α,γ) pair, for $\lambda = 1$ (which corresponds to no loss-aversion). Points in red on the surface indicate (α,γ) pairs for which the model is able to calculate valid payoffs for at least 99% of contracts. The plot indicates that the model is able to generate valid payoffs for at least 99% of contracts for a band of (α,γ) pairs that lie along the line $\alpha=\gamma$ and are less than 0.8. The top right panel in the figure plots the same quantities, for $\lambda = 1.25$ (which corresponds with moderate loss-aversion). For each (α,γ) pair, the corresponding proportion of contracts for which the model computes valid implied payoffs is lower

[31] The optimality conditions imply that contract expected returns must be negative, and that prospect theory utility must be positive. This provides us with necessary and sufficient conditions to evaluate if a choice of parameters Θ generates a valid payoff that satisfies the equilibrium and optimality conditions given the loss probability for a game. Particularly, for game i and a given Θ, a valid payoff exists if and only if the largest potential payoff that has negative returns (calculated using the loss probability) has a positive prospect theory value and provides greater utility than the spread contract.

than the $\lambda = 1$ case, and the model is only able to generate valid payoffs for more than 99% of contracts for a considerably smaller set of values.

The bottom left panel in the figure plots the negated MSE values against values of α and γ, for $\lambda = 1$. Points in red on the surface have an MSE < 0.014, while all other points are colored in gray. The observations with MSE lower than 0.014 all lie close to the diagonal line $\alpha = \gamma$, with values of α and γ less than 0.7. The bottom right panel plots the MSE surface for $\lambda = 1.25$, and once again reveals that the model fit is considerably worse when introducing loss-aversion. For each point on the grid, the MSE is higher than it is for the corresponding point with $\lambda = 1$.

Figure 5 indicates that *both* diminishing sensitivity, captured by α, and probability weighting, captured by γ, are important for explaining the data. Providing a strong fit for the data requires parameter values for both that are less than 0.8. Additionally, the parameter value for each feature is closely related to the parameter value of the other feature, as indicated by the fact that the model performs best for values of α and γ that are close to each other. Lastly, the figure also demonstrates that loss-aversion ($\lambda > 1$) reduces the ability of the model to explain the data. We provide more economic intuition for these results after first addressing whether the model can quantitatively fit the data.

5.5 How Well Does the Model Capture the Favorite-Longshot Bias?

To more closely study the ability of the model to explain the betting markets facts with a reasonable parametrization, and in a manner consistent with the evidence from financial markets, we fix the probability weighting parameter, using values of $\gamma = 0.65$ and $\gamma = 0.5$, and calibrate the corresponding values of α and λ that provide the lowest mean squared error of expected returns.[32] Probability weighting is well-studied, with substantial experimental and field evidence. Tversky and Kahneman (1992) report a value of $\gamma = 0.65$, and subsequent evidence confirms that similar values are able to explain the facts in financial markets, so the choice of $\gamma = 0.65$ seems like a natural starting point. Evaluating $\gamma = 0.5$ also provides another point of comparison with more extreme probability weighting (that still lies within reasonable values reported in experimental evidence, for example see Booij et al. (2010)), which also allows us to better understand how α varies with γ, as the two are closely related.

Table 3 reports the results from the estimation. Panel A reports estimates of α, the mean squared error (MSE) of expected returns, and the number of contracts for which the model is able to calculate valid payoffs (positive payoffs that satisfy the bettor's optimality conditions). Standard errors constructed from 2,000 bootstrap samples are reported in parentheses. The lower bound for loss-aversion, $\lambda = 1$, tightly binds

[32] α, γ, and λ are tightly linked and not well-identified from one another without additional data, which is why we choose a calibration approach rather than an unrestricted estimation. We discuss the economic relationship between these features of preferences in more detail after presenting the results here.

in both cases. The estimated values of α are 0.650 and 0.502, very close to the assumed values of γ. The model does a very similar job in explaining the data for $\gamma = 0.65$ and $\gamma = 0.5$, with similar mean squared errors (0.012 for both), and a similar number of contracts with valid payoffs (73,188 versus 73,187, out of a total of 73,218).

Panel B reports more details from the estimation, splitting contracts into deciles based on their Moneyline, with decile 1 corresponding to extreme underdogs and decile 10 extreme favorites. The panel also reports the fitted payoffs, expected returns, and proportion of contracts with valid payoffs across the deciles. For both values of γ, the model is able to capture the optimal decision to bet a positive and finite amount, calculating valid payoffs for more than 99.9% of betting observations in deciles 1 through 9, and 99.6% of observations in decile 10.

Moreover, the model does reasonably well qualitatively and quantitatively capturing contract prices and the Favorite-Longshot bias in expected returns. Focusing on $\gamma = 0.65$, the model slightly over-estimates the payoff that the bettor demands to bet on extreme longshots in decile 1 (the average value of $\log y_i$ is 2.15 in the data, versus 2.36 implied by the model), resulting in slightly higher expected returns for extreme longshots implied by the model (-16.2%) versus the data (-22.5%). The model also underestimates the returns for more modest underdogs, but more closely captures the returns of moderate to heavy favorites. The patterns of implied returns are similar for $\gamma = 0.5$, which does a slightly better job capturing the returns of larger underdogs at the cost of doing a slightly worse job of capturing the returns of more moderate underdogs. We plot the model implied expected returns versus estimated expected returns by decile in Figure 6 for values corresponding with $\gamma = 0.65$.

5.6 Bet Sizes

The model also delivers implications for the amount bettors choose to wager on different contracts. Equation (10) derives the *optimal* amount a bettor chooses to wager, which is novel to previous models of the Favorite-Longshot bias that typically assume bettors wager a fixed amount across contracts.[33] For the majority of our sample (32,685 games), we have data on the proportion of bets placed on the two teams in a game, where we can compare the ratio of the average wager placed on each of the two teams against the model-implied values. Since we do not explicitly target bet size in our estimation, this comparison provides an independent assessment of the model by focusing on a feature of the data we do not use in our calibration.

[33] Weitzman (1965) estimates the preferences of a "Mr. Avmart" (average man at the race track), whose "wagers are allocated among the entrants in a race exactly in the same proportion as the entire crowd apportions its money among the various horses." The most recent generation of work on the Favorite-Longshot bias (Jullien and Salanié (2000), Snowberg and Wolfers (2010), Gandhi and Serrano-Padial (2015), and Chiappori et al. (2019)) assumes that bettors decide between wagering a fixed dollar amount on each betting contract available. Under a risk-loving Expected Utility function or a risk-neutral value function with probability weighting (e.g., as considered by Snowberg and Wolfers (2010)), or under the Cumulative Prospect Theory specification studied by Jullien and Salanié (2000), there is no finite, positive optimal wager. Bradley (2003) makes this point with reference to Jullien and Salanié (2000).

For a given game, assuming that the bookmaker takes no risk, market clearing for the Moneyline contract requires that the payoffs to winning bettors come from the pool of wagers. This can be formally expressed as two market clearing conditions:

$$(1 + y_h)\, n_h b_h = n_h b_h + n_a b_a \tag{15}$$

$$(1 + y_a) n_a b_a = n_h b_h + n_a b_a \tag{16}$$

where n_j is the number of bets placed on team j, b_h and b_a are the *average* amount per bet wagered on the home and away teams, and y_h and y_a are the payoffs offered for winning bets on the home and away team in a game. We can re-write these conditions to derive the expression,

$$\frac{b_h}{b_a} = \frac{(1 + y_a)\, n_a}{(1 + y_h) n_h}. \tag{17}$$

The proportion of bets in our data directly allows us to compute $\frac{n_h}{n_a}$, which in turn allows us to estimate the ratio $\frac{b_h}{b_a}$ implied by Equation (17).

The first panel in Figure 8 plots a binned scatterplot of the log Moneyline ratio ($\frac{1+y_h}{1+y_a}$) against $\log\left(\frac{b_h}{b_a}\right)$ (the "log bet size ratio") where games are grouped into twenty equally spaced bins based on the log Moneyline ratio. The points in dark blue correspond with the Moneyline ratio estimated using observed Moneyline payoffs and the bet size ratio implied by Equation (17). The plot shows that the bet size ratio is decreasing in the Moneyline ratio, where bettors tend to wager more on the home team the more favored the home team is (as higher Moneyline ratios correspond with less favored home teams). The points in orange are the corresponding model implied values, using the parameter values $(\alpha, \gamma, \lambda) = (0.65, 0.65, 1)$. The model-implied values capture the decreasing relationship between the Moneyline ratio and the bet size ratio. In the region of games where the home team is favored (the log Moneyline ratio is negative), the model-implied values also closely match the slope of the relationship between the bet size ratio and the Moneyline ratio in the data. However, when the home team is less favored, the model-implied values display a steeper slope than in the data. The model-implied values for the Moneyline payoffs and bet size ratio also cover a wider range of values than those observed in the data.

The second panel in Figure 8 plots a binned scatterplot of the log bet size ratio estimated using the betting volume data against the model-implied bet size ratio for each game. The plot reveals that the model-implied bet sizes are monotonically increasing with the bet size ratio, and correspond reasonably well with the bet size ratio estimated from the volume data, with points lying very close to the 45 degree line.

One limitation of our calculation is that it requires the assumption that bookmakers take no risk in betting markets. This assumption is based on the idea that bookmakers primarily seek to profit from the "vig", or implicit transaction cost embedded in the prices of bets. This assumption appears to be true on average in practice (Moskowitz (Forthcoming)), though it need not hold in each game. To test whether this assumption matters and provide a robustness test of our results using a completely out-of-sample dataset, we obtain betting contracts from another betting market – Betfair – that is an exchange and has no bookmaker. Using betting data for a sample of soccer matches from 2006 to 2011 from Betfair, one of the world's largest betting exchanges, we measure the average wager size for a sample of games without having to make any assumptions about bookmaker behavior.[34] This test not only rules out the influence of bookmakers or assumptions about their behavior on our results, but also provides a completely out-of-sample test on independent data from a completely different sport from another venue.

The first panel of Figure 8 adds a plot of the binned scatterplot of the log Moneyline ratio versus the log bet size ratio in the Betfair sample in light blue. The plot shows that the same decreasing relationship between the bet size ratio and Moneyline ratio persists in the Betfair sample. Moreover, the slope of the relationship between bet sizes and Moneylines in the Betfair sample mirrors the slope of the relationship implied by the model even more closely than our data. To the extent that bookmakers may take some limited risk to accommodate bettor preferences, for example bettor preferences to place larger wagers on the home team, the Betfair sample may even more accurately capture the average size of wagers made by bettors. The Betfair sample provides strong out-of-sample evidence consistent with the model's predictions.

Figure 8 suggests that the model qualitatively and quantitatively captures the fact that bettors optimally make larger wagers on more favored teams. This finding is notable given that accurately capturing the bet size is not explicitly part of the objective function of the model. The ability of our model to capture patterns in the size of the average wager across games provides additional evidence in favor of our model's ability to explain the data.

5.7 Probability Weighting and Diminishing Sensitivity

Having calibrated the model, we seek to better understand what role the parameters in the model play in explaining the results. The two main parameters we focus on are the diminishing sensitivity parameter, α, and the probability weighting parameter, γ.

The probability weighting function increases the weight that the bettor gives to *unlikely* events in her decision, and decreases the weight that the bettor gives to *likely* events. The parameter γ governs the amount of probability weighting, where lower values generally correspond with increasing the weight that tail events play in a person's decision. Relative to the benchmark of objective probabilities, this generates a preference for lotteries with skewed payoffs, which in turn can help explain the lower equilibrium return the bettor

[34] Details on the sample of games in the Betfair sample are described in Appendix C.3. We thank Angie Andrikogiannopoulou for graciously providing these data.

demands for underdogs that offer high payoffs with low probability. As Barberis and Huang (2008) note, probability weighting can generate a positive demand for assets with negative returns that have positively skewed payoffs (like the underdogs in our sample), so probability weighting also helps explain why bettors may enter into the betting market on the extensive margin, and wager on underdogs in the first place.

Diminishing sensitivity, governed by the parameter α, leads the bettor to worry less about marginal losses the further she is from her reference point (here, her initial wealth), with smaller values of α corresponding with increased diminishing sensitivity. In the context of bets on the favorite, increased diminishing sensitivity leads the small, but high probability gains from betting to look more attractive relative to the possibility of losing the wager, to the point where the bettor gets positive utility from the bet. In the absence of diminishing sensitivity, bettors would not derive positive utility from betting on favorites, which offers negative returns that are only made less attractive when including probability weighting.

In our specification, there is a tension between probability weighting, which makes underdogs more attractive and favorites less attractive, and diminishing sensitivity, which makes favorites more attractive and underdogs less attractive. Overall, the two features are linked tightly, and increasing the amount of probability weighting also increases the amount of diminishing sensitivity required to explain the data. The relationship between probability weighting and diminishing sensitivity also helps to better understand some of the results in Table 3. For example, increasing probability weighting can potentially provide a better fit for the returns of extreme underdogs in decile 1, but doing so comes at the cost of being able to compute payoffs that satisfy the optimality conditions for the extreme favorites in decile 10. This relationship between probability weighting and diminishing sensitivity also explains why the fit of the model does not change substantially for difference choices of γ, when α is allowed to vary.

5.8 Probability Weighting, Loss-Aversion, and the Choice to Bet

In the non-Expected Utility component of preferences, we assume rank-dependent probability weighting (Quiggin (1982)). Losses relative to the reference point are weighted by $w(p)$, and gains relative to the reference point are weighted by $1 - w(p)$, where $w(\cdot)$ is the probability weighting function proposed by Tversky and Kahneman (1992). There are two notable points to make about this choice. First, this form of probability weighting has some differences from implementations of Cumulative Prospect Theory often used to explain financial markets facts, particularly in terms of the way that losses are treated. Second, recent experimental evidence casts doubt upon rank-dependence (Bernheim and Sprenger (2020)). Accordingly, we discuss how changing the assumed form of probability weighting may influence our conclusions.

First, note that the Tversky and Kahneman (1992) probability weighting function is not symmetric around 0.5, and crosses the line $w(p) = p$ at a point $\bar{p} < 0.5$. For $p < \bar{p}$, $w(p) > p$, and for $p > \bar{p}$,

$w(p) > p$. This property also holds for other commonly used probability weighting functions (e.g., the Prelec (1998) function). For our implementation, the implication of this property is that bettors underweight loss probabilities greater than $\bar{p}$ and overweight loss probabilities less than $\bar{p}$. Moreover, increasing the degree of probability weighting (lower γ) also corresponds with a lower $\bar{p}$, as illustrated in Figure 7. In effect, increasing probability weighting also carries an additional implication similar to decreasing loss-aversion, λ.[35] This, in turn, contributes to the model's ability to explain the decision to bet, including on Spread contracts, as we decrease γ.

These properties of our implementation can be contrasted with an alternative implementation of probability weighting commonly assumed in financial market applications of Cumulative Prospect Theory, where losses are weighted by $w(p)$ and gains are weighted by $w(1-p)$. Consider a bettor who evaluates a bet by computing

$$V = (1-\eta)\left[p(-b) + (1-p)yb\right] + \eta\left[w(p)v(-b) + w(1-p)v(yb)\right]. \tag{18}$$

The bettor is identical to the previous bettor in how he evaluates potential bets, except that he probability weights gains using $w(1-p)$, rather than $1-w(p)$. For $\lambda = 1$ (no loss-aversion), the previous bettor is willing to bet on a negative expected return Spread contract that has a 50% probability of paying off, while the present bettor is not. Additionally, for $\alpha = \gamma$ (close to what we find in our calibration), this bettor only receives positive non-Expected Utility value from betting if bets have positive expected returns, or if $\lambda < 1$.[36]

Would this alternative bettor choose to bet, if $\lambda \geq 1$? The answer is a qualified yes, on some contracts. This bettor will only be willing to wager on favorites, if diminishing sensitivity is stronger than probability weighting ($\alpha < \gamma$), or underdogs, if probability weighting is stronger than diminishing sensitivity ($\gamma < \alpha$), but not both. He is unwilling to bet on Spread contracts. Table 4 reports the proportion of favorite and underdogs contracts the bettor is willing to wager on for different parameterizations, given the estimated win probabilities and contract payoffs.[37] For $\alpha << \gamma$, the bettor is generally willing to bet on the favorite; for example, for $(\alpha, \gamma, \lambda) = (0.5, 0.65, 1)$ the bettor is willing to bet on 88% of favorites (and no underdogs). For $\gamma << \alpha$, the bettor is increasingly willing to bet on contracts on the underdog; for example, for $(\alpha, \gamma, \lambda) = (0.9, 0.65, 1)$, the bettor is willing to wager on 87% of underdogs (and no favorites). In a market populated

[35] The relationship between probability weighting and loss aversion in our implementation is conceptually related to the non-identification result in Barseghyan et al. (2013). They use the form of reference-dependent preferences proposed in Kőszegi and Rabin (2007), and show in that specification, loss-aversion and rank-dependent probability weighting are equivalent to a probability distortion of the form $\Omega(p) = w(p)\left[1 + \lambda(1 - w(p))\right]$.

[36] For $\alpha = \gamma$, the non-EU component of value is positive if and only if $(1-p)^\gamma (yb)^\gamma - \lambda p^\gamma b^\gamma > 0$. In turn, for $\lambda \geq 1$, restricting betting amounts to be positive ($b > 0$), this condition can only be satisfied if $(1-p)y - p > 0$, i.e., the contract has positive expected returns.

[37] In these calculations, we use the given contract payoffs for most contracts. For some contracts, which are estimated to have positive expected returns, we substitute the largest payoff for which the contract earns negative returns. We use the same estimated win probabilities as before in these calculations.

with bettors with preferences of this form, we may be able to partially explain the data if bettors have heterogeneous preferences, as studied by Chiappori et al. (2019), with bettors exhibiting more diminishing sensitivity wagering on favorites, and bettors exhibiting more probability weighting wagering on underdogs.[38] However, it is worth noting that introducing even mild loss aversion substantially reduces the willingness to bet – for $\lambda = 1.25$, the bettor is never willing to wager on more than 50% of contracts at the offered prices.

An alternative way to proceed is to allow λ to be less than one, i.e., the bettor may be *loss-tolerant* rather than *loss-averse*. This may be motivated by evidence that part of the overall US population appears to be loss-tolerant (Chapman et al. (2018), who suggest this may be true of more than half the population), or, alternatively, as a way to capture gambling preferences (for example, as in Conlisk (1993)), which may be domain specific. Loss-tolerance can help rationalize the decision of bettors to wager on negative expected return Spread contracts. To have similar explanatory power as the main specification we present, a substantial degree of loss-tolerance is required; the parameter values $(\alpha, \gamma, \lambda) = (0.65, 0.63, 0.86)$ deliver an MSE of returns of 0.012, similar to the model performance in the main specification.

5.9 Comparison with Evidence from Financial Markets

Returning to the initial motivation of our calibration exercise: are the set of preferences that capture the Favorite-Longshot bias in sports betting similar to those that are also able to explain the financial market low risk anomalies? Our evidence suggests that probability weighting and diminishing sensitivity may similarly feature in explaining the financial markets and betting markets facts. However, our calibrations also reveal differences, particularly in how bettors treat losses, relative to preferences used to explain the financial markets facts.

Given the tight relationship between γ and α, we fix the value of γ at 0.65 (and 0.5), and study the corresponding value of α. Studies find that a probability weighting parameter of $\gamma = 0.65$ is able to explain the pattern of risk and return in the cross-section of options (Baele et al. (2016)) and the patterns of risk-and return in the cross-section of stocks (Barberis et al. (Forthcoming)). The experimental values of γ reported by Booij et al. (2010) are similar in magnitude. We estimate the value of the diminishing sensitivity parameter, $\alpha = 0.65$. Booij et al. (2010) conduct a meta-study of experimental estimates of α. Most of the estimates lie between 0.5 and 0.95, and the average value they report is approximately 0.7. Barberis et al. (Forthcoming) use a value of $\alpha = 0.7$ in a dynamic model where bettors evaluate gains and losses in stocks relative to the purchase price that explains various stock market anomalies, including the relative under-performance of idiosyncratically risky stocks.

[38] It may not be straightforward to apply the approach of Chiappori et al. (2019) to endogenize the amount bettors choose to wager, as we seek to do here. They note "Whether nonparametric tests and identification like those we develop below could be constructed with endogenous bet amounts or would require information at the individual level is an open question."

Although we find similar values for α to those from other financial market applications, our model is static, and in our application of diminishing sensitivity, gains and losses are evaluated relative to the status quo.[39] The differences in the exact application of diminishing sensitivity in betting versus financial markets are natural, however, given the differences between the two settings. Bettors make decisions to bet hours (at most, days) before the outcome is revealed, making their wealth in the status quo a natural reference point against which to compute potential gains and losses. This is in contrast to financial markets, where individuals make investments that they may hold onto for months and years, and which they are able to continually reevaluate over that investment period.

A notable point is that while loss-aversion is one of the most well-known and well-studied components of Cumulative Prospect Theory, including loss-aversion in our setting reduces the fit of the model. For the particular application we are concerned with, the relative pricing of risk in the cross-section of risky assets, loss-aversion may not play an important role in explaining the facts in financial markets.[40] Additionally, recent evidence from a more representative sample suggests that there is substantial heterogeneity in loss aversion, and that perhaps half of the population does not demonstrate loss aversion (Chapman et al. (2018)). It is also possible that individuals with little loss aversion select into betting markets, or alternatively, that people behave in a less loss-averse way when gambling than when investing.

Overall, our calibration offers evidence of similarities in explaining the low-risk anomaly across financial and betting markets. In particular, our results suggest that probability weighting and diminishing sensitivity can reconcile the Favorite-Longshot Bias simultaneously with the low risk anomaly in financial markets.

5.10 Relationship with the Literature on the Favorite-Longshot Bias

The inclusion of both diminishing sensitivity and probability weighting is novel to the literature. For example, Snowberg and Wolfers (2010) assume a risk-neutral value function with no diminishing sensitivity, and estimate that very modest probability weighting (corresponding with approximately $\gamma = 0.9$ in the functional form we assume) is required to rationalize the Favorite-Longshot bias. However, when assuming a degree of

[39] Financial market applications of CPT are often applied in dynamic contexts, where diminishing sensitivity delivers the disposition effect that investors sell winners too quickly and hold on to losers too long as a result of them treating gains and losses relative to purchase price and caring about realized rather than "paper" gains and losses. In these applications, diminishing sensitivity, which leads to the disposition effect, can contribute to the lower returns of riskier stocks (Barberis et al. (Forthcoming)). This is different than the role diminishing sensitivity plays in our framework for static betting markets, where we find it leads bettors to demand higher returns for riskier bets. The different implications come from different reference points used to define gains and losses – in betting markets the reference point is zero and in financial markets applications it is the purchase price – *and* from using a static (betting markets) versus dynamic (financial markets) model.

[40] Baele et al. (2016) find that loss-aversion has little explanatory power for the cross-section of options returns and that probability weighting is the important component for generating the low returns of out-of-the money options. Similarly, decomposing the contribution of the components of CPT to the underperformance of idiosyncratically risky stocks, Barberis et al. (Forthcoming) report that probability weighting, along with diminishing sensitivity when evaluating gains and losses from the time an asset was purchased, are the primary contributors. Loss aversion, if anything, slightly detracts from CPT's ability to explain the low-risk anomaly. More generally, loss-aversion may potentially play a role in explaining the size of the equity risk premium (Benartzi and Thaler (1995) and Barberis et al. (2001)) and low stock-market participation (Barberis et al. (2006)), but appears less important for the relative pricing of risky assets.

probability weighting similar to the amount found in other economic applications (e.g., as summarized in Table 2a in DellaVigna (2018)), diminishing sensitivity is also required to make bets on the favorite attractive enough to bet on. Hence, our results also provide an economic application of diminishing sensitivity. As O'Donoghue and Sprenger (2018) note, while considerable focus is directed towards loss aversion in the study of reference-dependent preferences, diminishing sensitivity and its applicability across economic contexts is less well understood.[41] Diminishing sensitivity in our framework can help explain why bettors choose to wager on the favorite, despite negatively skewed returns (unattractive from the perspective of probability weighting) and a negative expected return in equilibrium.

Our specification for preferences in Equation (6) is also crucial for our model to explain betting behavior. Neither Expected Utility (EU) preferences alone ($\eta = 0$), nor non-EU preferences alone ($\eta = 1$), can generate positive and finite bet sizes. Under risk-neutrality (which we assume for the EU component of preferences), bettors never choose to wager on negative expected return contracts. More generally, with risk averse preferences, EU bettors optimally never choose to bet or with risk loving preferences, would bet an infinite amount. Similarly, with non-EU preferences, depending upon the chosen parametrization, the bettor will either not bet or bet an infinite amount. However, when a bettor cares both about the negative returns from betting and her non-EU value from betting, her focus on the return component modulates the otherwise infinite demand for betting coming from her non-EU preferences. Hence, our results also provide a unique perspective to the literature that seeks to assess whether risk-loving EU preferences or non-EU preferences do a better job of explaining the Favorite-Longshot bias (Jullien and Salanié (2000), Snowberg and Wolfers (2010), and Chiappori et al. (2019)). Our conclusion is that neither alone characterizes the data well, but by combining non-EU preferences with (weakly) concave EU preferences, we can capture bettor behavior in the data. Our differing conclusion comes from the fact that we endogenize the choice and amount to bet. Other studies estimate risk-preferences from aggregate betting data, which does not incorporate whether the choice to bet is optimal, and often assumes the same betting amount across contracts. As discussed in Section 5.6, the optimal betting amounts from the model are able to capture patterns in the size of average wagers across different contracts in the betting data.

6 Alternative Explanations for the Facts

We show that the empirical patterns we document are consistent with bettors having Cumulative Prospect Theory preferences. As we note earlier, there are two alternative, belief-based explanations that may also be

[41] One exception is in finance, where research focusing on the disposition effect, the tendency of investors to hold onto losers and sell winners first documented by Shefrin and Statman (1985), suggests that diminishing sensitivity may play a role. Barberis et al. (Forthcoming) study the role of diminishing sensitivity more broadly to a host of asset pricing anomalies. In the economics literature, Duraj and He (2020) propose a model of news utility where agents have diminishing sensitivity over the magnitude of news.

able to explain our results in betting markets. The first is that bettors may accurately perceive the expected point difference in a game, but mistakenly believe that the point distribution is more dispersed than it is in reality. A second alternative explanation is that bettors may have heterogeneous beliefs, and partition into betting contracts based on their beliefs. Here, we discuss and evaluate these alternative explanations. While we cannot completely eliminate either of these alternative explanations, we argue that CPT preferences likely provide a better characterization of the facts.

6.1 Misperceptions of the Point Difference Distribution

Misperceptions of the point-difference distribution may be able to explain the general pattern of returns, if bettors believe that the point-difference distribution is more dispersed than it actually is. This misperception leads bettors to believe erroneously that the probability of a favorite winning is lower, and the probability of an underdog winning is higher than reality. With such mistaken beliefs, bettors find bets on underdogs more attractive than bets on favorites, in a manner consistent with the Favorite-Longshot bias.

Conceptually, this bias operates similarly to the probability weighting, with the interpretation that bettors have mistaken beliefs. Just as with probability weighting, however, misperceptions alone are unable to capture bettor decisions to wager finite and positive amounts on negative expected return contracts. Hence, this bias alone is not a *sufficient* explanation for the betting market facts.

There are a few additional reasons that misperceptions are unlikely to play a dominant role in explaining these patterns. First, for each game i, consider $\hat{p}_i^f$ and $\hat{p}_i^u$, the perceived win probabilities for the favorite and underdog that set the (subjective) contract expected returns to be -4.5%, the empirically estimated average return for betting on a 'fairly-priced' spread contract. For a risk-neutral bettor with probability misperceptions (e.g., as considered by Snowberg and Wolfers (2010)), we expect $\hat{p}_i^f = 1 - \hat{p}_i^u$. In the data, we find that $\hat{p}_i^f - (1 - \hat{p}_i^u) = -0.93\%$ (statistically significant at the 0.1% level). What does this tell us? A risk-neutral bettor that is indifferent between betting on the favorite, underdog, and spread contracts on a game, on average, must underestimate the win probability of the favorite *more* than she overestimates the win probability of the underdog. This asymmetry is also seen in the implied volatility surfaces in Figure 4, where the implied volatilities are systematically higher for favorites than for corresponding underdogs. Put differently, a risk-neutral bettor with probability misperceptions can only be indifferent between betting on the favorite and underdog Moneyline contracts as well as a Spread contract, if and only if she uses inconsistent beliefs about the game's outcome probabilities to price the favorite and underdog contracts. This unappealing result suggests that rationalizing the data requires an additional feature.

Second, the distribution of point differences appears to be relatively constant across games. In Figure C.2, we plot the standard deviation of the point-differential minus Spread line of all games, sorted into deciles by the Moneyline ratio in each sport. The standard deviations are largely constant across deciles. There is no systematic pattern in standard deviations across the deciles (e.g., point-spreads being more or less dispersed in games where the home team is more favored). To the extent there are differences in the standard deviation across deciles, they are economically small compared to the implied volatilities shown in Figure 4 and Table 1. Notably, this is in contrast to the evidence in financial markets, where the standard deviations of asset returns vary substantially over time for the same assets, as well as across different assets. For misperceptions about the point distribution to matter, bettors would have to get the largely constant distribution of point differences consistently incorrect. This seems especially unlikely, given that, based on the evidence in Table 2, Spread and Moneylines capture nearly completely overlapping information for win probabilities, implying bettors recognize the point spread distribution is constant across games and treat the expected point difference as sufficient to determine win probabilities.

Third, we examine another contract written on the fundamental of each game – points scored – using the Over/Under contract, which is a contingent claim on total points scored. We show that market prices in the Over/Under market also accurately reflect the total number of points scored by both teams in games. We sort games into twenty bins based on the Over/Under line of the game, and plot a binned scatterplot of the Over/Under line versus the total number of points scored in Figure 9. All points lie on the 45 degree line, illustrating that the Over/Under contracts accurately capture point totals in games. The R^2 from the regression of the raw data (not binned) is 89.3%, suggesting a very good fit between expected points and actual points scored across all games in all sports.

The second panel in Figure 9 sorts games into deciles based on the Moneyline ratio, and plots the average Over/Under line and the average total score of games in each decile. The Over/Under lines are also accurate in capturing the total score of games regardless of how close or lopsided the game is expected to be. This evidence indicates that the betting market is very accurate at forecasting expected points scored, the key piece of fundamental information for all betting contracts. The fact that Spread and Over/Under contracts are not mispriced, and predict point differences (Spread) and point totals (Over/Under) extremely accurately, challenges the notion that the FLB in Moneyline contracts is due to erroneous understanding of the point distribution. For misperceptions of the point distribution to be the driver of the results, the market must consistently misperceive the relatively static point distribution across games, despite displaying extreme sophistication in forming expectations about the number of points each team will score in each game, which is a conditional random variable that varies substantially across games. It seems implausible that betting markets are so good at capturing a complex dynamic component of points scored while simultaneously and consistently misperceiving a very stable and static feature of points scored.

Overall, the evidence suggests that while it may be possible for misperceptions of the point distribution to play a role, they cannot characterize the betting data alone and seem unlikely to explain the facts.

6.2 Heterogeneous Beliefs

Belief heterogeneity is another potential explanation for the facts, and is a leading alternative explanation for the Favorite-Longshot bias. Previous studies link belief heterogeneity with the Favorite-Longshot Bias from a theoretical perspective (e.g., Ali (1977), Shin (1991, 1992), and Ottaviani and Sørensen (2009, 2010)). Gandhi and Serrano-Padial (2015) empirically estimate the degree of belief heterogeneity assuming a discrete choice environment, where the choice of which horse to bet on is isomorphic to a model of horizontally differentiated demand.[42] Green et al. (2020) suggest a slightly different, but related form of belief heterogeneity as an explanation for the Favorite-Longshot bias. In particular, they suggest that racetracks post morning-line odds that reflect a Favorite-Longshot bias, which are believed by gullible traders, and create a wedge between the valuations of these gullible traders and risk-neutral arbitrageurs.

In general, the logic behind the heterogeneous beliefs explanation for the Favorite-Longshot bias is that in fixed-odds markets, the dollar volume wagered on a team is increasing in how favored the team is. In equilibrium, only a small proportion of dollar volume is wagered on extreme underdogs, meaning that the low returns to betting on underdogs can be theoretically explained by the fact that only a small proportion of bettors with extreme beliefs choose to wager on underdogs. Note that with risk-averse bettors, heterogeneous beliefs can explain the decision to bet a finite and positive amount (as with the CPT model presented in Section 5), but bettors in such a model *mistakenly* believe that their bets have positive expected returns, while bettors in our model are aware that betting has negative expected returns.[43]

In the absence of data on betting volume, Gandhi and Serrano-Padial (2015) and Green et al. (2020) follow the common assumption made in the literature that bettors wager an exogenously fixed amount, which is assumed to be the same across contracts. With this assumption, one can pin down the distribution of beliefs using the equilibrium relationship between dollars wagered and odds, as Gandhi and Serrano-Padial (2015) do. Our data provide us with the proportion of the *number* of bets placed on each team, which allows us to test the logic behind a heterogeneous beliefs model as an explanation for the Favorite-Longshot bias. Our betting data provides an additional test of this explanation without assuming equal bet sizes across contracts.

Assuming the bookmaker takes no risk, the proportion of dollars bet on a team in a game (which we refer

[42] In a similar spirit, in recent work, Egan et al. (2020) use a revealed preference approach to estimate preferences and the distribution of beliefs about expected stock market returns using data on investor flows into ETFs that track the S&P500, exploiting differences in leverage across the ETFs.

[43] In the heterogeneous beliefs model, it is also possible that bettors receive utility from gambling and understand that their bets have negative expected returns, as Green et al. (2020) assume. However, this type of gambling preference can only explain the amount that bettors choose to wager if implemented in a preference specification similar to the one we present in the previous section. For example, if a bettor simply derives a fixed utility from gambling that makes participation sufficiently attractive, he would always optimally choose to wager an infinitesimally small amount when he bets.

to as *dollar market share*) is pinned down by the odds offered on the two teams. More favored teams should have a higher proportion of dollars bet on them, while less favored teams should have a lower proportion of dollars bet on them under this framework. Combined with the assumption of equal betting amounts across bettors, the Favorite-Longshot bias naturally emerges, with the marginal bettor choosing to bet on the underdog having more extreme beliefs.

We compare the estimated dollar proportion of bets with the proportion of the number of bets placed on a team in our data. We sort contracts into twenty bins based on the estimated dollar proportion of bets. Figure 10 plots a binned scatterplot of the average dollar market share of bets on the x-axis against the average proportion of bets on the y-axis for each bin. The proportion of bets is generally increasing in the dollar proportion of bets on a contract, though the reverse is true in the tails (for extreme favorites and extreme underdogs.) The relationship between the proportion of bets and the dollar proportion of bets is flatter than the 45 degree line. For example, while the dollar proportion of bets for contracts in the top and bottom bins (extreme favorites and underdogs) are approximately 90% and 10%, the average proportion of bets placed on contracts in these bins are 56% and 44%, respectively. Notably, the average returns for contracts in these bins are -0.40% and -24%, so that 44% of all bets are being placed on contracts with the most negative returns, though only about 10% of the *dollar* volume is being placed on these contracts. While theoretically, the Favorite-Longshot Bias may emerge with only a small proportion of bets on extreme underdogs (as noted in Gandhi and Serrano-Padial (2015)), our evidence suggests that many bettors bet on extreme underdogs. The evidence suggests that while in theory heterogeneous beliefs may help explain the Favorite-Longshot bias, the large proportion of bets on extreme underdogs with very negative returns is hard to rationalize in a heterogeneous beliefs model.

To illustrate this point, Appendix B presents a stylized model of risk-neutral bettors with heterogeneous beliefs about the expected point differential in each game, who partition into contracts based on their beliefs. We assume that the distribution of beliefs of bettors is the same across games. The model produces a non-monotonic pattern in expected returns, where the returns to betting on the most extreme underdogs yield higher expected returns than the returns to betting on more moderate underdogs. If a larger proportion of bettors choose to bet on more extreme underdogs, then the marginal bettor on such underdogs has less extreme beliefs than the marginal bettor in games with more moderate underdogs. Thus, belief heterogeneity has a hard time reconciling the increasing proportion of bettors choosing to bet on extreme underdogs when the belief distribution is the same across games. This result holds regardless of parametric assumptions about the shape of the distribution of beliefs.[44]

[44] A caveat to our conclusions on heterogeneous beliefs is that our analysis is conditional on the choice to bet on a game. In reality, bettors' choice sets involve all games offered on a date (as well as not betting), and bettors may sort across games based on the bets they find most attractive. This, in turn, may nullify the assumption that the distribution of beliefs is the same

We cannot, however, rule out belief heterogeneity as partially contributing to these patterns. Other studies have suggested heterogeneous beliefs matter in financial markets, and some form of heterogeneity is required for any betting or trading volume.[45] Therefore, it may be interesting to apply such models to understand why out-of-the-money options earn lower returns than in-the-money options. Similar to the Favorite-Longshot Bias, such a pattern can emerge under certain conditions if agents with heterogeneous beliefs about an asset's returns sort into options based on their optimism (Shefrin (2008)).[46] However, our results from both sports betting and financial markets appear most consistent with preferences playing a major role in explaining these facts. We add additional evidence to the idea that non-Expected Utility preferences may play an important role in option prices (Kliger and Levy (2009), Boyer and Vorkink (2014), and Baele et al. (2016)), and can simultaneously capture the low risk and Favorite-Longshot bias in financial and sports betting markets.

7 Conclusion

We use unique data in betting markets on two contingent claims written on the same game outcome to infer bettor preferences and beliefs. We use these inferences to bridge two well-known and similar empirical phenomena in financial markets and betting markets: in financial markets, risky assets earn lower risk-adjusted returns, and in betting markets the Favorite-Longshot bias represents a similar phenomenon. The pattern in betting returns disappears when comparing equal-risk contracts across the same games. When cross-sectional differences in risk are removed, we find no difference in returns, highlighting that risk, and not any other characteristic of the game or teams, is the chief attribute driving these patterns. Quantitatively comparing the returns in betting markets with the low risk anomalies in financial markets, we find very similar magnitudes.

Synthesizing these results informs which theories are able to simultaneously explain all of the facts. Rational theories, whether they rely on capital market frictions or the presence of undiversifiable risk, predict no relationship between idiosyncratic risk and returns, and cannot explain these findings. We calibrate a model with reference-dependent preferences to match the data, and find that similar parameter values for diminishing sensitivity and probability weighting used to match the low risk anomaly in equity and options markets are also able to rationalize the betting market facts. However, our analysis also highlights

across games. We consider some of the implications of a wider choice set on our results in Appendix B. Our analysis suggests that the choice sets of bettors are unlikely to explain our findings, however.

[45] There is a substantial literature documenting that belief heterogeneity in the presence of short-sale constraints can contribute to trading volume and can lead to assets being overvalued relative to fundamentals (Miller (1977), Harrison and Kreps (1978), Scheinkman and Xiong (2003), and Diether et al. (2002)).

[46] The pricing results in such a model are not consistent with the dominant, no-arbitrage paradigm in option pricing. However, there is evidence against no-arbitrage in options markets. For example, Garleanu et al. (2008) show theoretically and empirically that customer demand pressures, combined with risk-averse market makers that cannot perfectly hedge their inventory, may lead to no-arbitrage violations, such as the low returns of out-of-the-money options.

a difference between preferences that are able to rationalize betting market behavior and preferences that have been used to explain the behavior of investors in traditional financial markets, namely in how people treat losses. Alternative models, such as heterogeneous beliefs or misperceptions of probabilities, have more difficulty explaining the facts, particularly the unique betting volume data we examine. Overall, we find that non-expected utility preferences likely play an important role in explaining the facts.

We find that using betting market data in connection with financial market evidence is helpful to identify explanations that can fit both markets when seeking a unifying explanation for investor decision making. While the markets do differ along several dimensions, their commonality and, in particular, the commonality of similar return patterns, suggests that we can learn something about decision making under uncertainty from examining both markets simultaneously.

Tables and Figures

Figure 1: Risk and Return in the Cross-Section of Assets

The figure presents returns versus measures of risk in the cross-section of assets in both sports betting markets and financial markets. The first panel plots the average returns of Moneyline contracts, sorted into deciles based on the Moneyline. Decile 1 corresponds with contracts betting on the most extreme underdogs and Decile 10 corresponds with contracts on the most extreme favorites. The error bars correspond with 95% confidence intervals for the average return within a given decile. The second panel plots the CAPM Alpha (the monthly risk-adjusted return, in basis points) of US equities, sorted into deciles based on their CAPM Beta. Decile 1 corresponds with contracts that have the highest CAPM Beta (the riskiest stocks) and Decile 10 corresponds with the the lowest CAPM Beta (the least risky stocks). The numbers used in the plot are taken from Table III in Frazzini and Pedersen (2014) and the sample runs from January 1926 to March 2012.

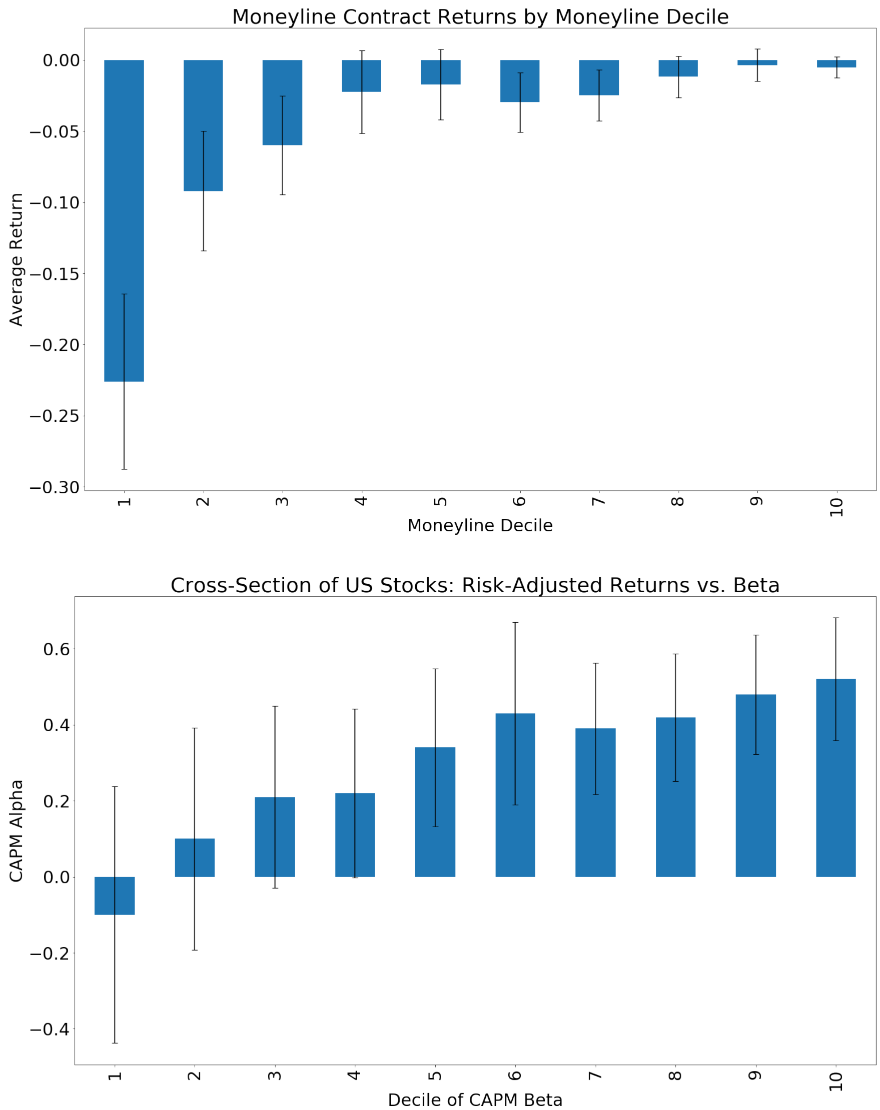


Figure 2: Risk and Return in the Cross-Section of Contracts

The figure presents the average return of contracts on the same asset, sorted based on characteristics related to their risk. The first panel plots the average returns of Spread and Moneyline contracts on the underdog and favorite of each game, with one column corresponding with each contract type. The error bars correspond with 95% confidence intervals. The second panel plots delta hedged excess returns, in percent terms, of one month maturity options sorted into one of five portfolios based on the absolute value of their Black-Scholes Delta values: Deep out of the Money (Absolute Delta of 0-20), Out of the money (Absolute Delta of 20-40), At the Money (Absolute Delta of 40-60), In the Money (Absolute Delta of 60-80), and Deep in the Money (80-100). Equity options correspond with all options on equities in the OptionMetrics Ivy DB, subject to a set of liquidity filters. Index options correspond with options on a set of 13 indices, also subject to liquidity filters. Each portfolio is rebalanced on the first trading day following expiration Saturday, and all options are value weighted within a given portfolio based on the value of their open interest. The data in the plot are taken from Table III of Frazzini and Pedersen (2020).

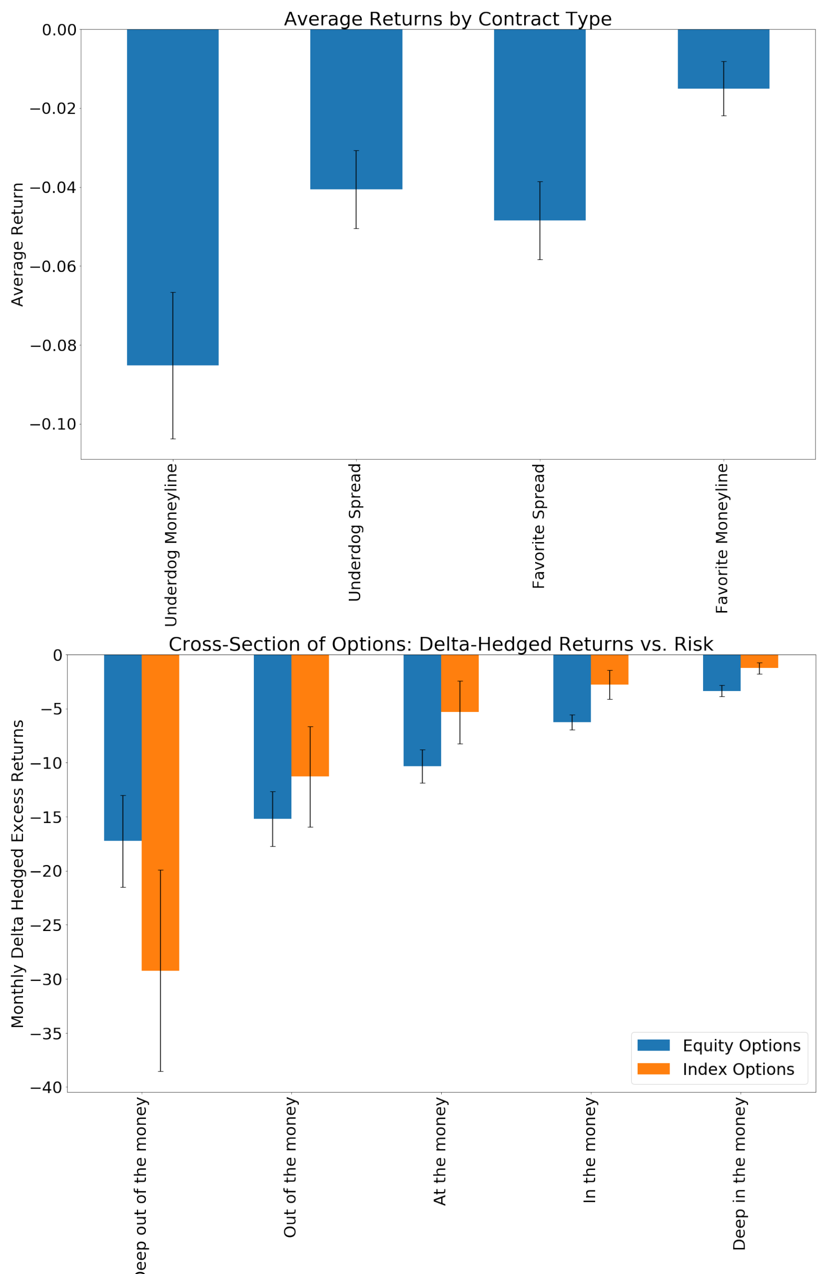


Figure 3: Average Spread Contract Return Across Games

The figure plots the average returns of contracts sorted into deciles based on the Moneyline. Decile 1 corresponds with contracts betting on the most extreme underdogs and Decile 10 corresponds with contracts on the most extreme favorites. The darker bars in the foreground correspond with Spread contract returns. The error bars correspond with two standard errors above and below the average return for Spread contracts in a given decile. The lighter bars in the background correspond with Moneyline contract returns, as presented in Figure 1.

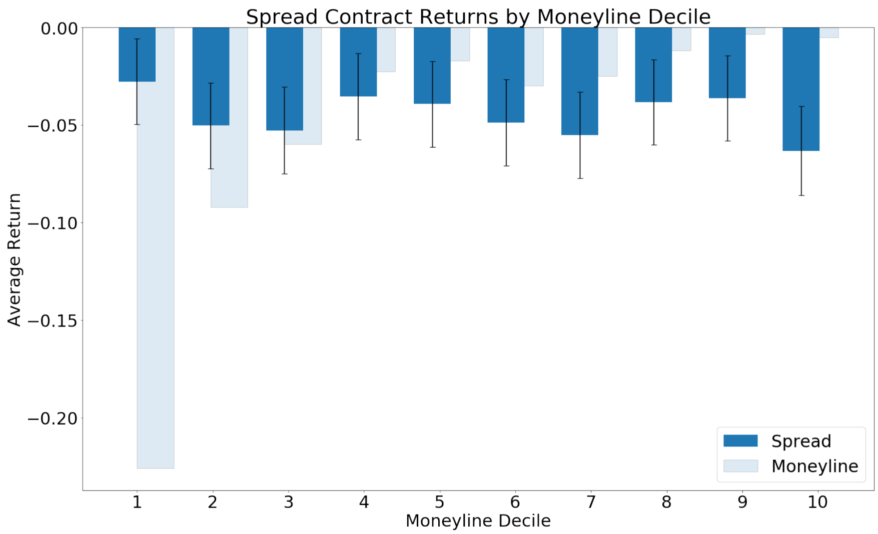


Figure 4: Implied Volatility Surface by Sport

The figure plots the Implied Volatility Surface for betting contracts, with one panel for each sport. For each game, we calculate the the win probabilities non-parametrically. The "implied win probability" for a contract is calculated as the probability that makes the contract's expected return equal to a Spread contract with a 50% change of paying off. The implied volatility for a contract is the value σhat that satisfies Equation (4), using the win probability and implied probability of the contract. All contracts are placed into one of 90 equally spaced bins based on the win probability. Each plot plots a scatterplot of the average implied volatility against the average win probability in each bin. Each plot also has a dotted red horizontal line, which corresponds with the sample Maximum Likelihood Estimate of the standard deviation of the point-differential minus Spread line within each sample.

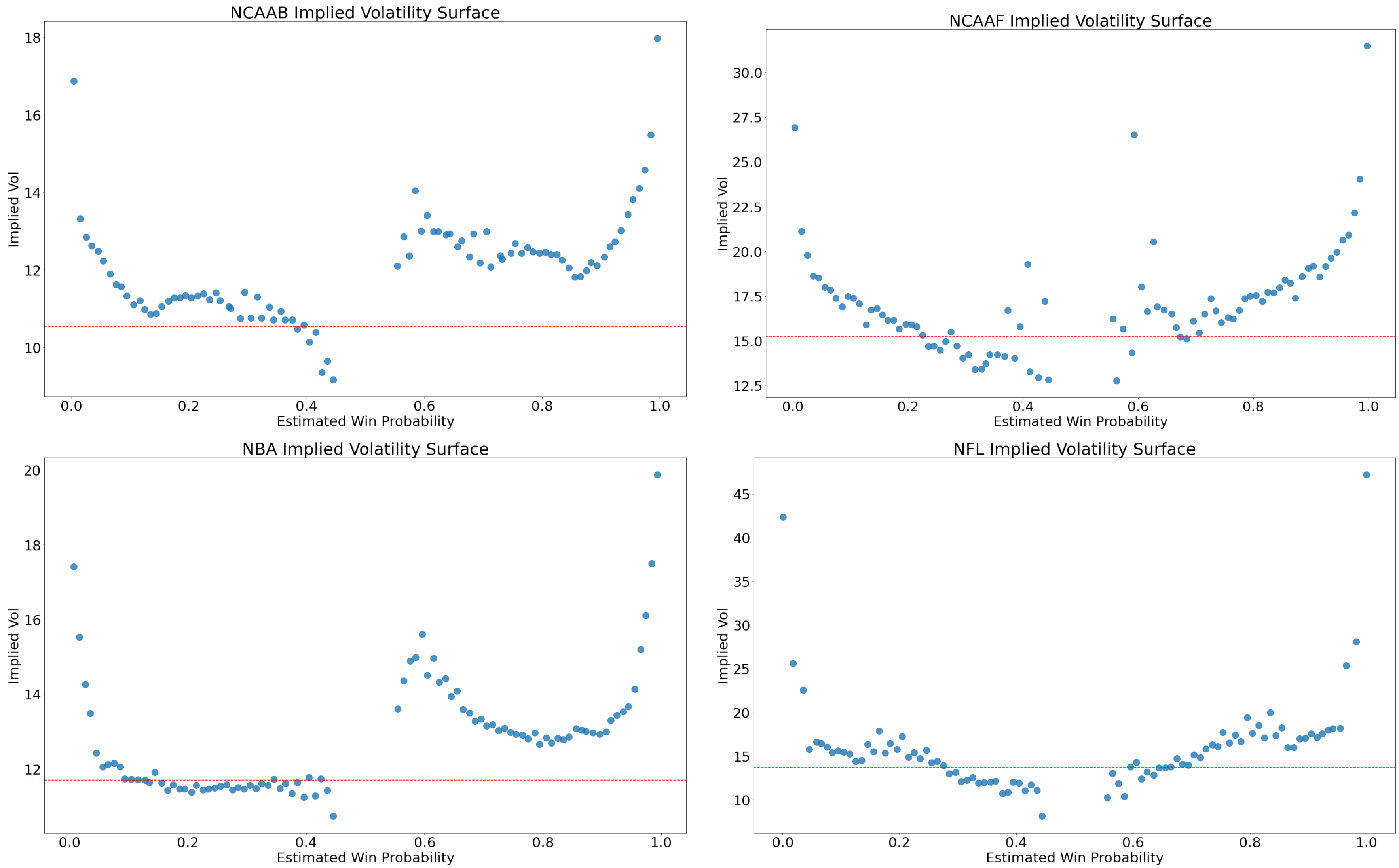


Figure 5: Surface of Mean Squared Errors and Valid Contract Proportions

The top two panels in the figure plot the proportion of contracts against values of α, the diminishing sensitivity parameter, and γ, the probability weighting parameter, for which the model computes a valid payoff that satisfies the optimality conditions. The top left panel corresponds with values where the loss aversion parameter, $\lambda = 1$. The top right panel corresponds with values for $\lambda = 1.25$. The bottom two panels in the figure plot the mean squared error of expected returns against α and γ. The bottom left panel corresponds with values for $\lambda = 1$ and the bottom right panel corresponds with values for $\lambda = 1.25$.

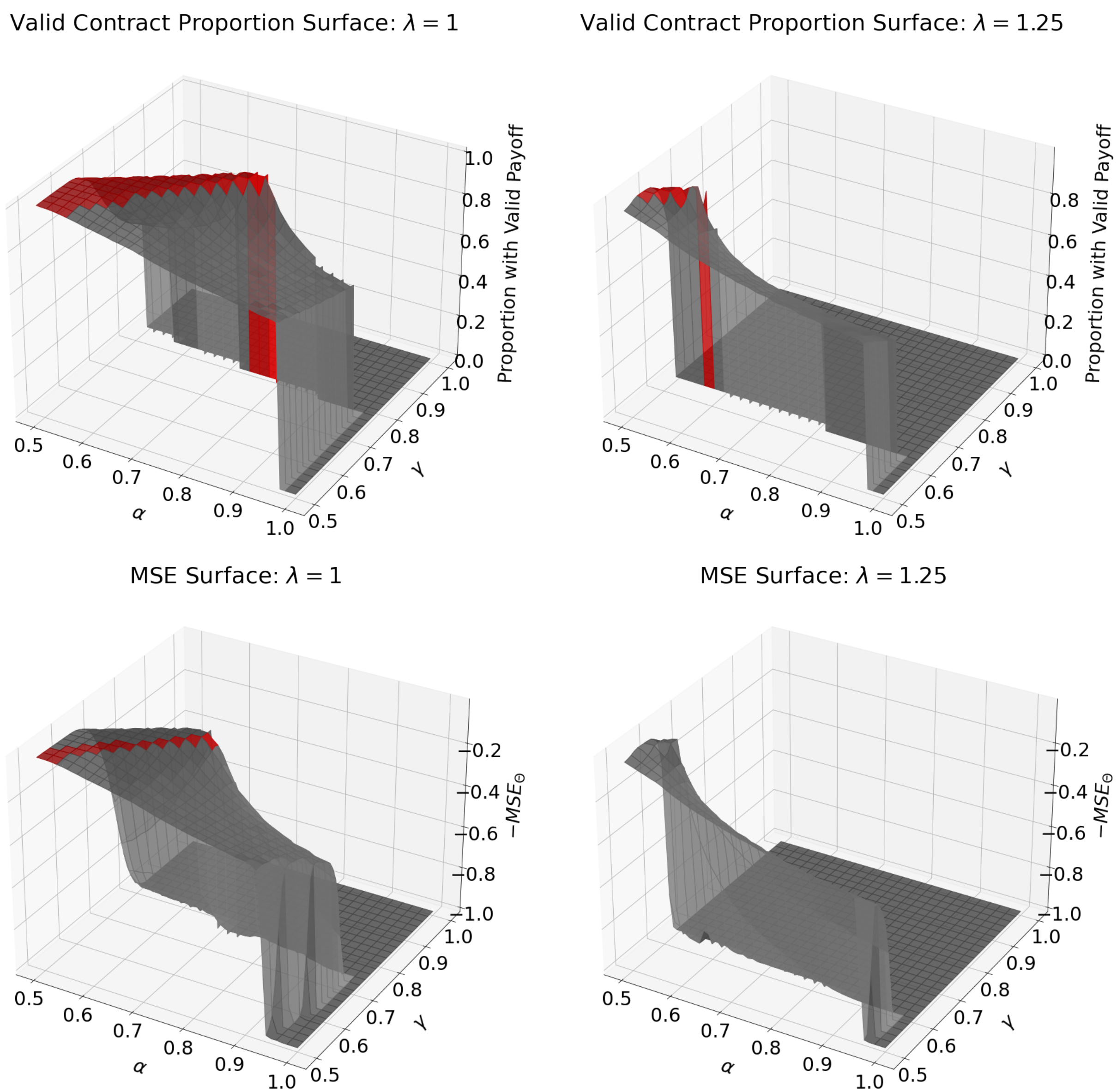


Figure 6: Preference-Implied Returns and Expected Returns

The figure plots the average expected returns and preference-implied expected returns of Moneyline contracts sorted into deciles based on the Moneyline. Decile 1 corresponds with contracts betting on the most extreme underdogs and Decile 10 corresponds with contracts on the most extreme favorites. Model-implied expected returns are estimated by assuming a Cumulative Prospect Theory bettor that is indifferent between betting on each contract offered and a spread contract. The parameter values used are $(\alpha, \gamma, \lambda) = (0.65, 0.65, 1)$.

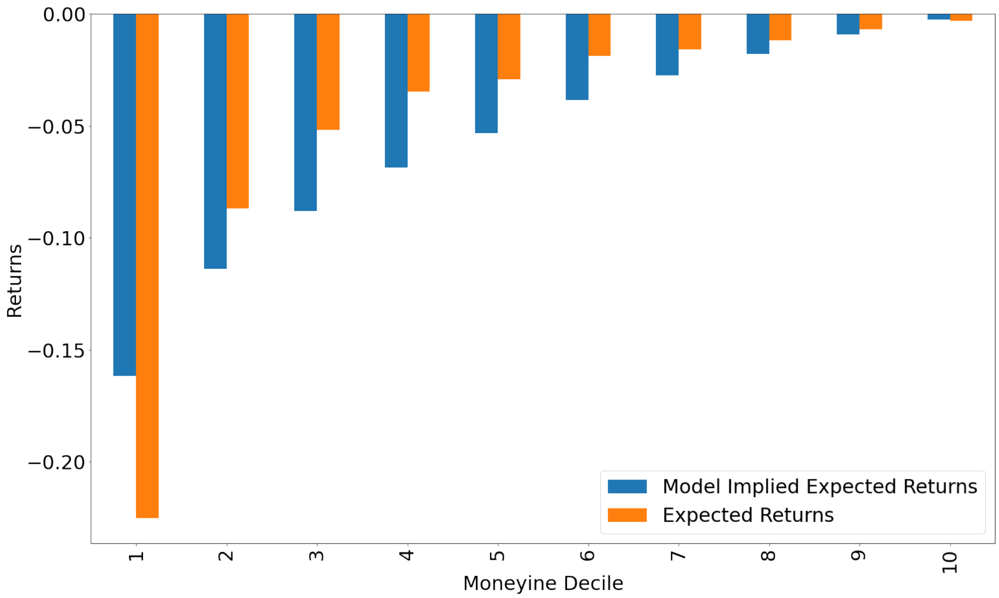


Figure 7: The Probability Weighting Function

The figure plots objective probabilities p, on the x-axis, versus weighted probabilities, $w(p)$ on the y-axis. The lines plotted correspond with $w(p) = p$ and the Tversky and Kahneman (1992) probability weighting function, $w(p) = \frac{p^\gamma}{(p^\gamma + (1-p)^\gamma)^{1/\gamma}}$, for $\gamma = 0.65$ and $\gamma = 0.5$.

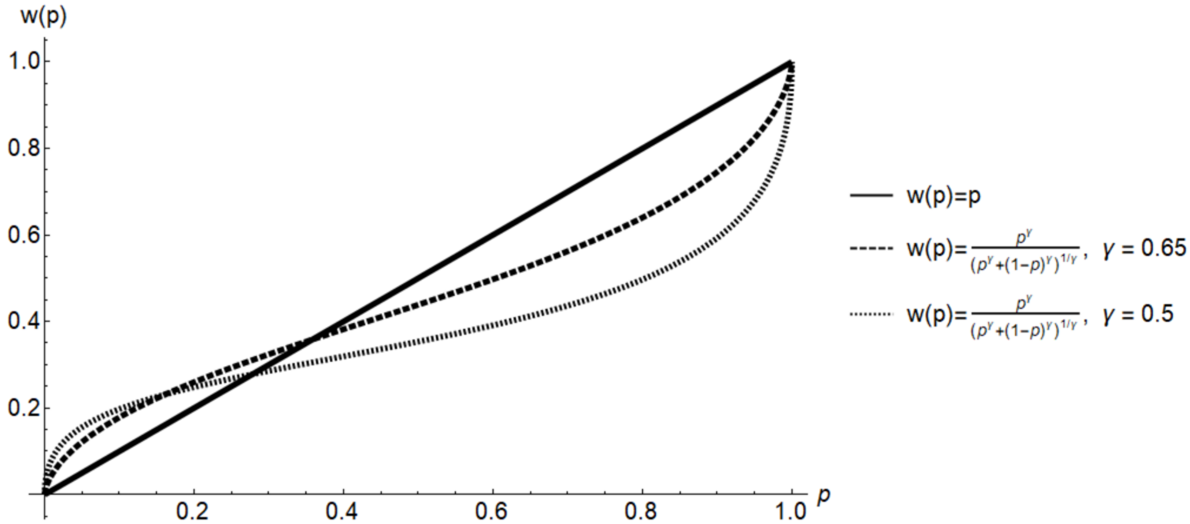
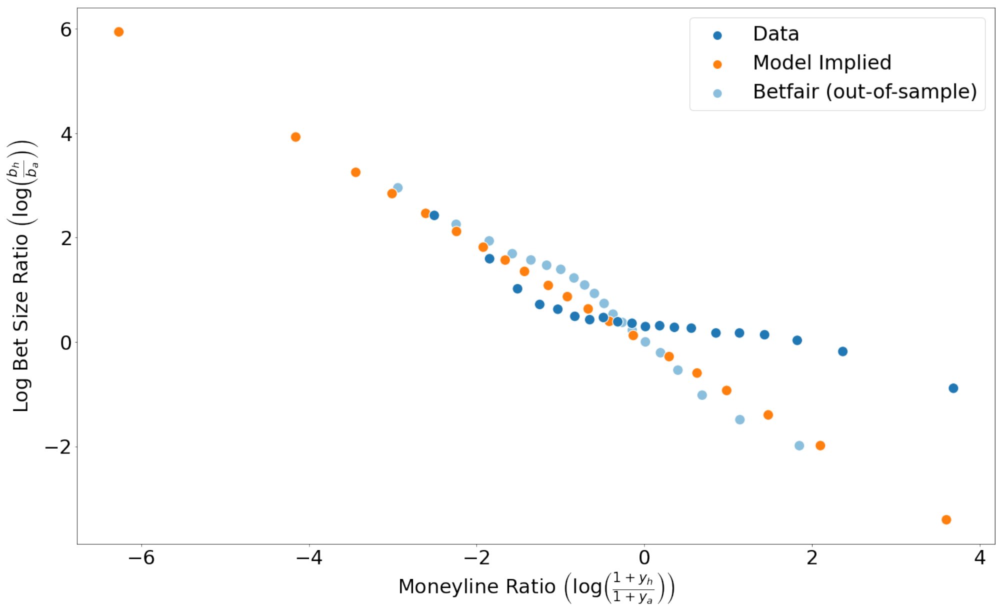


Figure 8: Average Bet Sizes

The first panel in the figure is a binned scatter plot of the log ratio of $\log\left(\frac{1+y_h}{1+y_a}\right)$ (the "Moneyline Ratio") versus the estimated average bet size for bets placed on the home team relative to the average bet size for bets placed on the away team (the "bet size ratio"), where y_h and y_a are the payoffs associated with winning Moneyline bets on the home and away teams at close. Each game is sorted into one of twenty equally sized bins based on the bet size ratio of the game. Each point corresponds with the average log bet size ratio and average Moneyline ratio within a bin. The dots in blue correspond with values calculated using the proportion of bets (estimated using Equation (17)) and empirically observed payoffs offered on games, the dots in orange correspond with values implied by the Cumulative Prospect Theory model estimated in Section 5, and the dots in light blue correspond with values calculated using the observed average bet sizes and payoffs for a sample of soccer games from Betfair (described in Appendix C.3). The second panel is a binned scatterplot of the bet size ratio from the data versus the bet size ratio implied by the model, where the bins are once again constructed by sorting games into one of twenty equally sized bins.



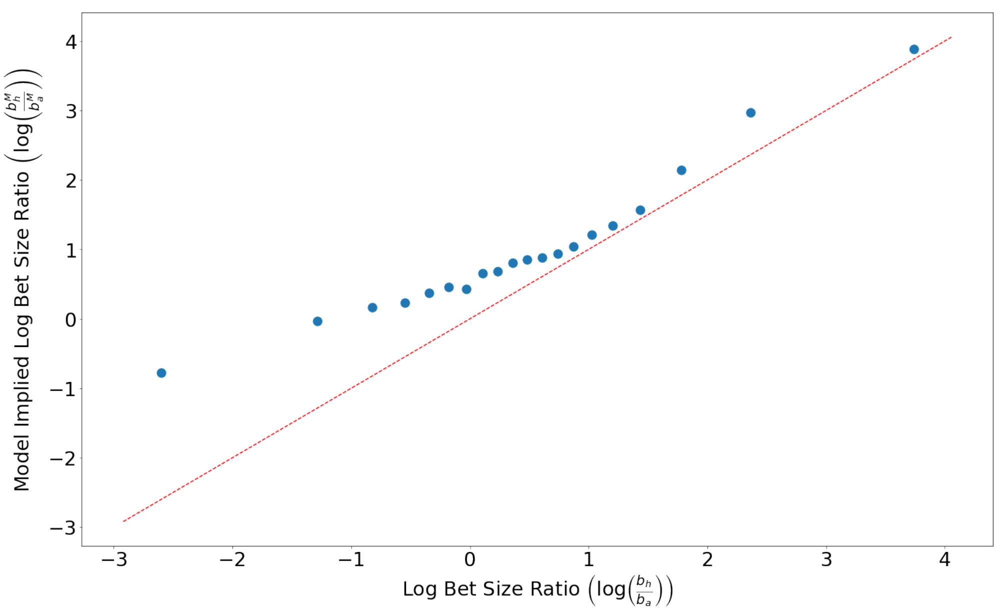


Figure 9: Accuracy of Over/Under Contracts

The first panel in the figure is a binned scatter plot of the Over/Under line versus the total number of point scored in a game. Each game in our sample is sorted into 20 equal sized bins based on the Over/Under line of the game. Each point on the plot corresponds with the average Over/Under Line and the average point total of each game in one of the bins. The 45 degree line is also plotted on the graph in red. The second panel in the figure sorts each game into deciles based on the quantity $\log\left(\frac{1+y_h}{1+y_a}\right)$ ("the Moneyline Ratio"), where y_h and y_a are the payoffs associated with winning Moneyline bets on the home and away teams, and plots the average total number of points scored and Over/Under line in each decile. The error bars correspond with plus/minus two standard deviations relative to the mean.

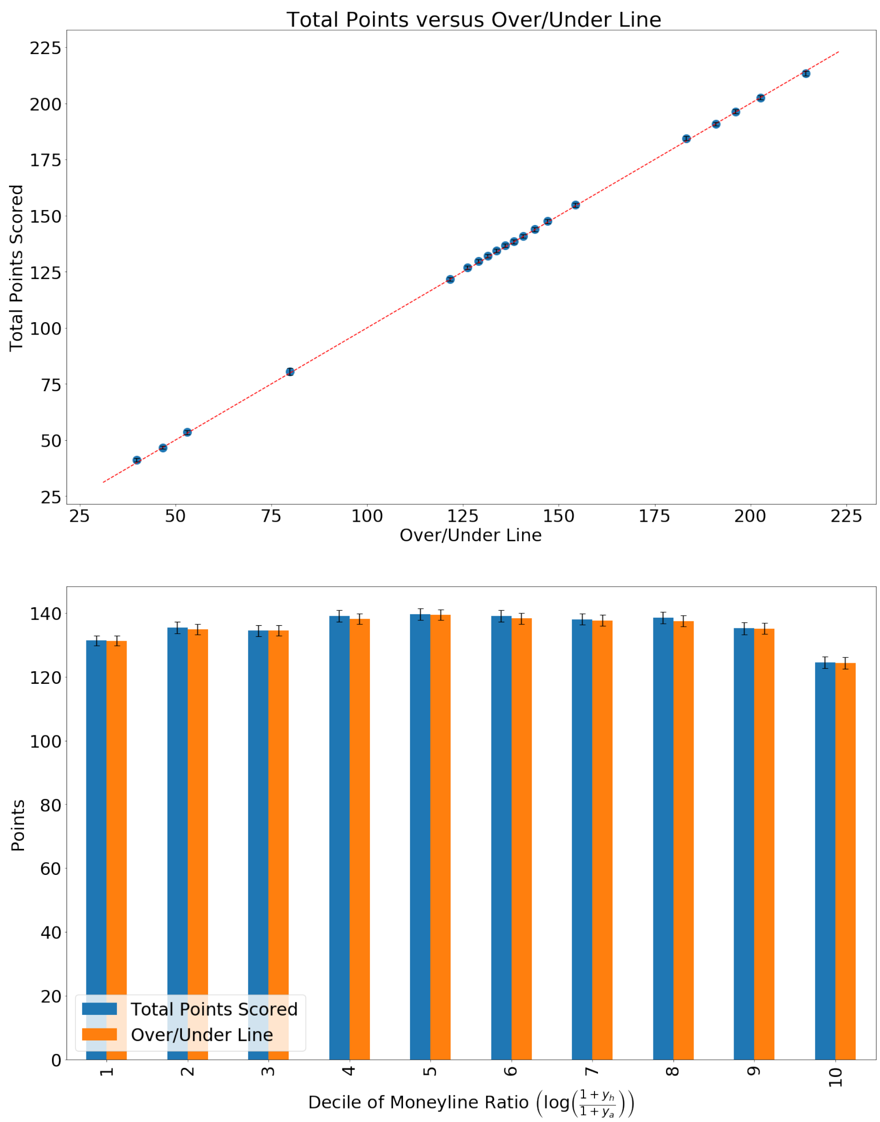


Figure 10: Comparison of Market Share Measures

The figure plots a binned scatterplot of the proportion of bets placed on a Moneyline contract in a game versus the estimated dollar proportion of bets placed on the contract. The estimated dollar proportion of bets is constructed by comparing the payoffs offered on the two Moneyline contracts on a game, using the market clearing conditions in Equations (B.1) and (B.2) which assume that bookmakers take no risk on a game. All observations are grouped into twenty bins based on the estimated dollar proportion of bets. The scatterplot plots the average proportion of bets versus the average estimated dollar proportion of bets for each bin. The figure also plots the 45 degree line in red.

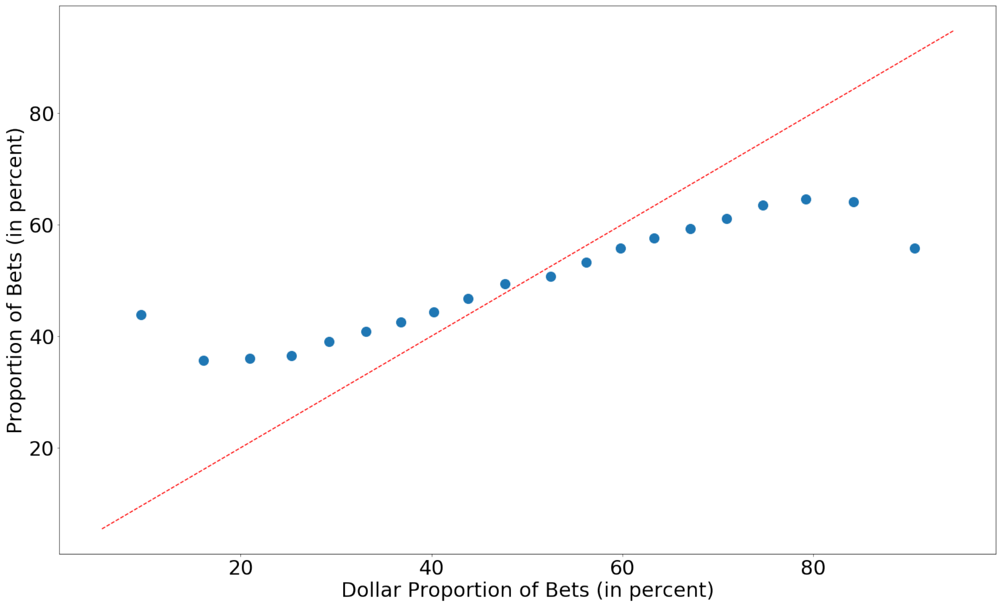


Table 1: Implied Volatility Premium: Sports Betting Contracts and Options

The table presents data on the implied volatilities in sports betting contracts and options contracts. The first panel of the table presents the sample standard deviation of the point-spread minus the spread line, and the average implied volatility of Moneyline contracts expected to pay off 5%-15% and 85%-95% of the time within each sport. The payoff probabilities are estimated nonparametrically, as described in Appendix A. The last two rows of the panel correspond with the Implied Volatility premia, as the percent different between the average implied volatilities and the sample volatility. Panel B presents the average implied volatility for standardized, one month maturity 10, 50, and 90 delta call options and -10, -50, -90 delta put options across a set of 13 indices. The data are from the OptionMetrics Implied Volatility surface file, and are calculated by interpolating options of different maturities and deltas. Panel C presents the average implied volatility for standardized, one month maturity 10, 50, and 90 delta call and -10, -50, -90 delta put options for all equities for which data is available in the OptionMetrics Implied Volatility surface file. The last two rows of Panels B and C present the "IV Premium" for call and put options, which we define as the percent difference between an option and the corresponding 50 delta option.

Panel A: Sports Betting Contracts

	NCAAB	NCAAF	NBA	NFL	Average
Sample Volatility	10.5	15.3	11.7	13.7	-
IV for 5%-15% WP Contract	11.3	17.2	11.9	15.4	-
IV for 85%-95% WP Contract	12.4	18.9	13.1	17.1	-
IV Premium for 5%-15% WP	7.6%	12.8%	1.4%	12.4%	8.6%
IV Premium for 85%-95% WP	17.4%	23.7%	12.4%	24.8%	19.6%

Panel B: Index Options

	Call Options	Put Options	Average
IV for 50 Delta	0.22	0.21	0.21
IV for 10 Delta	0.19	0.28	0.24
IV for 90 Delta	0.30	0.19	0.24
IV Premium for 10 Delta	-4.5%	43.6%	19.6%
IV Premium for 90 Delta	48.0%	-0.9%	23.6%

Panel C: Equity Options

	Call Options	Put Options	Average
IV for 50 Delta	0.46	0.47	0.46
IV for 10 Delta	0.54	0.59	0.56
IV for 90 Delta	0.55	0.52	0.54
IV Premium for 10 Delta	17.0%	26.7%	21.9%
IV Premium for 90 Delta	21.1%	12.4%	16.7%

Table 2: Spread lines, Moneylines, and Predicting Wins

Panel A of the table presents regression results from OLS regressions where the independent variable is a 0/1 indicator variable capturing whether the home team won the game and the independent variables are the closing Spread line of the game and $\log\left(\frac{1+y_h}{1+y_a}\right)$ (the "Moneyline Ratio"), where y_h and y_a are the payoffs associated with winning Moneyline bets on the home and away teams at close. Panel B reports results from regressions of changes in the Moneyline ratio from open to close of betting on changes in the Spread line from open to close of betting for each sport. Independent variables are standardized to have zero mean and unit standard deviation within each sport, and separate regression coefficients are estimated by sport, with t-statistics are reported in parentheses.

Panel A: Wins, Spreads, and Moneylines

	Dependent variable = $\mathbb{1}_{\text{home team wins}}$		
	(1)	(2)	(3)
$\beta_{\text{ML,NCAAB}}$	-0.22		0.01
	(-76.49)		(0.36)
$\beta_{\text{ML,NCAAF}}$	-0.26		-0.14
	(-40.01)		(-2.40)
$\beta_{\text{ML,NBA}}$	-0.19		-0.04
	(-41.12)		(-1.33)
$\beta_{\text{ML,NFL}}$	-0.18		-0.11
	(-17.68)		(-1.27)
$\beta_{\text{Spread,NCAAB}}$		-0.22	-0.24
		(-76.98)	(-7.78)
$\beta_{\text{Spread,NCAAF}}$		-0.26	-0.12
		(-40.04)	(-2.06)
$\beta_{\text{Spread,NBA}}$		-0.20	-0.16
		(-41.57)	(-5.91)
$\beta_{\text{Spread,NFL}}$		-0.18	-0.07
		(-17.68)	(-0.80)
Sport FE	Yes	Yes	Yes
N	36,609	36,609	36,609
R^2	20.66%	20.85%	20.87%
F-statistic	1361.37	1377.75	877.72

Panel B: Changes in Spread Lines and Moneylines

	Dependent variable = ΔMoneyline Ratio$_{\text{open-to-close}}$			
	NCAAB	NCAAF	NBA	NFL
ΔSpread Line$_{\text{open-to-close}}$	0.61	0.53	0.63	0.62
	(114.49)	(41.31)	(73.49)	(34.08)
N	21,742	4,363	8,390	1,843
R^2	0.38	0.28	0.39	0.39

Table 3: Model Estimates

The table presents details of the estimation of the main preference specification, calculated by fixing a value of γ and minimizing Equation (14). Panel A presents the estimate of α, the number of contracts (out of 73218) for which the model is able to calculate payoffs that satisfy the equilibrium and optimality conditions, and the mean squared error of expected returns from the model, with standard errors from 2000 bootstrap samples reported in parentheses. Panel B groups observations into deciles based on the Moneyline, and presents the average log payoff ($\log y_i$) and average expected returns of contracts in the decile, as well as the corresponding values output from the model (super-scripted by M) for the assumed value of γ. The panel also includes the proportion of contracts for which the model is able to calculate implied payoffs in each decile.

Panel A: Model Estimate Details

γ=0.65			γ=0.5		
α	Valid Bets	MSE	α	Valid Bets	MSE
0.650	73,188	0.012	0.502	73,187	0.012
(0.0004)	(49.4)	(0.005)	(0.002)	(53.9)	(0.005)

Panel B: Model Estimates By Decile

γ=0.65							γ=0.5						
Decile	ML	$\log y_i$	$\log y_i^M$	$E(r_i)$	$E(r_i)^M$	Valid	Decile	ML	$\log y_i$	$\log y_i^M$	$E(r_i)$	$E(r_i)^M$	Valid
1	1003.68	2.15	2.36	-22.5%	-16.2%	100.0%	1	1003.68	2.15	2.33	-22.5%	-18.4%	100.0%
2	355.96	1.26	1.22	-8.7%	-11.4%	100.0%	2	355.96	1.26	1.21	-8.7%	-12.1%	100.0%
3	224.25	0.80	0.75	-5.2%	-8.8%	100.0%	3	224.25	0.80	0.74	-5.2%	-9.1%	100.0%
4	157.06	0.45	0.39	-3.5%	-6.9%	100.0%	4	157.06	0.45	0.39	-3.5%	-7.0%	100.0%
5	90.85	0.12	0.08	-2.9%	-5.3%	100.0%	5	90.85	0.12	0.08	-2.9%	-5.4%	100.0%
6	-125.91	-0.23	-0.27	-1.9%	-3.8%	100.0%	6	-125.91	-0.23	-0.27	-1.9%	-3.8%	100.0%
7	-173.93	-0.55	-0.58	-1.6%	-2.8%	100.0%	7	-173.93	-0.55	-0.58	-1.6%	-2.7%	100.0%
8	-255.09	-0.93	-0.95	-1.2%	-1.8%	100.0%	8	-255.09	-0.93	-0.95	-1.2%	-1.8%	100.0%
9	-430.70	-1.44	-1.46	-0.7%	-0.9%	100.0%	9	-430.70	-1.44	-1.46	-0.7%	-0.9%	100.0%
10	-1570.88	-2.52	-2.61	-0.3%	-0.3%	99.6%	10	-1570.88	-2.52	-2.61	-0.3%	-0.3%	99.6%

Table 4: Alternate Preference Specification: Contracts with Positive Valuations

The table presents the proportion of contracts that have positive valuations under the alternative preference specification given in Equation (18), which differs from the main preference specification in the form that probability weighting takes. Panel A fixes the value of α as 0.65, and presents the proportion of contracts with positive valuations for a set of (α, γ, λ) triples. Panel B fixes the value of γ as 0.65 and presents the proportion of contracts with positive valuations for a set of (α, γ, λ) triples.

Panel A: Proportion with Positive Value, $\alpha = 0.65$

γ	$\lambda = 1$		$\lambda = 1.25$	
	Favorites	Underdogs	Favorites	Underdogs
0.50	0%	83%	0%	6%
0.55	0%	66%	0%	1%
0.60	0%	30%	0%	0%
0.65	0%	0%	0%	0%
0.70	66%	0%	1%	0%
0.75	79%	0%	8%	0%
0.80	85%	0%	19%	0%
0.85	89%	0%	33%	0%
0.90	90%	0%	43%	0%
0.95	92%	0%	50%	0%

Panel B: Proportion with Positive Value, $\gamma = 0.65$

α	$\lambda = 1$		$\lambda = 1.25$	
	Favorites	Underdogs	Favorites	Underdogs
0.50	88%	0%	21%	0%
0.55	82%	0%	8%	0%
0.60	68%	0%	1%	0%
0.65	0%	0%	0%	0%
0.70	0%	28%	0%	0%
0.75	0%	58%	0%	1%
0.80	0%	76%	0%	4%
0.85	0%	84%	0%	14%
0.90	0%	87%	0%	26%
0.95	0%	89%	0%	39%

References

Ali, Mukhtar M, "Probability and utility estimates for racetrack bettors," *Journal of Political Economy*, 1977, *85* (4), 803–815.

Andrikogiannopoulou, Angie and Filippos Papakonstantinou, "Heterogeneity in risk preferences: Evidence from a real-world betting market," *Swiss Finance Institute Research Paper*, 2016, (13-53).

_ and _ , "History-dependent risk preferences: Evidence from individual choices and implications for the disposition effect," *Review of Financial Studies, forthcoming*, 2019.

Ang, Andrew, Robert J Hodrick, Yuhang Xing, and Xiaoyan Zhang, "The cross-section of volatility and expected returns," *Journal of Finance*, 2006, *61* (1), 259–299.

_ , _ , _ , and _ , "High idiosyncratic volatility and low returns: International and further US evidence," *Journal of Financial Economics*, 2009, *91* (1), 1–23.

Asness, Cliff, Andrea Frazzini, Niels Joachim Gormsen, and Lasse Heje Pedersen, "Betting against correlation: Testing theories of the low-risk effect," *Journal of Financial Economics*, 2020, *135* (3), 629–652.

Avery, Christopher and Judith Chevalier, "Identifying investor sentiment from price paths: The case of football betting," *Journal of Business*, 1999, *72* (4), 493–521.

Baele, Lieven, Joost Driessen, Sebastian Ebert, Juan M Londono, and Oliver Spalt, "Cumulative prospect theory, option prices, and the variance premium," *Review of Financial Studies*, 2016.

Barberis, Nicholas, "The psychology of tail events: Progress and challenges," *American Economic Review*, 2013, *103* (3), 611–16.

_ , Abhiroop Mukherjee, and Baolian Wang, "Prospect theory and stock returns: An empirical test," *Review of Financial Studies*, 2016, *29* (11), 3068–3107.

_ and Ming Huang, "Mental accounting, loss aversion, and individual stock returns," *Journal of Finance*, 2001, *56* (4), 1247–1292.

_ and _ , "Stocks as lotteries: The implications of probability weighting for security prices," *American Economic Review*, 2008, *98* (5), 2066–2100.

_ and Wei Xiong, "What drives the disposition effect? An analysis of a long-standing preference-based explanation," *Journal of Finance*, 2009, *64* (2), 751–784.

_ , Lawrence J Jin, and Baolian Wang, "Prospect Theory and Stock Market Anomalies," *Journal of Finance*, Forthcoming.

_ , Ming Huang, and Richard H Thaler, "Individual preferences, monetary gambles, and stock market participation: A case for narrow framing," *American Economic Review*, 2006, *96* (4), 1069–1090.

_ , _ , and Tano Santos, "Prospect theory and asset prices," *Quarterly Journal of Economics*, 2001, *116* (1), 1–53.

Barseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua C Teitelbaum, "The nature of risk preferences: Evidence from insurance choices," *American Economic Review*, 2013, *103* (6), 2499–2529.

_ , _ , _ , and _ , "Estimating risk preferences in the field," *Journal of Economic Literature*, 2018, *56* (2), 501–64.

_ , Jeffrey Prince, and Joshua C Teitelbaum, "Are risk preferences stable across contexts? Evidence from insurance data," *American Economic Review*, 2011, *101* (2), 591–631.

Benartzi, Shlomo and Richard H Thaler, "Myopic loss aversion and the equity premium puzzle," *Quarterly Journal of Economics*, 1995, *110* (1), 73–92.

Bernheim, B Douglas and Charles Sprenger, "On the Empirical Validity of Cumulative Prospect Theory: Experimental Evidence of Rank-Independent Probability Weighting," *Econometrica*, 2020, *88* (4), 1363–1409.

Black, Fischer, "Capital market equilibrium with restricted borrowing," *Journal of Business*, 1972, *45* (3), 444–455.

_ , Michael C Jensen, Myron Scholes et al., "The capital asset pricing model: Some empirical tests," *Studies in the Theory of Capital Markets*, 1972, *81* (3), 79–121.

Bondarenko, Oleg, "Why are put options so expensive?," *Quarterly Journal of Finance*, 2014, *4* (03), 1450015.

Booij, Adam S, Bernard MS Van Praag, and Gijs Van De Kuilen, "A parametric analysis of prospect theory's functionals for the general population," *Theory and Decision*, 2010, *68* (1-2), 115–148.

Boyer, Brian H and Keith Vorkink, "Stock options as lotteries," *Journal of Finance*, 2014, *69* (4), 1485–1527.

Bradley, Ian, "The representative bettor, bet size, and prospect theory," *Economics Letters*, 2003, *78* (3), 409–413.

Brown, William O and Raymond D Sauer, "Does the basketball market believe in the hot hand? Comment," *American Economic Review*, 1993, *83* (5), 1377–1386.

Brunnermeier, Markus K, Christian Gollier, and Jonathan A Parker, "Optimal beliefs, asset prices, and the preference for skewed returns," *American Economic Review*, 2007, *97* (2), 159–165.

Camerer, Colin F, "Does the basketball market believe in the hot hand?," *American Economic Review*, 1989, *79* (5), 1257–1261.

Chapman, Jonathan, Erik Snowberg, Stephanie Wang, and Colin Camerer, "Loss attitudes in the US population: Evidence from dynamically optimized sequential experimentation (DOSE)," Technical Report, National Bureau of Economic Research 2018.

Chiappori, Pierre-André, Bernard Salanié, François Salanié, and Amit Gandhi, "From aggregate betting data to individual risk preferences," *Econometrica*, 2019, *87* (1), 1–36.

Conlisk, John, "The utility of gambling," *Journal of Risk and Uncertainty*, 1993, *6* (3), 255–275.

Cox, John C, Stephen A Ross, and Mark Rubinstein, "Option pricing: A simplified approach," *Journal of Financial Economics*, 1979, *7* (3), 229–263.

DellaVigna, Stefano, "Structural behavioral economics," in "Handbook of Behavioral Economics: Applications and Foundations 1," Vol. 1, Elsevier, 2018, pp. 613–723.

Diether, Karl B, Christopher J Malloy, and Anna Scherbina, "Differences of opinion and the cross section of stock returns," *Journal of Finance*, 2002, *57* (5), 2113–2141.

Dorn, Daniel and Paul Sengmueller, "Trading as entertainment?," *Management Science*, 2009, *55* (4), 591–603.

Duraj, Jetlir and Kevin He, "Dynamic information design with diminishing sensitivity over news," 2020.

Durham, Gregory R, Michael G Hertzel, and J Spencer Martin, "The market impact of trends and sequences in performance: New evidence," *Journal of Finance*, 2005, *60* (5), 2551–2569.

Egan, Mark L, Alexander MacKay, and Hanbin Yang, "Recovering Investor Expectations from Demand for Index Funds," Technical Report, National Bureau of Economic Research 2020.

Einav, Liran, Amy Finkelstein, Iuliana Pascu, and Mark R Cullen, "How general are risk preferences? Choices under uncertainty in different domains," *American Economic Review*, 2012, *102* (6), 2606–38.

Fehr-Duda, Helga and Thomas Epper, "Probability and risk: Foundations and Economic implications of probability-dependent risk preferences," *Annu. Rev. Econ.*, 2012, *4* (1), 567–593.

Frazzini, Andrea and Lasse H Pedersen, "Embedded leverage," Technical Report 2020.

_ **and Lasse Heje Pedersen**, "Betting against beta," *Journal of Financial Economics*, 2014, *111* (1), 1–25.

Gandar, John M, William H Dare, Craig R Brown, and Richard A Zuber, "Informed traders and price variations in the betting market for professional basketball games," *Journal of Finance*, 1998, *53* (1), 385–401.

Gandar, John, Richard Zuber, Thomas O'brien, and Ben Russo, "Testing rationality in the point spread betting market," *Journal of Finance*, 1988, *43* (4), 995–1008.

Gandhi, Amit and Ricardo Serrano-Padial, "Does belief heterogeneity explain asset prices: The case of the longshot bias," *Review of Economic Studies*, 2015, *82* (1), 156–186.

Garleanu, Nicolae, Lasse Heje Pedersen, and Allen M Poteshman, "Demand-based option pricing," *The Review of Financial Studies*, 2008, *22* (10), 4259–4299.

Geanakoplos, John, "The leverage cycle," *NBER Macroeconomics Annual*, 2010, *24* (1), 1–66.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus, "Five facts about beliefs and portfolios," *American Economic Review*, 2021, *111* (5), 1481–1522.

Golec, Joseph and Maurry Tamarkin, "The degree of inefficiency in the football betting market: Statistical tests," *Journal of Financial Economics*, 1991, *30* (2), 311–323.

Gormsen, Niels Joachim and Eben Lazarus, "Duration-driven returns," 2020.

Gray, Philip K and Stephen F Gray, "Testing market efficiency: Evidence from the NFL sports betting market," *Journal of Finance*, 1997, *52* (4), 1725–1737.

Green, Etan, Haksoo Lee, and David M Rothschild, "The favorite-longshot midas," 2020.

Griffith, Richard M, "Odds adjustments by American horse-race bettors," *American Journal of Psychology*, 1949, *62* (2), 290–294.

Grinblatt, Mark and Bing Han, "Prospect theory, mental accounting, and momentum," *Journal of Financial Economics*, 2005, *78* (2), 311–339.

_ **and Matti Keloharju**, "Sensation seeking, overconfidence, and trading activity," *Journal of Finance*, 2009, *64* (2), 549–578.

Harrison, J Michael and David M Kreps, "Speculative investor behavior in a stock market with heterogeneous expectations," *Quarterly Journal of Economics*, 1978, *92* (2), 323–336.

Hausch, Donald B, William T Ziemba, and Mark Rubinstein, "Efficiency of the market for racetrack betting," *Management Science*, 1981, *27* (12), 1435–1452.

Hudson, James, "UK Sports Betting Market Growth and Mobile Betting," *Favourit*, 2014.

Imas, Alex, "The realization effect: Risk-taking after realized versus paper losses," *American Economic Review*, 2016, *106* (8), 2086–2109.

Ingersoll, Jonathan, "Non-Monotonicity of the Tversky-Kahneman Probability-Weighting Function: A Cautionary Note," *European Financial Management*, 2008, *14* (3), 385–390.

Jullien, Bruno and Bernard Salanié, "Estimating preferences under risk: The case of racetrack bettors," *Journal of Political Economy*, 2000, *108* (3), 503–530.

Kahneman, Daniel, "Maps of bounded rationality: Psychology for behavioral Economics," *American Economic Review*, 2003, *93* (5), 1449–1475.

Kliger, Doron and Ori Levy, "Theories of choice under risk: Insights from Financial markets," *Journal of Economic Behavior & Organization*, 2009, *71* (2), 330–346.

Kőszegi, Botond and Matthew Rabin, "A model of reference-dependent preferences," *Quarterly Journal of Economics*, 2006, *121* (4), 1133–1165.

_ **and** _ , "Reference-dependent risk attitudes," *American Economic Review*, 2007, *97* (4), 1047–1073.

Kuypers, Tim, "Information and efficiency: an empirical study of a fixed odds betting market," *Applied Economics*, 2000, *32* (11), 1353–1363.

Lettau, Martin and Jessica A Wachter, "Why is long-horizon equity less risky? A duration-based explanation of the value premium," *The Journal of Finance*, 2007, *62* (1), 55–92.

Levitt, Steven D, "Why are gambling markets organised so differently from Financial markets?," *Economic Journal*, 2004, *114* (495), 223–246.

Miller, Edward M, "Risk, uncertainty, and divergence of opinion," *Journal of Finance*, 1977, *32* (4), 1151–1168.

Moskowitz, Tobias J, "Asset pricing and sports betting," *Journal of Finance*, Forthcoming.

Ni, Sophie Xiaoyan, "Stock option returns: A puzzle," 2008.

O'Donoghue, Ted and Charles Sprenger, "Reference-dependent preferences," in "Handbook of Behavioral Economics: Applications and Foundations 1," Vol. 1, Elsevier, 2018, pp. 1–77.

Ottaviani, Marco and Peter Norman Sørensen, "Surprised by the parimutuel odds?," *American Economic Review*, 2009, *99* (5), 2129–34.

_ **and** _ , "Noise, information, and the favorite-longshot bias in parimutuel predictions," *American Economic Journal: Microeconomics*, 2010, *2* (1), 58–85.

Pankoff, Lyn D, "Market efficiency and football betting," *Journal of Business*, 1968, *41* (2), 203–214.

Peta, Joe, *Trading Bases: How a Wall Street Trader Made a Fortune Betting on Baseball*, Penguin, 2014.

Pope, Peter F and David A Peel, "Information, prices and efficiency in a fixed-odds betting market," *Economica*, 1989, pp. 323–341.

Prelec, Drazen, "The probability weighting function," *Econometrica*, 1998, pp. 497–527.

Quiggin, John, "A theory of anticipated utility," *Journal of Economic Behavior & Organization*, 1982, *3* (4), 323–343.

Sauer, Raymond D, Vic Brajer, Stephen P Ferris, and M Wayne Marr, "Hold your bets: Another look at the efficiency of the gambling market for National Football League games," *Journal of Political Economy*, 1988, *96* (1), 206–213.

Scheinkman, Jose A and Wei Xiong, "Overconfidence and speculative bubbles," *Journal of Political Economy*, 2003, *111* (6), 1183–1220.

Shefrin, Hersh, *A behavioral approach to asset pricing*, Elsevier, 2008.

_ **and Meir Statman**, "The disposition to sell winners too early and ride losers too long: Theory and evidence," *The Journal of Finance*, 1985, *40* (3), 777–790.

Shin, Hyun Song, "Optimal betting odds against insider traders," *Economic Journal*, 1991, *101* (408), 1179–1185.

_ , "Prices of state contingent claims with insider traders, and the favourite-longshot bias," *Economic Journal*, 1992, *102* (411), 426–435.

Simsek, Alp, "Belief disagreements and collateral constraints," *Econometrica*, 2013, *81* (1), 1–53.

Snowberg, Erik and Justin Wolfers, "Explaining the favorite–long shot bias: Is it risk-love or misperceptions?," *Journal of Political Economy*, 2010, *118* (4), 723–746.

Thaler, Richard H and William T Ziemba, "Anomalies: Parimutuel betting markets: Racetracks and lotteries," *Journal of Economic Perspectives*, 1988, *2* (2), 161–174.

Tversky, Amos and Daniel Kahneman, "Prospect theory: An analysis of decision under risk," *Econometrica*, 1979, *47* (2), 263–291.

_ **and** _ , "Advances in prospect theory: Cumulative representation of uncertainty," *Journal of Risk and Uncertainty*, 1992, *5* (4), 297–323.

Vissing-Jørgensen, Annette, "Limited asset market participation and the elasticity of intertemporal substitution," *Journal of Political Economy*, 2002, *110* (4), 825–853.

Weinberg, Ari, "The case for legal sports gambling," *Forbes. com*, 2003.

Weitzman, M, "Utility analysis and group behavior: an experimental study," *Journal of Political Economy*, 1965, *73*, 18–26.

Woodland, Linda M and Bill M Woodland, "Market efficiency and the favorite-longshot bias: the baseball betting market," *Journal of Finance*, 1994, *49* (1), 269–279.

Zuber, Richard A, John M Gandar, and Benny D Bowers, "Beating the spread: Testing the efficiency of the gambling market for National Football League games," *Journal of Political Economy*, 1985, *93* (4), 800–806.

Appendices for

What Can Betting Markets Tell Us About Investor Preferences and Beliefs? Implications for Low Risk Anomalies

A Non-Parametric Estimation of Win Probabilities

We describe the procedure to estimate win probabilities for Moneyline contracts used in Section 3.2 and Section 5.

Within each sport, we run a non-parametric regression of the home team contract returns on the Spread line of the game, $\log(1 + y_{h,i})$ and $\log(1 + y_{a,i})$, where $y_{j,i}$ is the payoff associated with a winning dollar bet on team j in game i, h is the home team, and a is the away team. Each regression is run using the `KernelReg` function in the `statsmodels` package in Python. Each regression is a local linear regression using a Gaussian kernel. The bandwidth of the kernel is selected via least-squares cross-validation.

We use the fitted values from the regressions as estimates of the expected returns of the home team. We can write the expected return in terms of p_i, the probability of the home team winning game i, and $y_{h,i}$, where p_i is the value we are interested in estimating.

$$E(r_{h,i}) = p_i y_{h,i} + (1 - p_i)(-1) \tag{A.1}$$

Re-writing Equation (A.1) in terms of p_i, we get $p_i = \frac{E(r_{h,i})+1}{y_{h,i}+1}$. The win probability of the away team is $1-p_i$. The non-parametric regression used to produce $E(r_{h,i})$ does not have any restrictions that the implied p must be less than 0 or greater than 1, and there are some instances where the implied win probabilities lie outside this range (588 games out of 36,609). We truncate any values of p_i less than 0 to $\underline{p}$ and any values greater than 1 to $\bar{p}$, where $\underline{p}$ and $\bar{p}$ are the minimum and maximum implied values of p within the range $(0, 1)$. We use these truncated values of p_i to re-estimate the expected returns for the home contract, and to form expected returns for the away contract in each game.[47]

We next validate the accuracy of the estimated expected returns and win probabilities. We form expected returns for the away contracts, and sort contracts into deciles based on the Moneyline (as in Figure 1), and plot the average estimated expected returns and average realized returns of each contract in each decile in

[47] An alternative way to proceed is to use the same methodology except that the away team contract returns are the independent variable in the non-parametric regressions, where p_i is estimated as the away team's win probability. This approach yields similar results.

Figure A.1: Moneyline Contract Expected Returns v. Realized Returns

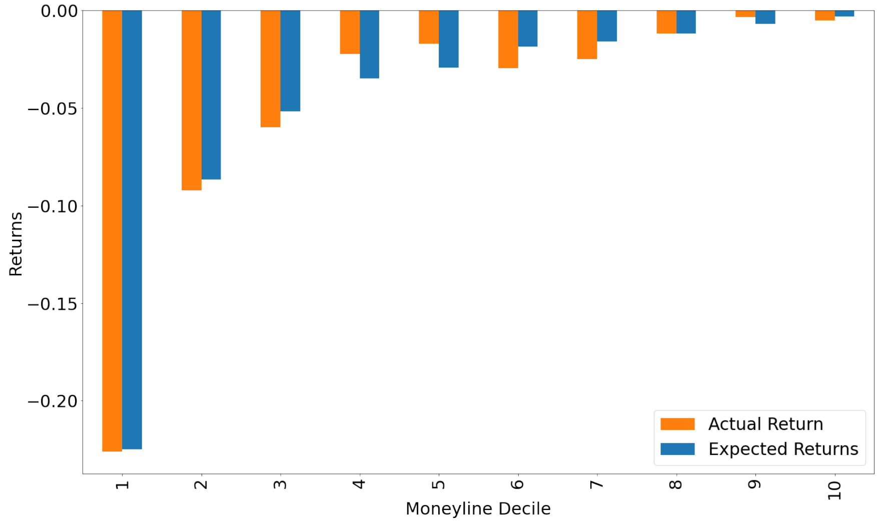


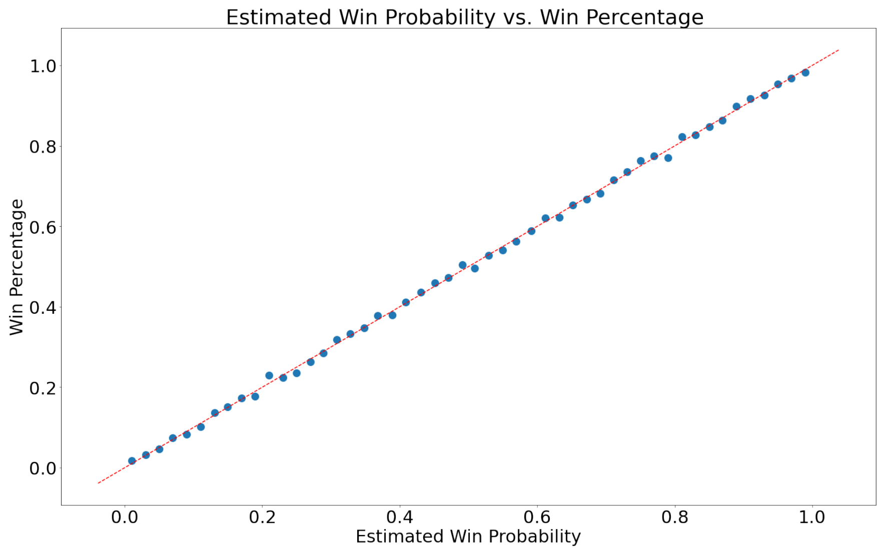
Figure A.1. The expected returns match up with the average realized returns of contracts within each decile and capture the patterns observed in the data.

To assess the estimated win probabilities, we categorize each contract into 50 equally-spaced bins based on the estimated win probability (i.e., one bin for contracts with estimated win probabilities of 0 to 2%, one bin for estimated win probabilities of 2 to 4%, etc.), and plot the average win probabilities against the win percentage of each bin. The plots are presented in Figure A.2. We find that the estimated and realized win probabilities match up along a 45 degree line using both methodologies.

Lastly, one point to note is that we also run alternate versions of this procedure, including non-payoff related characteristics in the regressions. The inclusion of these characteristics generally improves the explanatory power for contract returns, but does not do a substantially better job of explaining the Favorite-Longshot bias, the primary pattern of interest in our study. Additionally, the estimated win probabilities when including non-payoff related characteristics introduce noise into the estimation of the model in Section 5, where it is difficult to accommodate preferences for non-monetary characteristics into the preferences of bettors. Accordingly, we proceed using the estimated win probabilities that do not incorporate non-payoff related characteristics.

Figure A.2: Estimated Win Probabilities vs. Actual Win Percentages

The figure plots a scatterplot of estimated win probabilities against win percentages. The methodology to estimate win probabilities is described in Appendix A. For each sport, all contracts are split into 50 equally spaced bins. The average estimated win probability for each bin is plotted on the x-axis and the actual percentage of games won on the y-axis. A 45 degree line is plotted for reference (dotted red line).



B Matching the Favorite-Longshot Bias with Heterogeneous Beliefs

Here we present the setup for a stylized model of risk-neutral bettors that have heterogeneous beliefs about the expected point-difference in each game. Similar to Gandhi and Serrano-Padial (2015), bettors in the model partition themselves into one of the contracts based on their beliefs about the game's outcome. This type of model can theoretically rationalize the Favorite-Longshot Bias, if the marginal bettor in games where the favorite is more heavily favored than the underdog has more extreme beliefs (in favor of the underdog). We present details on estimating the belief distribution from the model using data on the actual proportion of bets on both teams in the game.

B.1 Setup

We denote the team that is favored to win the game as team 1 and the underdog as team 2. The probability distribution function and cumulative distribution function for the difference in points scored by team 1 and team 2 are f and F. We assume f is normal, with a mean that varies across games and standard deviation σ, which is common across all games in a sport. For each game, there are a continuum of risk-neutral bettors, each indexed by their probability distortion, b. Bettor b perceives the probability distribution as $\hat{f}(x) = f(x - b\sigma)$. A value of $b = 0$ indicates that a bettor has the correct beliefs about the point-spread distribution. More positive and negative values of b correspond with bettors believing the mean point-spread will be higher or lower than the truth (e.g., $b = 1$ indicates the bettor believes the expected point-spread will be one standard deviation higher than the true expected point spread). Bettors are dogmatic in their beliefs, in the sense that viewing market prices does not change their beliefs (they agree to disagree). We denote the probability distribution function of belief distortions b as $h(b)$, and assume that $h(b)$ is atomless and continuous, and that beliefs are drawn from the same distribution for each game. The goal of this section is to estimate $h(b)$, and to understand what the implied probability distribution of beliefs looks like to match prices and volume.

For each game, the bookmaker fixes prices for the Moneyline contracts exogenously. Bettors choose between betting on team 1 and team 2 in the Moneyline (and do not have the option not to bet, meaning we only capture the distribution of beliefs conditional on the choice to bet).[48] As before, betting on team 1 offers a potential payoff of y_1 per dollar wagered and betting on team 2 offers a potential payoff of y_2 per

[48]Theoretically, we could expand the choice set of contracts to include Spread contracts as well. However, our empirical exercise relies upon being able to calculate market shares in each betting contract for each game. Unfortunately, we do not have reliable data to calculate the relative share of betting in Spread versus Moneyline contracts. Anecdotally, it also appears that the choice between betting in the Spread and Moneyline contracts might be driven by factors other than beliefs about the game's outcome.

dollar wagered. Because bettors are risk-neutral, they purchase the contract that provides them with the highest subjective expected return. When both contracts trade in equilibrium, this means there is a $\bar{b}$ such that $b > \bar{b}$ bets on team 1 and $b < \bar{b}$ bets on team 2 in the Moneyline, since subjective expected returns for betting on team 1 (team 2) are increasing (decreasing) in b.[49]

Bettor $\bar{b}$ is the marginal bettor in each game that prices contracts in equilibrium, in the sense that his belief is reflected in the price of the contracts offered. More negative values of $\bar{b}$ mean that the marginal bettor is more (incorrectly) optimistic about the underdog winning the game, and accordingly corresponds with more negative returns for betting on the underdog. In order to reproduce the Favorite-Longshot Bias in the data, we expect $\bar{b}$ to be more negative for games in which the outcome is more extreme.

Given the assumption that bettors sort into contracts based on their valuations, we can re-write the market shares for each contract (the proportion of betting on each contract) in terms of a cutoff bettor, $\bar{b}$, and the belief distribution $h(b)$.

$$s_1 = \int_{\bar{b}}^{\infty} h(b) \qquad \text{(Favorite Market Share)}$$

$$s_2 = \int_{-\infty}^{\bar{b}} h(b) \qquad \text{(Underdog Market Share)}$$

We measure market shares in two ways. First, we measure the market share of a contract as the proportion of the number of bets placed on the contract in a game (which is directly provided to us in the data). Second, we estimate the market share of a contract as the proportion of dollars bet on the contract assuming the bookmaker takes no risk. To estimate market shares in this way, we re-write the market clearing conditions in Equations (15) and (16) in terms of market shares (below). We then directly solve for market shares for each game by equating the market clearing conditions, where $s_1 + s_2 = 1$. Both measures of market shares have different interpretations for our results, which we discuss in more depth below.

$$s_1(1 + y_1) = s_1 + s_2 \tag{B.1}$$

$$s_2(1 + y_2) = s_1 + s_2 \tag{B.2}$$

[49]Partitioning by beliefs also occurs in a context where each bettor is risk-averse and makes a discrete choice to either bet the same amount on team 1 or team 2. This is because the subjective expected utility of a risk-averse bettor wagering a fixed amount on team 1 (team 2) is increasing (decreasing) in b. The discrete choice assumption with equal betting amounts is made elsewhere in related studies (Gandhi and Serrano-Padial (2015) and Chiappori et al. (2019)). Without imposing discrete choice, a risk-averse bettor may choose to bet on multiple contracts, because of a hedging motive. Similarly, while partitioning by beliefs would occur for risk-averse bettors that decide between betting the same amount on either team, it does not necessarily occur when bettors have the choice to bet different amounts on the two contracts.

The likelihood function of bettor b choosing contract j can be expressed in terms of the cutoff agent and the distribution of beliefs

$$\text{likelihood}(j;b) = \begin{cases} \int_{-\infty}^{\bar{b}} h(b) & \text{if } j = 1 \\ \int_{\bar{b}}^{\infty} h(b) & \text{if } j = 2 \end{cases} \tag{B.3}$$

Using the market shares, we can express the likelihood of a sample of games $k = 1, \ldots, K$ as

$$\mathcal{L} = \prod_{k=1}^{K} \prod_{j_k \in J_k} \text{likelihood}(j_k; b_k)^{s_k^j} \tag{B.4}$$

and the corresponding log-likelihood of the sample as

$$\log \mathcal{L} = \sum_{k=1}^{K} \sum_{j_k \in J_k} s_k^j \log(\text{likelihood}(j_k; b_k)). \tag{B.5}$$

B.2 Empirical Implementation

To estimate belief heterogeneity in the data, we need to express the likelihood function in Equation (B.3) in terms of parameters we can estimate. Because bettors are risk-neutral, $\bar{b}$'s indifference condition is expressed as

$$\bar{p}y_1 - (1 - \bar{p}) = (1 - \bar{p})y_2 - \bar{p} \tag{B.6}$$

Re-arranging terms yields

$$\bar{p} = \frac{y_2 + 1}{y_1 + y_2 + 2} \tag{B.7}$$

We then convert the corresponding value of $\bar{p}$ for each contract to $\bar{b}$ by computing

$$\bar{b} = F^{-1}(\bar{p}) - F^{-1}(p) \tag{B.8}$$

where F^{-1} is the inverse CDF of the standard normal distribution, and p is the objective probability that team 1 wins the game. As in Section 3.2 (and discussed in Appendix A), we use the nonparametrically estimated win probabilities for the objective win probabilities.

We assume $h(b)$ is logistically distributed with mean zero. The parametric assumption is used primarily for simplicity, and our interpretation of the results does not vary with this assumption. Given the parametric assumptions imposed, the scale parameter, which controls the dispersion of the belief distribution, is the only parameter to be estimated.

B.3 The Two Market Share Measures

We estimate the model two measures of market shares: the proportion of bets placed on a team, and the proportion of the dollar value of bets placed on a team. In the case that each bettor places wagers of the same size, as is commonly assumed in the literature, both of these measures of market shares are exactly the same. However, given our evidence, it is unlikely that bettors place wagers of the same size on each contract.

The estimated proportion of dollars bet on contracts is strictly increasing in the probability of the contract paying off. Bets on underdogs offer a high payoff with a low probability, while bets on favorites offer a low payoff with a high probability. For the market to clear without the bookmaker taking any risk, more dollars must be bet on the favorite than the underdog in equilibrium, as can be observed via Equations (B.1) and (B.2). However, the relationship between the proportion of bets placed on a contract and the probability of the contract paying off is an empirical question, since in practice, bettors may wager different amounts when betting on the favorite and the underdog.

As it applies to the interpretation of our estimated belief distribution, measuring a contract's market share as the proportion of bets placed on the contract yields an estimate of the distribution of beliefs in the population, giving each individual bettor equal weight. Measuring a contract's market share as the proportion of dollars wagered on the contract yields an estimate of the dollar-weighted distribution of beliefs in the population, where each bettor's belief is weighted by the number of dollars wagered by the bettor.

The empirical relationship between the proportion of bets placed and the dollar proportion of bets placed on a contract, as presented in Figure 10 and discussed in the main text, foreshadow the results of the estimation results. First, the flat relationship between the two suggests that the estimated distribution of beliefs is more dispersed than the dollar-weighted distribution of beliefs. Second, the non-monotonicity of the relationship suggests that belief heterogeneity alone will not be sufficient to explain the distribution of beliefs. Belief heterogeneity can explain the under-performance of extreme underdogs when the marginal bettor betting on extreme underdogs has more extremely distorted beliefs. However, belief heterogeneity cannot explain the *increase* in the proportion of bettors for extreme underdogs that we observe in the data.

B.4 Expected Returns from Estimated Belief Distribution

We construct the implied expected returns for each contract using the estimated belief distribution and observed market shares. Using Equation (B.8), we estimate the objective probability that team 1 wins (and hence expected returns) as $p = F\left(F^{-1}(\bar{p}) - \bar{b}\right)$, where F and F^{-1} are the standard normal CDF and inverse CDF, $\bar{b}$ is the belief of the marginal bettor and $\bar{p}$ is team 1's win probability as perceived by $\bar{b}$. We directly calculate $\bar{p}$ from contract payoffs in each game via Equation (B.7). We calculate $\bar{b}$ using observed market shares for each game and the calibrated belief distribution via the equation $\bar{b} = H^{-1}(s_2)$, where H^{-1} is the inverse CDF function of the calibrated belief distribution.

B.5 Empirical Results

We calibrate the belief distribution using both market share measures via Maximum Likelihood, which provide estimates of the scale parameter of the belief distribution. When using dollar market shares, we estimate the scale parameter as 0.0718 (standard error of 0.0012) - the corresponding estimated standard deviation of b is $\sigma_b = 0.13$. When using market shares, we estimate the scale parameter as 0.302 (standard error of 0.014) - the corresponding estimated standard deviation of b is $\sigma_b = 0.547$.

To assess the ability of the model to explain the Favorite-Longshot bias with both assumptions, we compare the model implied expected returns with the realized returns across contracts. We sort each contract into deciles based on the Moneyline. We plot the model implied expected returns and the average realized returns for each decile using both market share measures in Figure B.1. The first plot in the figure corresponds with the model implied returns assuming bettors each wager the same amount. The plot reveals a Favorite-Longshot Bias in expected returns that is quantitatively similar to that observed in realized returns, though the magnitude of the estimated Favorite-Longshot Bias is slightly larger than observed in the data.

The second plot in the figure corresponds with expected returns implied using the actual proportion of bets. The model-implied returns exhibit a non-monotonic Favorite-Longshot Bias and the model does a considerably worse job of capturing Favorite-Longshot Bias than using the dollar weighted belief distribution. The model underestimates the returns for extreme underdogs and overestimates the returns for more moderate underdogs. Extreme underdogs (in the first decile) actually earn higher equilibrium returns than more moderate underdogs in the model calibration, opposite what we observe in the data. The failure of heterogeneous beliefs to explain the Favorite-Longshot bias using the actual proportion of bets stems from the fact that the proportion of bets placed is not monotonic in odds, with a substantial uptick of bettors choosing to wager on the underdog in games where the outcome is expected to be more extreme.

B.6 Heterogeneous Beliefs and Choice Sets

One potential obstacle for the conclusions we draw is that they assume the distribution of beliefs across games is the same, with bettors choosing either to bet on the underdog or favorite in a particular game. In reality, bettors have a wider choice set available, which can challenge the assumption that the distribution of beliefs of bettors choosing to wager on a game is the same across games. For example, bettors may have subjective beliefs about the outcome of each game that is played on a particular date, and, based on these beliefs, may choose the betting contract *across* games that they find most attractive. In this example, heterogeneous beliefs may be consistent with our finding that, relative to games with more moderate underdogs, a higher proportion of bettors choose to wager on the underdog in games with more extreme underdogs. It is possible that with other alternative games to bet on, bettors that would otherwise bet on a game with an extreme expected outcome choose to bet on games with more moderate expected outcomes, which they subjectively believe offer more attractive returns. With an appropriately shaped distribution that beliefs are drawn from, where many bettors have moderately biased beliefs that lead them to find extreme favorites to be less attractive than other contracts, and where a sufficient number of bettors have beliefs that make them find extreme underdogs to be the most attractive bet, we may be able to find the pattern of increasing proportion of bettors choosing the underdog in more extreme games.[50]

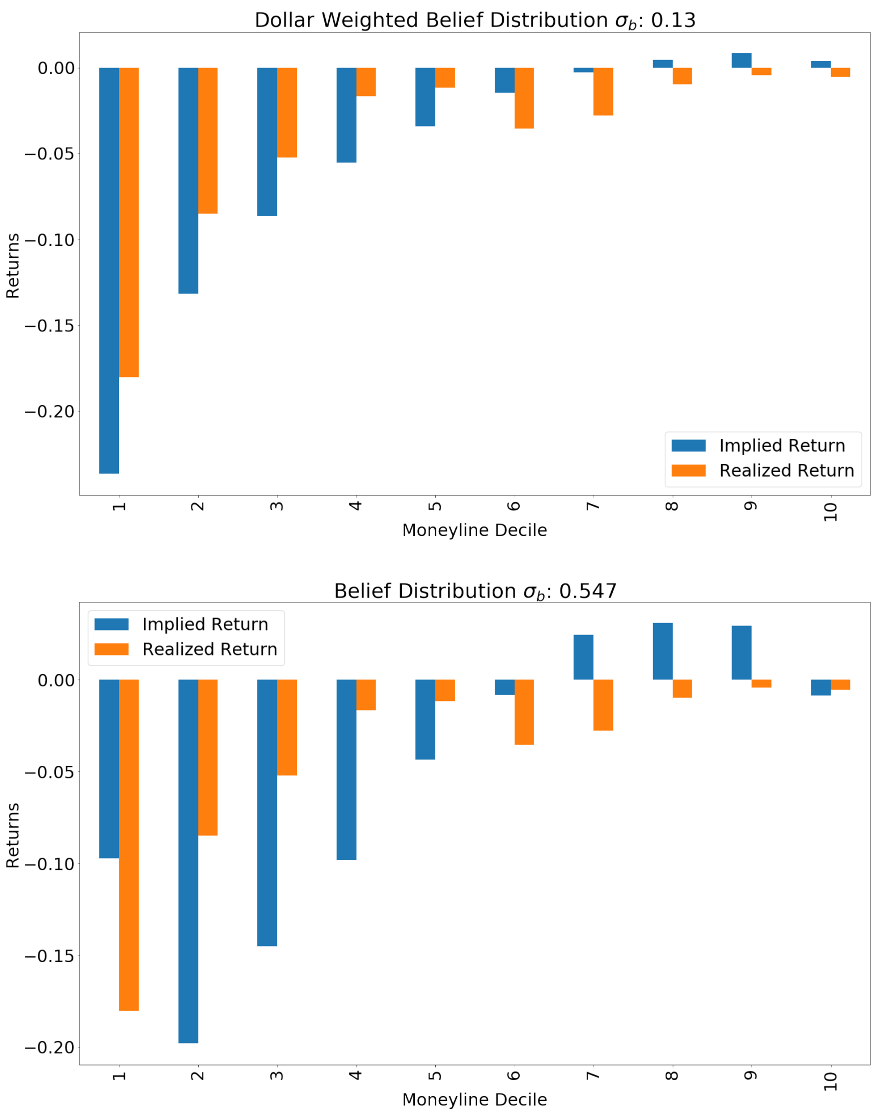
We do not have the betting volume data required to estimate a model in which the choice set includes all potential games on a date. However, an implication of this type of heterogeneous beliefs model is that the allocation of bettors into different contracts will vary depending upon the choice set available on each date. For example, when there is only one game available to bet on, all bettors must choose to bet either on the favorite or the underdog. When there are two games available to bet on, bettors with more extreme beliefs may choose to bet on the more extreme underdog (which earns the lowest returns), while bettors with more moderate beliefs may choose to bet on the other three available contracts. Under this alternative explanation for our empirical evidence, we expect the proportion of bets placed on extreme underdogs and the returns of these underdogs to vary based on the other betting opportunities available on a given date.

To test this alternative explanation, we consider the decile of the most extreme underdogs, and analyze (1) how the proportion of bettors choosing to bet on the underdog and (2) how betting returns, vary with the available choice set on a given date. We construct two measures to capture the relevant choice set for a particular contract. The first is simply the number of other betting contracts available on the same date as the contract considered. The second is the proportion of bets on underdogs available on the same date with a higher probability of paying off (with zeros if no other games are available to bet on). We regress

[50]This alternative may be closer to the studies of belief heterogeneity at the horse racetrack, where bettors choose from a number of horses in a race.

Figure B.1: Heterogeneous Beliefs and Expected Returns

The figure plots the average returns and expected returns of Moneyline contracts for games that we have betting volume data for, sorted into deciles based on the Moneyline. Decile 1 corresponds with contracts betting on the most extreme underdogs and Decile 10 corresponds with contracts on the most extreme favorites. Expected returns are computed using the observed market shares on each contract and the calibrated belief distribution. In the first plot, expected returns are estimated using the dollar-weighted belief distribution, estimated by assuming that bettors each wager the same amount. In the second plot, the expected returns are estimated using the belief distribution, weighing each bettor equally in the calculation for the belief distribution. Each plot also includes σ_b, the estimated standard deviation of the belief distribution implied by the model.



contract returns and the proportion of bets placed on the underdog on these variables, including a control for the contract odds. In effect, these regressions capture whether, for a given odds contract, the betting returns and proportion of bets placed on the contract vary with the set of alternative contracts that are available. We standardize all independent variables in the regressions to have zero mean and unit standard deviation. If sorting across games explains our results, we expect lower returns and a higher proportion of bettors betting on the underdog corresponding with increases in the independent variables. This explanation for our results would suggest that with a greater amount of alternative betting opportunities in games with more moderate expected outcomes, a smaller (larger) proportion of bettors choose to bet on the favorite (underdog) in games with extreme expected outcomes, and the bettors choosing to bet on extreme underdogs will have more extreme beliefs.

Table B.1 reports the results from the regressions. The first three columns correspond with results where betting returns are the dependent variable, and the last three columns correspond with results where the proportion of bets placed are the dependent variable. The regression results suggest that returns are decreasing in odds, and the proportion of bets placed are increasing in odds. Since higher odds correspond with more extreme underdogs, these results are consistent with the Favorite-Longshot bias, and the increasing proportion of bettors betting on more extreme underdogs. Next, we turn to the variables of interest in the regression. First, the coefficients on the number of other betting contracts available is not statistically significant in any of the regressions, and in fact has the opposite sign as expected when the betting proportion on the underdog is the dependent variable. Turning to the proportion of other underdog contracts with a greater probability of paying off, the regressions suggest that a one standard deviation change in this variable corresponds with a 1% increase in the proportion of bettors wagering on the underdog. This is small compared with the average proportion of bettors (41%) choosing to bet on the underdog in this decile of contracts. Additionally, the coefficient on this variable is negligible in regressions where the independent variable is contract returns, suggesting that choice sets do not appear to explain any variation in the returns of extreme underdogs. Overall, the regression results do not support an alternative explanation for our results where bettors sort into different games based on their beliefs.

An additional point of note in this discussion is that bettors are also free to bet on multiple events, and are not restricted to betting on a single event. In a sample of 336 individual bettors, Andrikogiannopoulou and Papakonstantinou (2016) report that when betting, the average bettor in their sample places a little more than five bets per day. This provides some evidence against the idea that bettors select a single game where they find the expected returns to be most attractive, as sorting into games by beliefs might suggest.

Table B.1: Betting on Underdogs and Choice Sets

The first three columns in the table present regression results from a series of regressions of contract betting returns on the number of other contracts offered on the same date ("Num Contracts") and the proportion of contracts on underdogs with greater probability of paying off offered on the same date ("Proportion"), where the sample is the decile with the most extreme underdogs. The last three columns in the table report regression results with the same sample and independent variables, where the dependent variable is the proportion of bets placed on the underdog. t-statistics are reported in parentheses.

	Betting Returns			Proportion of Bets		
	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	-0.23	-0.23	-0.23	0.41	0.41	0.41
	(-7.33)	(-7.33)	(-7.33)	(126.70)	(125.89)	(126.10)
Proportion	0.00		0.00	0.01		0.01
	(-0.01)		(0.03)	(3.91)		(3.93)
Num Contracts		-0.05	-0.05		-0.01	-0.01
		(-1.69)	(-1.69)		(-1.92)	(-1.95)
Odds	-0.11	-0.11	-0.11	0.03	0.03	0.03
	(-3.30)	(-3.52)	(-3.35)	(6.20)	(7.81)	(6.13)
N	7349	7,349	7,349	5,451	5,451	5,451

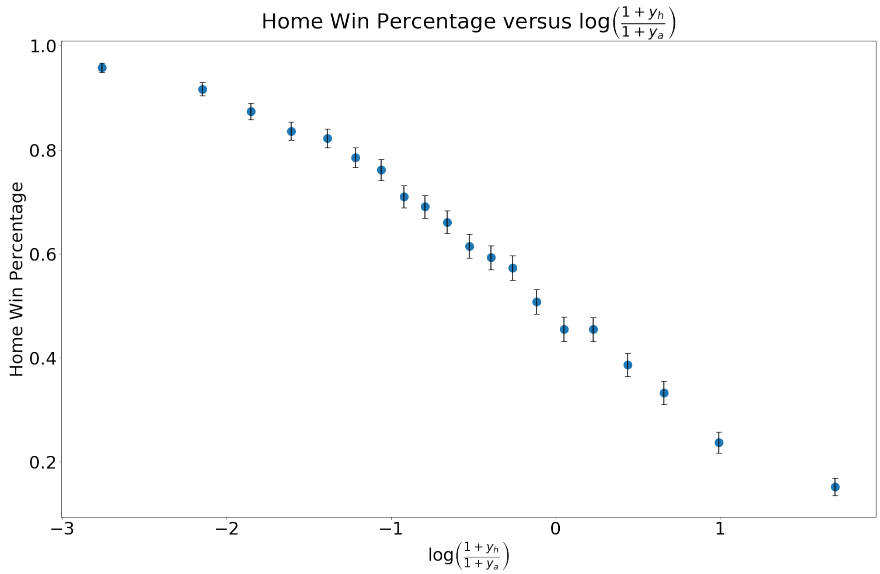
B.7 Discussion

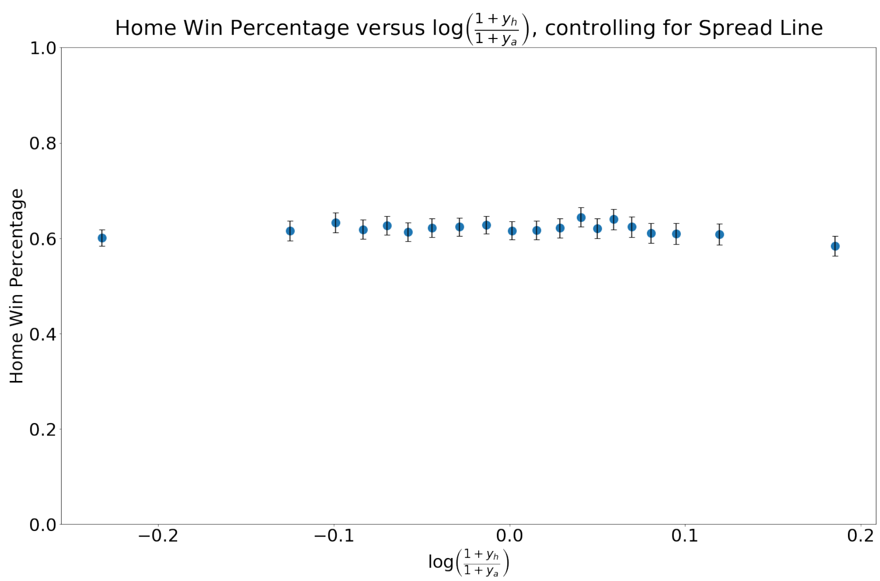
While we believe preferences play an important role in the results, the results do suggest a potential role for belief heterogeneity in the Favorite-Longshot Bias as well. In our stylized model of heterogeneous beliefs, we estimate that the standard deviation of the dollar-weighted belief distribution and the unweighted belief-distribution, expressed in units of the standard deviation of point-spread distribution, are 13% and 54.7%. Compared with numbers found in financial markets, these numbers imply a reasonable degree of belief heterogeneity. For example, Giglio et al. (2021) report in a survey of investors that the standard deviation of expectations of the one-year return of the S&P 500 is 5.2%. The annualized volatility of monthly S&P 500 returns is about 15%, so belief heterogeneity about stock market returns, as a proportion of stock market volatility, is about 35%. Similarly, using the I/B/E/S analyst database, which provides analyst estimates of quarterly stock level earnings per share for a number of firms, (analyzing all listed US stocks in the database from 1990 to 2017), we find that the average standard deviation of quarterly earnings forecasts, as a percentage of the trailing 2-year standard deviation of a firm's earnings, is 35%. With these numbers from financial markets providing context, the degree of belief heterogeneity revealed by our model is quite reasonable. While belief heterogeneity is unlikely to be the whole story, it may play a role in explaining the patterns in the data in combination with preference-based explanations.

C Additional Tables, Figures, and Analyses

Figure C.1: Wins, Moneylines, and Spreads

The figure plots the relationship between the probability the home team wins and the quantity $\log\left(\frac{1+y_h}{1+y_a}\right)$ (the "Moneyline ratio"), where y_h and y_a are the payoffs associated with winning bets on the favorite and the underdog in the Moneyline. The first plot in the figure plots a binned scatterplot, where each game is sorted into twenty bins by sport based on the Moneyline ratio, and the figure plots the average Moneyline ratio versus the home win percentage in each bin. The second plot in the figure plots the same quantities, including a control for the Spread Line of the game. The error bars correspond with plus/minus two standard errors relative to the home win percentage in each bin.





C.1 Distribution of Point Differential Minus Spread Line

In this section, we discuss the assumption that the the point-differential minus spread line is distributed *iid* normal across different games within a sport.

First, as we document in our main results, spread contracts are priced such that there is approximately an equal probability of winning and losing the bet, with approximately equal returns regardless of the spread line (as captured by Figure 3). This suggests that the quoted spread line reasonably accurately captures the median of the distribution. We similarly analyze how the second moment of the realized-point differential minus spread line varies across games. We divide games into deciles based on the quoted Spread Line in each sport, and plot the standard deviation of each decile in Figure C.2. The standard deviation does not exhibit any systematic pattern across games and appears to be relatively constant within each sport. Lastly, one implication of the assumption of the the distribution of the realized point differential minus the quoted spread line being identically distributed is that the quoted spread line is a sufficient statistic to summarize the probabilities of either team winning the game. This, in turn, means that, the spread line should be able to explain the prices of Moneyline contracts. We find that spread lines are able to explain more than 89% of the variation in Moneyline contract prices in each sport, consistent with this implication. Regression results are presented in Table C.1.

We plot a histogram of the realized point differential minus the quoted spread line for each sport in Figure C.3. The point-spread distributions are reasonably well behaved and appear to be approximately symmetric and bell-shaped. For each sport, the figure also overlays the Maximum Likelihood Estimate of the probability distribution function assuming that the data follow a normal distribution. The normal distribution provides a reasonably good, though certainly far from perfect fit of the data.

We can also estimate the implied win percentages for games using the Spread Line and the assumption that the realized point differential minus Spread line distribution is mean zero and identically normally distributed across games in each sport. To evaluate the identical normality assumption, in Figure C.4, we estimate the implied win percentage based on the normality assumption, then group games into 50 equally spaced bins based on their estimated implied win percentage, and plot the average win percentage against the average implied win percentage of each bin. In general, the estimated win percentages line up with the implied win percentages on a 45 degree line, especially for college sports. The estimates are more noisy for the NFL, and for the NBA, the implied win probabilities do appear to be higher than the true win probabilities for strong underdogs and lower than the true win probabilities for strong favorites.

Figure C.2: Point-Spread Standard Deviation By Spread Decile

The figure plots the standard deviation of the difference between the point-spread and spread-line for games grouped into deciles based on the quoted Spread Line on the game. Decile 1 correspond with the home team being strongly favored and Decile 10 corresponds with the home team being strong underdogs. Each panel in the figure corresponds with the values for a particular sport. The error bars corresponds with 95% confidence intervals for the the decile standard deviation.

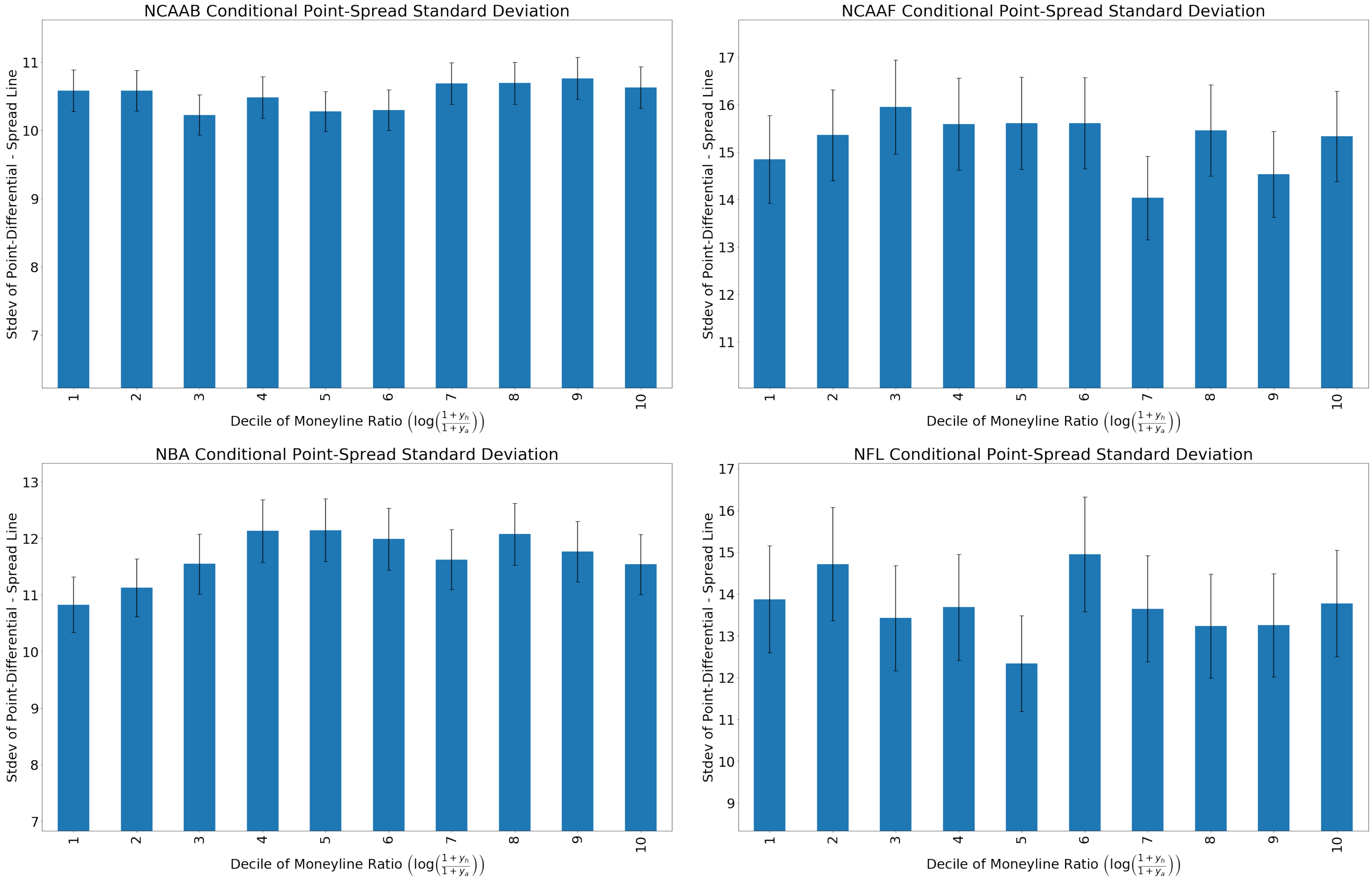


Table C.1: Relationship between Moneyline and Spread Lines

The table returns results from the regression of transformed Moneylines on Spread Lines within each sport. The transformed Moneyline variable is the log of the payoff for a winning $1 bet implied by the Moneyline. The sample includes both the Moneyline contract on the home team and the Moneyline contract on the away team for each game. For observations that correspond with the Moneyline contract on the away team, we multiply the spread line by negative one. Standard Errors are reported in parentheses.

	NCAAB	NCAAF	NBA	NFL	Combined
Intercept	0.845	0.885	0.801	0.767	0.836
	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)
Spread Line	0.076	0.056	0.078	0.071	0.071
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
R^2	89.2%	88.0%	89.3%	93.0%	87.5%
N	43,964	8,784	16,784	3,686	73,218

Figure C.3: Conditional Point-Spread Distribution by Sport

The figure plots a histogram of the difference between the point-spread and the posted spread line across games, with one panel for each sport. Each panel also overlays the best fit normal distribution for the data, calculated via Maximum Likelihood.

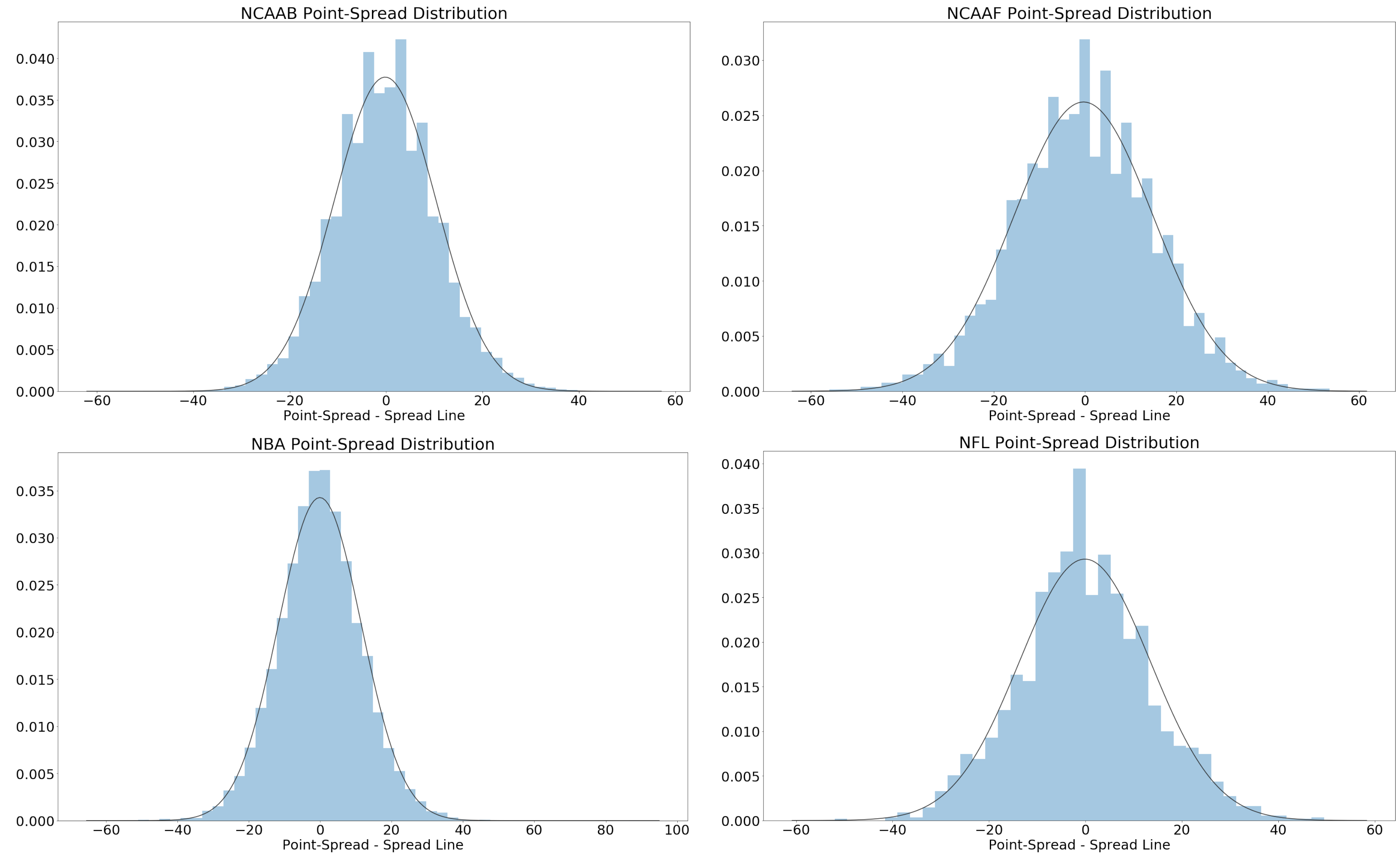


Figure C.4: Win Probabilities Implied by Normality

The figure plots the implied win probability of games assuming that the Spread line minus point-differential distribution is identically normally distributed across games versus actual win percentages. The plot is formed by estimating the standard deviation of the Spread Line minus point-differential distribution for each sport, σ_s. Then, for each game an implied win percentage is calculated, based on the Spread Line and the assumption that the Spread Line minus point-differential is distributed $N(0, \sigma_s)$. All contracts are split into 50 equally spaced bins. Each plot plots the average estimated win probability for each bin on the x-axis and the percentage of games one on the y-axis. Each plot also plots the 45 degree line as a dotted red line.

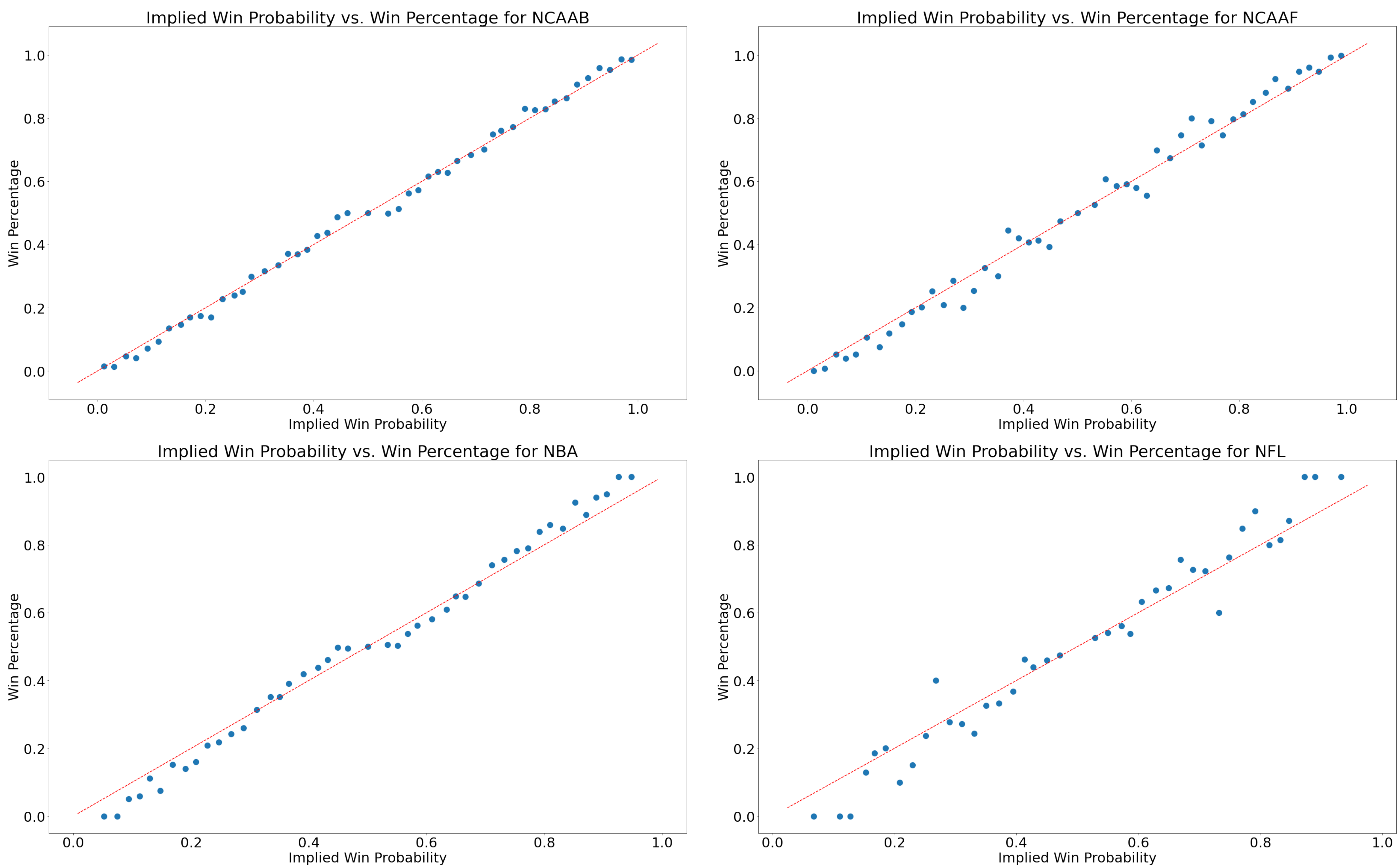


Figure C.5: Accuracy of Over/Under Contracts by Sport

The figure presents binned scatter plot of the Over/Under line versus the total number of point scored in a game for each sport. Each game in the sample is sorted into 20 equal sized bins based on the Over/Under line of the game. Each point on the plot corresponds with the average Over/Under Line and the average point total of each game in one of the bins. The 45 degree line is also plotted on the graph in red. The error bars correspond with plus/minus two standard errors relative to the mean number of points scored in a bin.

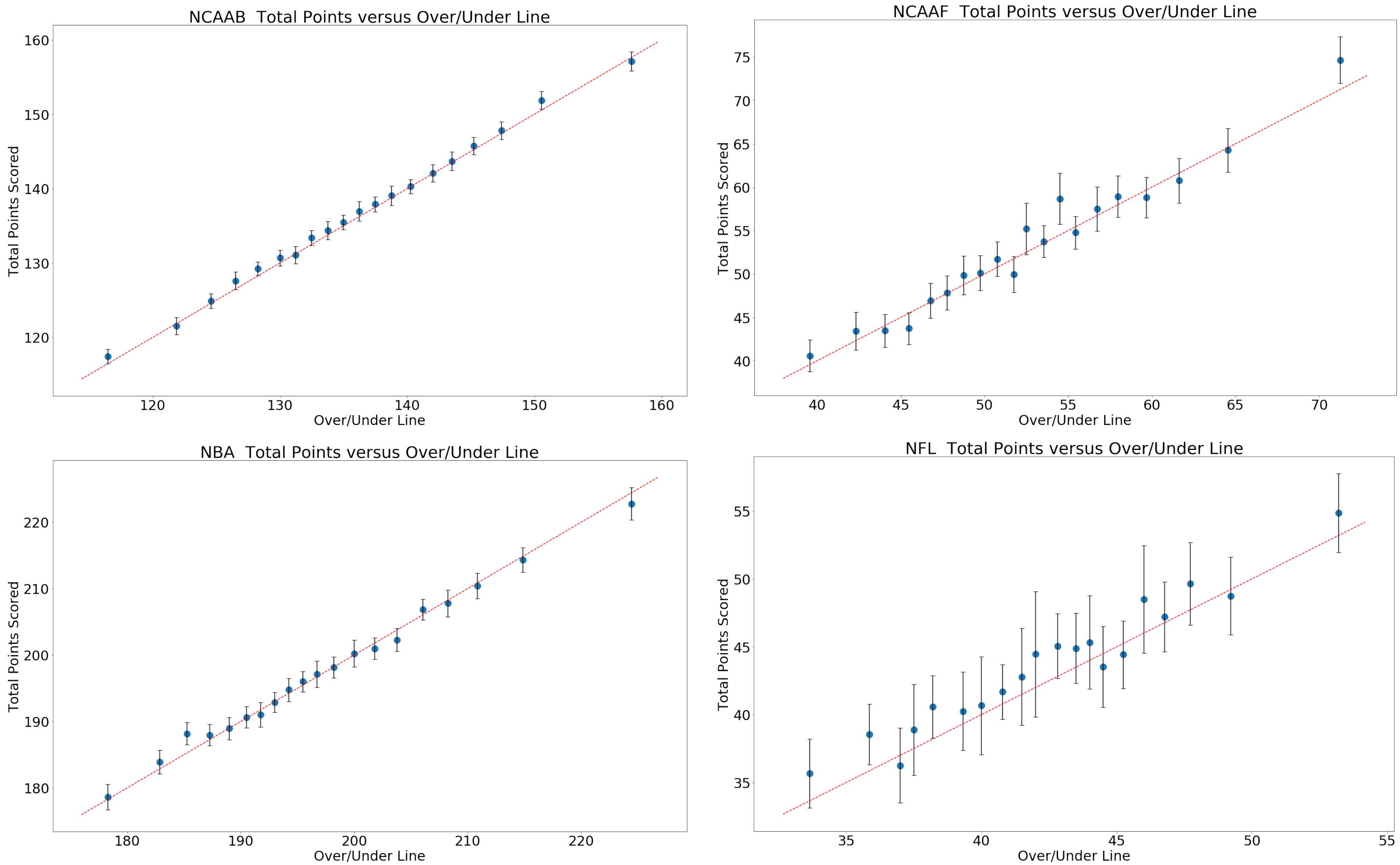


Table C.2: Index Option Sample

The table reports the list of instruments included in our sample of index options along with the corresponding date ranges and ticker symbols.

Name	Ticker	Start Year	End Year
CBOE INT RATE 30 YR T B	TYX	1996	2010
CBOE TREASURY YIELD OPT	TNX	1996	2011
DOW JONES INDEX	DJX	1997	2019
NASDAQ 100 INDEX	NDX	1996	2019
NYSE ARCA MAJOR MARKET	XMI	1996	2008
PSE WILSHIRE SMALLCAP I	WSX	1996	2000
RUSSELL 2000	RUT	1996	2019
S&P100 INDEX	OEX	1996	2019
S&P500 INDEX	SPX	1996	2019
S&P MIDCAP 400 INDEX	MID	1996	2019
S&P SMALLCAP 600 INDEX	SML	1996	2019

C.2 Details on Options Data

For our analysis on the options implied volatility surface, we use all available equity options in the Option Metrics database, and we use data on index options for eleven indices, whose details are listed in Table C.2.

Listed index options are European options, while listed equity options are American. Implied volatilities for European options are calculated using the Black-Scholes formula. Implied volatilities for American options are calculated by OptionMetrics using the Cox et al. (1979) binomial tree model.

The volatility surface file is provided by OptionMetrics and constructed for each security on each day by using a kernel smoothing algorithm that interpolates implied volatilities from listed options to provide implied volatilities for call options and put options with fixed expirations and option deltas.

Figure C.6: Cross-Section of Betting Returns with Opening Lines

The figure reproduces first Panel of Figure 1, the first Panel of Figure 2, and Figure 3, using opening lines to sort contracts and calculate returns rather than the closing lines.

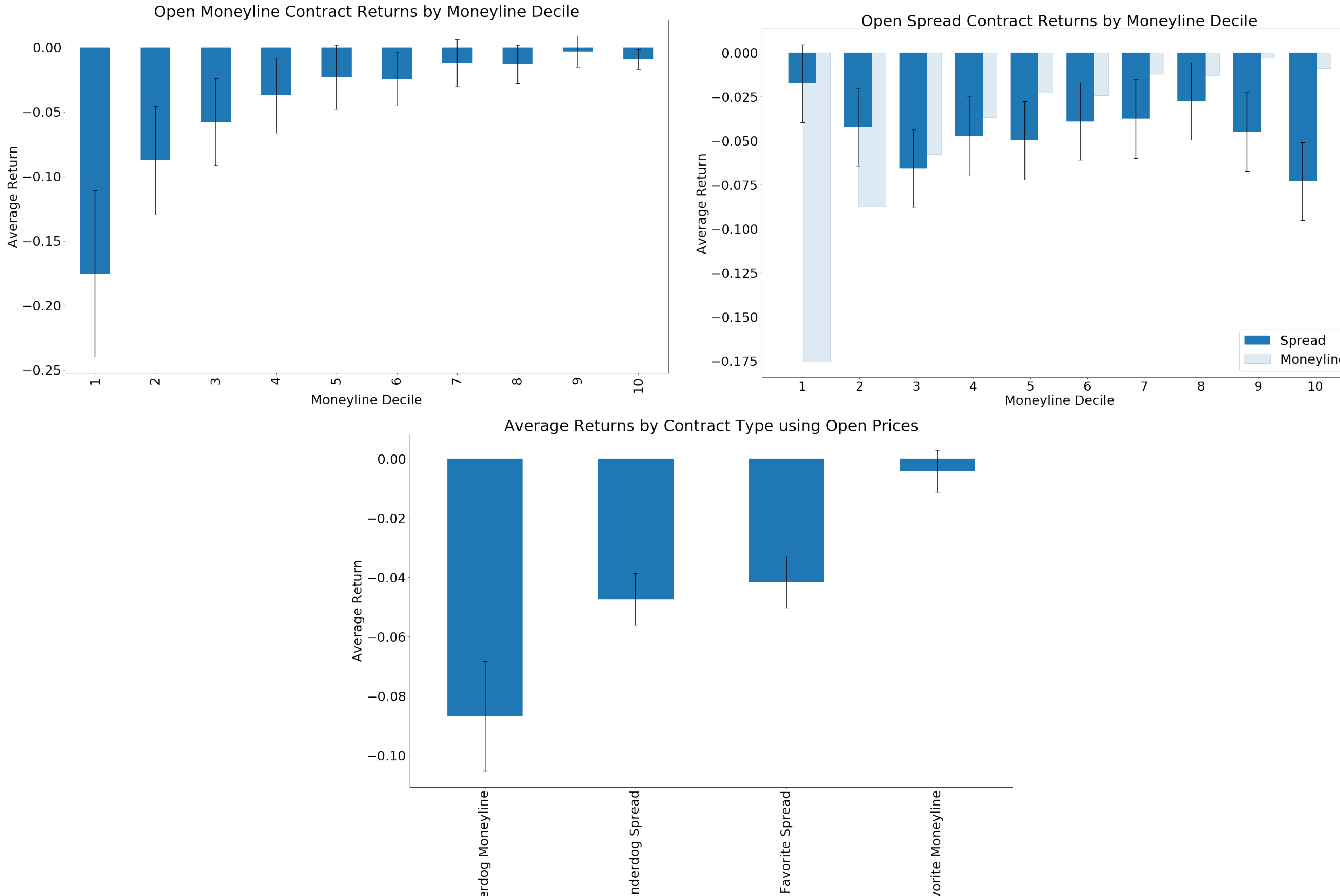


Figure C.7: Average Moneyline Contract Return Across Games

The figure plots the average returns of Moneyline contracts, sorted into deciles based on the Moneyline and split into bets on home and away teams. Decile 1 corresponds with contracts betting on the most extreme underdogs and Decile 10 corresponds with contracts on the most extreme favorites. The error bars correspond with two standard errors above and below the average return for a particular decile. The first plot is the subset of bets made on the home team and the second plot is the subset of bets made on the away team.

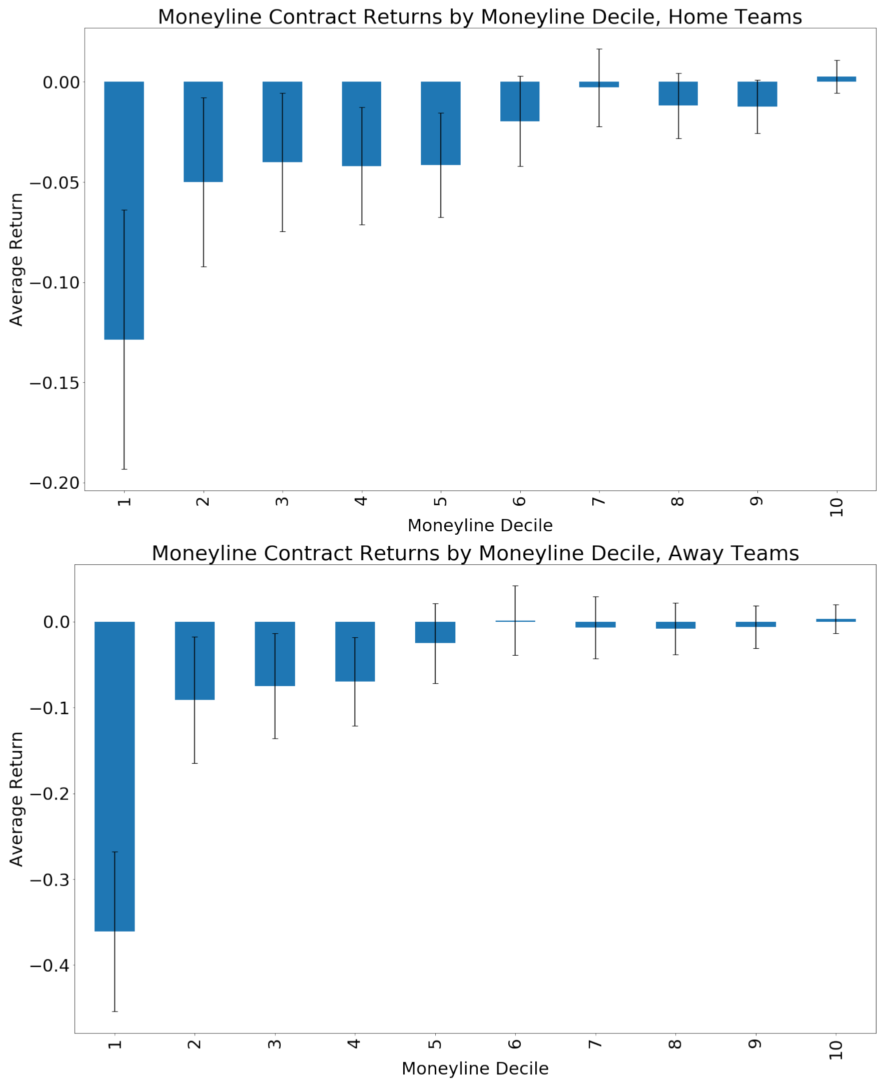
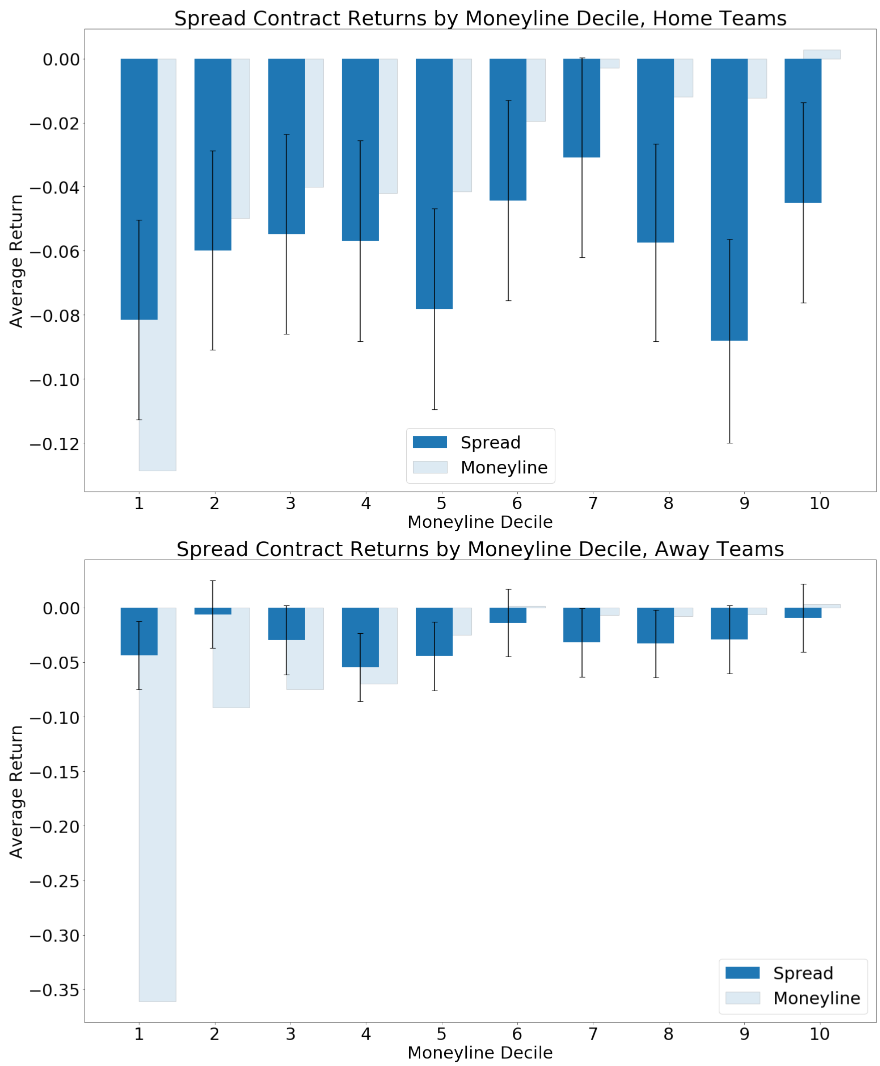


Figure C.8: Average Spread Contract Return Across Games

The figure plots the average returns of contracts sorted into deciles based on the Moneyline and split into bets on home and away teams. Decile 1 corresponds with contracts betting on the most extreme underdogs and Decile 10 corresponds with contracts on the most extreme favorites. The darker bars in the foreground correspond with Spread contract returns. The error bars correspond with two standard errors above and below the average return for Spread contracts in a given decile. The lighter bars in the background correspond with Moneyline contract returns, as presented in Figure C.7. The first plot is the subset of bets made on the home team and the second plot is the subset of bets made on the away team.



C.3 Bet Sizes and Odds

In order to estimate bet sizes in our sample, we use the market clearing condition assuming that bookmakers take no risk. Naturally, this assumption may affect the patterns of bet sizes that we estimate in the data. To alleviate potential concerns associated with this assumption, we conduct an out of sample test of the relationship between odds and bet sizes in a sample of soccer matches from Betfair, an exchange with no bookmakers; the results are plotted in the main text in Figure 8. Here, we present additional details about the sample and the data.

For the sample of soccer matches, our data contain the total dollar volume of bets placed on the home and away teams for each game in our sample, as well as the total number of bets on the home and away teams for each game. This allows us to directly measure the average bet size of wagers placed on the home and away teams without making assumptions about bookmakers.

The sample ranges from 2006 to 2011 and contains data on games in eighteen major soccer leagues: the Belgian Jupiler league, the Dutch Eredivisie, the English Championship League, the English Premier League, the French Ligue 1, the French Ligue 2, the German Bundesliga 1, the German Bundesliga 2, the Greek super league, the Italian Serie A, the Italian Serie B, the Portuguese Super Liga, the Scottish Premier League, the Spanish Primera Division, the Spanish Segunda Division, and the Turkish Super League.